Topics in the Accuracy and Resolution of Superresolution Systems

David Brandwood

A thesis submitted for the degree of

Doctor of Philosophy

of

University College London

Department of Electronic and Electrical Engineering

University College London

September 2009

Declaration

I, David Herbert Brandwood, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: D.H. Brandwood

12 September 2009

Abstract

Since their introduction in the 1970s and 1980s superresolution systems for point source parameter estimation have received theoretical attention regarding their potential performance. Two aspects of performance in particular are of interest, the accuracy of the parameter estimation and the resolution achievable. Limitations on performance may be considered to be due to noise affecting the data, or to errors in the system.

Superresolution methods divide roughly into two groups – 'spectral' methods and maximum likelihood (ML) methods. MUSIC is perhaps the most effective example of a spectral method and has been studied in considerable detail, in both performance measures, but mainly only for the case of a single parameter. In this study the accuracy of MUSIC in the application of two-dimensional direction finding (DF) has been analysed, with and without system errors, using a general array. Theoretical results are confirmed by simulations. An aim has been to produce simpler results for use in estimating the potential performance of practical systems.

Little work has been reported on the resolution of ML methods and this is the second main topic of this work, particularly for the two-dimensional DF case using a general array, with a ML method (IMP) similar to the better known Alternating Projection. Some results are obtained for resolution with and without errors for the case of non-coherent signals. For coherent signals (including the standard radar case) the performance is found to depend on the relative phase of the signals, varying from the quadrature case, where the performance is as for the non-coherent case, to the in-phase (or antiphase) case where only one signal peak is seen.

Principal Results and Original Contributions

1. MUSIC accuracy with errors

(a) The study of DF accuracy using MUSIC has been extended to the case of two angle parameters (azimuth and elevation) using general arrays (2- or 3-dimensional, not necessarily regular). The main study is of the common case of arrays of similar elements, similarly oriented. (This is termed the EPP case in the thesis.)

(The study is limited to the case where the limitations on accuracy are due to errors alone - the effects of noise and finite sampling are taken to be negligible compared with the effects of errors. This corresponds to the large signal, high sample number case.)

(b) The study has been further extended to the case of general arrays of non-EPP elements.

(c) The method used is independent of eigenanalysis, using the equivalence of the vector space basis given by the principal eigenvectors (of the system covariance matrix) and the basis given by the true point source response vectors (PSVs) of the signals. (This is a valid approach for the condition of the study given in (a) above.) This leads to considerably simpler and more convenient expressions for the statistics of the measurement errors.

(d) It is shown that in the case where only a single target is present (for which both the superresolution methods MUSIC and IMP are equivalent to simple beamforming) the errors in angle estimation depend only on the system phase errors and are independent of amplitude errors (at least to first order, which is valid up to quite substantial errors, covering likely error levels in most well engineered systems). For two or more targets amplitude errors also affect the measurement accuracy. These theoretical results are confirmed in simulation. If only phase errors are present the two signal acccuray is lose to that of the targets present singly.

(e) An array sensitivity function is defined (following Friedlander [15]) relating the errors in angle measurement to the system errors. Contour plots of this function can be drawn over all angle space for any given array.

(f) A very simple expression for a rule of thumb for the sensitivity of a circular array with elements at half-wavelength spacing is obtained, and another for a more general planar array.

(g) The structure of the angle error covariance matrix due to system (PSV) errors is

shown to be similar to the CRB matrix, determined by noise. This enables an equivalence to be drawn between the levels of system errors and receiver noise.

2. Array moment matrix

An array second moment matrix is introduced and shown to be significant and useful in expressions of (i) array error sensitivity, (ii) CRB, (iii) approximate array beamwidth and (iv) resolution.

3. Cramér-Rao Bound

(a) The CRB is given in a form based on the array moment matrix

(b) A relatively simple form for the CRB for the case of two signals and two parameters is derived from the more general (multiple signal, multiple parameter) expression.

4. IMP resolution

(a) Developing the approach of Speirs and colleagues [41], an expression for the resolution of two non-coherent signals using a ML method (such as IMP) was obtained, showing the dependence on the power of the weaker signal, assuming the other signal is strongly dominant. This expression uses the array moment matrix. The method was extended to the case of two signals in general, showing the dependence on both signal powers.

(b) An approximate resolution improvement factor is defined as the ratio of the beamwidth to the signal separation at the limit of resolution, which is remarkably simple, and independent of the array moment matrix.

(c) A modified expression for the resolution improvement factor when significant errors are present was obtained.

(d) The resolution of two coherent signals (also covering the case of processing a single data frame, or snapshot, as in the radar case) was investigated. It was found that when the coherent signals are in phase quadrature the resolution is the same as for non-coherent signals. When they are in phase or in antiphase only one signal is seen.

5. Other topics

(a) Clear and accessible descriptions of the IMP and MUSIC superresolution methods are given.

(b) The CRB for multiple signals and multiple parameters is derived from basic definitions and special cases derived.

Acknowledgements

I would like to acknowledge gratefully the encouragement and help of Professor Hugh Griffiths, from the initial decision to embark on this work to finally seeing it reach a conclusion.

I would also like to thank my friends and colleagues at Roke Manor Research, in particular Chris Tarran and David Salter, for their support. Finally thanks are due to my wife Pat for her patience and understanding during this rather unsocial period of study.

Contents

List of Figures

List of Tables

Chapter 1

Chapter 3

Acronyms and Abbreviations

Superresolution methods:

ASPECT – **A**daptive **S**ignal **P**arameter **E**stimation and **C**lassification **T**echnique

DOSE – **D**irection **O**f arrival by **S**ignal **E**limination

IMP - **I**ncremental **M**ulti-**P**arameter

MODE – **M**ethod **O**f **D**irection **E**stimation

MUSIC – **Mu**ltiple **Si**gnal **C**lassification

Other abbreviations:

CML – conditional (or deterministic) ML

CRB – Cramér-Rao (lower) Bound

DF – direction finding

EPP – equal, parallel patterns

i.i.d. – independent, identically distributed

ML – maximum likelihood

PSV – point source (response) vector

s.d. – standard deviation

SNR (and S/N) – signal to noise ratio

iSNR – integrated signal to noise ratio

ULA – uniform (or regular) linear array

UML – unconditional (or stochastic) ML

Chapter 1: Introduction and Aims of the Study

1.1 SUPERRESOLUTION

Superresolution in general is a term applied to parameter estimation methods which aim to achieve better resolution than standard methods. To take a concrete example, a parameter of importance in the radar case is the direction of each target. Point targets are found by scanning a beam in azimuth (taking just the single dimensional case) and if two targets are present, sufficiently large and more than a beamwidth apart, then two beam patterns will be seen as the radar antenna beam scans past the targets. The target directions are taken to be the azimuth pointing positions of the beam at the peaks of the two responses. If the targets are moved closer together the two responses will eventually merge, forming a single peak, and the targets will be considered not to be resolved. The resolution achievable is generally taken to be given by the Rayleigh criterion, that is, that the targets are at the limit of resolution when the peak of the response from one target is at the position of the first null of the response from the other. (This is for the case of equal strength targets. For different target strengths there seems to be no simple definition, but the resolution will be poorer – the minimum separation for resolution will be greater.) For a rectangular, uniformly weighted aperture the beam power response will be a sinc^2 function and the separation will correspond closely to the 4dB beamwidth. Of course if the targets are separated by a smaller amount than this, although there is only a single peak, the response will be broader than the basic beamshape, perhaps with a shoulder on it, and in principle the presence of two targets could be deduced and their angular positions estimated. However to detect and analyse this broadening (which becomes more difficult at low signal to noise levels) will require extra processing, and if this is applied it could be considered a form of superresolution. We can see that if this method were tried we would expect the performance to improve with signal to noise ratio (SNR), as the response shape would be better defined, either in the minimum resolvable separation for given strength signals, or in the increased SNR difference at which signals could be resolved for a given separation.

We consider here superresolution methods (though not actually using the approach above, of analyzing the response shape explicitly) with the ability of improving resolution above that of simple beamforming, which are based on using a sampled, or discrete, aperture followed by digital processing. The advantage of using a sampled aperture, with digitization, is that the data can be recorded and then more complex processing can be applied. Also, for a range of directions of interest, beamforming can

be performed electronically, without having to rotate the antenna physically. We note that, although this discussion so far has been in terms of direction finding, and this is indeed an important application of superresolution, the principle is applicable in other areas, for example in finding the frequencies of sinewaves in noise (an early application of MUSIC, to be described below). In this case the parameter to be found is frequency and the aperture, more generally termed a window, extends in the time domain. Another application, relevant to high performance radar, is for high range resolution (HRR). In this case the wide spectrum required for HRR is discretized in the form of a set of pulses distributed in frequency over the band. Here frequency is the aperture domain and range (effectively delay) is the parameter required. Another radar target parameter which could be found using superresolution is velocity.

An early application of superresolution in direction finding was to sonar, often with long towed arrays, basically linear in form, and looking in only one angle. Application to radio, particularly at HF and VHF, and radar followed, where generally a planar array is used to find signals in a two dimensional angle domain (azimuth and elevation). (The array need not necessarily be planar; it could be a volume (3D) array, including the case of an uneven, approximately planar, array.) In this study we assume that we are interested in the 2D angle case, whether for radio, sonar or radar application.

There was considerable interest in the subject of superresolution in the 1970s and 1980s, when a number of methods were devised and published, dividing into two broad groups, the 'spatial spectrum' methods and maximum likelihood methods. The methods in the first group generally required less computation and were simpler to implement and received most attention. In this group one method tended to become dominant, MUSIC ('**Mu**ltiple **Si**gnal **C**lassification', introduced in 1979, followed by a number of variations on it), being effective, reasonably robust and not too computationally demanding. Although some related methods (Root-MUSIC, ESPRIT, etc.) may be improvements in some respects, MUSIC remains quite a practical and popular system and we take this method for the study here.

The methods of the other group, based on maximum likelihood (ML) parameter estimation, are potentially very demanding computationally and techniques or algorithms for simplifying the search for the set of parameters maximizing the likelihood functions were presented. In fact the only statistics for the additive noise corrupting the data, and so limiting the parameter estimation accuracy, were taken to be normal (Gaussian) in which case the problem becomes a form of least squares

minimization, simplifying the problem to the point where it becomes rather more practical. Even in this case, more work is required than for the spatial spectrum methods. However, the ML method has an important capability that MUSIC does not have. In general MUSIC is not able to separate two coherent signals, while in principle this can be achieved with ML. This is important if superresolution is to be applied to radar target echoes, which are effectively coherent, being copies of the same signal – the radar transmitted pulse. Furthermore MUSIC requires at least as many sets of data samples (snapshots) as there are (non-coherent) signals present, while ML can operate in principle on a single snapshot. (In this case even non-coherent signals are equivalent to coherent.) This means that, in the radar case, superresolution can be applied to the output of a range gate (applied across the set of receiver channels) following a single pulse transmission. Where these potential advantages of ML are not required MUSIC is likely to be preferred for its simpler and perhaps more robust implementation. The ML (or 'Approximate Maximum Likelihood') method taken for study here is known as IMP. Both MUSIC and IMP are described in Chapter 2 below.

Papers on the accuracy and resolution of superresolution followed, mainly in the 1980s and 1990s. These were very largely based on MUSIC and related methods and almost entirely considered only the single angle parameter case, using a linear array. There seemed to be little on the subject of the two-dimensional angle case (source azimuth and elevation) and it seemed this should be investigated for the benefit of users of systems designed to give this information. There was also rather little on the resolution of ML methods, perhaps because these were not seen as being applied in practical systems to any significant degree, and this also seemed to be a possible gap to be covered.

1.2 BRIEF REVIEW OF PUBLISHED WORK ON THE PERFORMANCE OF SUPERRESOLUTION METHODS.

1.2.1 MUSIC

Table 1.1 lists some of the papers published in connection with theoretical studies of the performance of MUSIC. (The papers are labelled here by the initials of the authors' surnames and the last two digits of the year of appearance, to emphasize the chronological sequence. The square bracket numbers are the numbers of the papers in the overall reference list at the end of the thesis.) There are basically two aspects of performance of interest – the accuracy of the parameter estimates, given as a variance, or standard deviation (the methods generally give zero bias), and the ability of the

method to resolve two close targets. The focus of the paper on one or the other, or both, of these is given in the second column. In the third column is shown whether the study is of more than one parameter per target. In this table we see that every paper but one listed considers the case of only a single parameter (azimuth angle).

It is important to consider the potential performance available from a system, limited only by the presence of additive noise, with a finite number of data samples per channel (or 'snapshots'), and many papers explored this. However in practice there will be some degree of error in the system due to, for example, unknown phase and amplitude variations in the receiver channels, uncertainties in the exact positions of the sensors, variations in the sensor responses, and so on. In some cases, particularly with weak signals (low signal to noise ratio (SNR)) and low integration (a small number of samples) noise will dominate errors in determining performance, and the effect of errors may be neglected. Thus it is also important to determine the performance as limited by errors, and whether the performance in the presence of errors (or errors plus noise) is the subject of a paper is indicated in column 4. In other cases, particularly with high integration, the effect of noise can be made negligible, and the performance can be considered limited only by the errors.

Paper	Acc/ Res	2 par	Errors	Notes
JF85 [17]	(R)	×	×	Mean MUSIC function expression
KB86 [21]	R	×	×	Expression for resolution threshold
FJM 88 [13]	R	×	×	Mean MUSIC function expression, 2nd order approx. (Low integ'n case.)
ZHJ89 [55]	R	×	×	Alternative to KB86; similar result for resolution threshold
SN89 [42]	A (R)	×	×	Compares MUSIC (and ML) with CRB. Very influential paper.
F90 [15]	A (R)	×	✓	Accuracy and resolution. Very good paper.
LW90 [24]	R	×	×	Based on KB86. Shows beamspace operation gives better resolution.
LW91 [25]	R	×	×	Very good results for probability of resolution, but expressions complex.
SK92 [48]	A	×	✓	More types of error, including multipath.
WF94 [52]	R	×	✓	Excellent paper; with simple expressions for resolution of ULA and UCA.
KSS94 [20]	A	×	✓	Variance in terms of eigencomponents. Compares MUSIC and ESPRIT.
SNLZ96-1 [45]	R	×	✓	Expressions complex. Only ULA considered.
HF95 [19]	A	✓	×	Polynomial rooting algorithm (PRIME) using regular 2D arrays
Z95 [54]	R	×	×	Good, but expressions complex (no errors)
SNLZ96-2 [46]	A (R)	×	✓	Errors due to filter mismatch taken.
SGNLZ00 [47]	A (R)	×	✓	Very complex expressions (giving good results).
ND04 [33]	R	×	×	Planar arrays but elements close. Wavefield modelling – rather specialized.

Table 1.1 Papers on MUSIC

Early work by Seidman [40] looked at the CRB (and also the Ziv-Zakai bound, not considered in any later papers referenced here) for the case of simple beamforming,

though only for one parameter and one signal, using a linear array. His CRB expression (eq. (12)) shows this bound as (inversely) dependent explicitly on the array second moment of position, a result not clearly shown by subsequent authors, in general, but confirmed and made more general in Chapter 3, below. (Note: although simple beamforming is not a superresolution method, in the case of a single signal both MUSIC and IMP reduce to this case, and it may be of some interest to know how accurate these methods are when only a single signal happens to be present.)

Jefferies and Farrier in 1985 (JF85, [17]) began to tackle the question of the resolution of MUSIC by obtaining expressions for the mean value of the MUSIC function. This function is scanned over the angular region of interest (in the DF case; this may be the whole azimuth range of 360º) and searched for minima (ideally zeros) at the positions of signal directions, or for maxima (peaks), in the case where the reciprocal of the function in plotted. Searching for zeros is more satisfactory in the computation algorithm, but terminology is often in terms of peaks, as this is a more attractive form for display. Using the peak terminology, we note that when signals are well separated (i.e. some beamwidths apart) we obtain distinct peaks but as two signals are brought close together the space between the peaks tends to fill in and eventually the two peaks merge into one, and clearly the signals are not resolved. Obtaining a mean form for the MUSIC function showed at what separation the sources were on the edge of resolution, though no analytic form was found for this separation, in terms of system and model parameters (such as the array structure or signal to noise ratios).

Kaveh and Barabell in 1986 (KB86, [21]), however, did produce an expression for this threshold, in a very good paper. However, as for all the papers of this time it was for the case of a single parameter – azimuth direction – and for a uniform (equispaced) linear array (ULA). This model was widespread, perhaps because of early work related to sonar, using long towed arrays. Some papers (e.g. SK92, [48]) refer to ‘a towed array’ and one (KSS94, [20]) even took for its error model in simulation a basically linear array with sinusoidal transverse position errors. Other papers (e.g. BM86, [8]) refer to using ML not for DF but for extracting ‘sinusoids in noise’ from long time series data - presumably another sonar application.

Farrier, Jefferies and Mardani (FJM88, [13]) followed up the 1985 paper with a more accurate (second order) approximation to refine their mean MUSIC function expression, and in principle this could be used to find the separation of two targets at the limit of resolution for a specific case, but no theoretical expression for this limit was given. The

second order approach gave an improved result for the mean MUSIC function in the case of small sample numbers.

Zhou, Haber and Jaggard (ZHJ89, [55]) also obtained an expression for the resolution threshold based on equating the function values at the target positions with the value midway between these positions.

In 1989 Stoica and Nehorai published a very influential paper (SN89, [42]) on the parameter measurement accuracy as given in theoretical and algorithm-independent form by the CRB, as well as obtaining expressions for the accuracy of MUSIC and Maximum Likelihood methods. They obtained an expression for the Cramér-Rao Bound from the rather theoretical definition, stating that although some CRB expressions had been obtained for specific cases 'there does not appear to be available in the literature' an expression for the CRB in general. Although this takes the generalization a considerable way, it is still only for a single parameter per target, although it is for multiple targets. They also obtained expressions for the accuracy of (single) parameter estimation of MUSIC and ML methods, and showed that asymptotically (for large sample numbers and for arrays with large numbers of elements) the performances of these methods approach the CRB, and gave other significant results. More precisely, they conclude that the CRB gives the approximate limit on the accuracy of MUSIC for uncorrelated signals (very often taken to be the case). In the ML case, they show that the estimator is 'inefficient' (i.e. does not meet the CRB) for a finite number of elements, but this is for the 'deterministic' (or 'conditional') ML case, where all the signal complex amplitude samples are included as parameters. In the 'stochastic' or 'unconditional' ML case, where only the signal statistics (in particular the covariances of the signals) are considered as parameters, they 'conjecture' that the MLE (maximum likelihood estimator) is statistically efficient for large sample numbers.

This (Stoica and Nehorai's) seems to be the first substantial paper on the accuracy of MUSIC in the absence of errors. The first substantial paper on the performance of MUSIC with errors was that of Friedlander in 1990 (F90, [15]), again for a single parameter. However, planar arrays were also included as well as the usual linear arrays, but in this case the signal sources are taken to be in the plane of the array, and there is no consideration of a second angle parameter. In order to investigate the effect of errors, rather than noise, on the performance, he takes the case of infinite integration, so forming the actual system covariance matrix, rather than the estimated covariance

matrix (required by MUSIC for its eigencomponents) resulting from finite integration, and taken as the starting point for previous analyses. (In practice sufficient integration could always be taken, in principle, to ensure that errors, rather than noise, dominated the performance.) This condition is also taken in Chapter 3 below. A (multiple source) error sensitivity, relating the parameter estimation error variances to the input system errors variances is defined, an idea also used in Chapter 3 (and not apparently in other papers of those listed here). A weakness, from the point of view of easy application of the results, is that the expressions are complicated in form, and not simply related to the array specification.

Lee and Wengrovitz (LW90 and LW91, [24, 25]) returned to the study of resolution of MUSIC without errors, and in the 1990 paper showed that resolution is improved using beamspace MUSIC – i.e. forming a set of beams from the outputs of the array elements, and then applying MUSIC. This is not a surprising result as we now have, in effect, a set of high gain (or high directivity) elements but there may be problems of spatial coverage, so this is not necessarily a preferred approach. For simulation a planar array was used (crossed linear arrays with logarithmic spacing, rather than the usual uniform spacing) but again the source was in the plane of the array. (The reference point for the element positions was explicitly the centroid, as is the case for the general arrays taken in Chapter 3). The 1991 paper presented results on resolution giving very good agreement with simulation, and including different signal power levels, but again the expressions are very demanding to evaluate and there is no clear relationship to the array form. Swindlehurst and Kailath (SK92, [48]) included, in principle, more sources of error including mutual coupling. Good results are obtained but only for the single parameter case. (The authors remark that the results are ‘easily extended’ to the multiple parameter case.)

Weiss and Friedlander (WF94, [52]) took a different approach from Friedlander in 1990, and produced remarkably simple expressions for the resolution of uniform arrays, both linear and circular (again with the signal sources in the plane of the array). Again this was for single parameter systems but the authors stated there was the capability for multiple parameter estimation. The results were limited to the case of uncorrelated and equal power signals, but these are commonly taken conditions. Preprocessing (i.e. beam space processing) was also considered, and shown, as in LW90, to improve the resolution threshold.

The main aim of Kangas, Stoica and Söderström (KSS94, [20]) was to compare the

accuracy of MUSIC with a related method, ESPRIT, with errors present. They showed that generally MUSIC was better than ESPRIT. The expressions for accuracy (as in several other papers) were in terms of the eigencomponents (eigenvectors and eigenvalues) of the covariance matrix, which may not be a convenient form for making a reasonable estimate of performance.

Zhang (Z95, [54]) returned to the study of resolution in the error free case, obtaining excellent results (comparing with simulation) but again at the cost of very complex expressions.

Hatke and Forsythe (HF95, [19]) are the only authors here to take the case of two angle parameters. However they require a regular planar array (i.e. elements on a regular grid) and use a rooting method – an extension of the RootMUSIC method for linear arrays (and one angle) so is not standard MUSIC. An expression for the 2D spectral MUSIC accuracy is given but this is rather complex and requires the system eigenvalues. Errors are not considered.

Su *et al* (SNLZ96-1, [45]) analysed resolution further in the presence of errors, producing very complex expressions, limited to the case of the uniform linear array (and, as in all these cases, a single angle parameter). The same authors (SNLZ96-2, [46]) looked at the case of errors (or channel mismatches) over the receiver bandwith, taking an approach based on the frequency spectrum. This may have some relevance in the case of signals with a relatively high bandwidth, but is a rather special case, as in most applications such mismatches will be very small compared with other sources of error. A third paper of Su *et al* (SGLNZ00, [47]) again looked at accuracy (and resolution), and again good results (comparing with simulations) were achieved at the cost of complex expressions.

In 2004 Nemirovsky and Doron (ND04, [33]) looked at MUSIC and a related method, Root-MUSIC, in a very special case, using arbitrary planar arrays and analysing by a technique termed wavefield modelling. However their example array placed elements very close together (many separated by as little as 0.15 wavelengths) which would be quite unrealistic (in generating strong mutual coupling) in most arrays, and it seems that a larger aperture array (their 16-element array was contained within a circle of one wavelength radius) would rapidly require more computation.

In conclusion we see that the questions of both the resolution and accuracy obtainable using MUSIC, with and without realistic system errors, have been answered fairly

comprehensively but *only for the case of a single angle parameter* (apart from HF95 which is not for standard MUSIC). In most cases the arrays taken for simulations confirming the theoretical results were the simple uniform linear array, and even when planar arrays were used the signals were taken in the plane of the array. After some relatively early papers (particularly KB86,[21], F90,[15] and WF94,[52]) many others gave more refined, but also generally rather complex expressions, which only marginally advanced the knowledge of the problems. Later papers covered various special cases, which is not the aim in this thesis, which is to concentrate on essentially common, practical systems. From this point of view it appears that the case of direction finding with two angle parameters, using a general array (two-dimensional (planar) or three dimensional) is not yet covered. Thus the accuracy of MUSIC in the two dimensional angle case is the subject of Chapter 3, below. One assumption that seems widespread is that the array consists of elements which all have the same response (apart from phase shift due to position) to each signal. This is reasonable, in the sense that practical arrays may be designed in general to consist of similar, essentially identical, elements. We keep this assumption, in general, in the analysis of Chapter 3, except for one circular array consisting of outward looking directional elements. The analysis there shows how to handle the case of arrays with elements which do not all have the same gain to each signal.

1.2.2 Maximum Likelihood

Fewer papers have been published on Maximum Likelihood methods for DF than on MUSIC and variations on it. A list is given in Table 1.2, using the same column format as for MUSIC. Bresler and Macovski's early paper (BM86, [8]) applies ML to the problem of sinusoids in noise (a sonar type of problem) which has much in common with single parameter DF using a ULA. The paper is concerned with presenting a method, rather than theory on accuracy or resolution. The relatively early paper by Jaffner (J88, [16]) was important in showing that the ML function can be arranged in a form separating the signal waveform parameters and the angle parameters; so for DF only the part involving the angle parameters is required, thus reducing considerably the difficulty of finding the maximum of the likelihood function. Subsequent ML studies (for DF) in general begin with this first step (or with the resulting 'reduced' likelihood function obtained). Jaffner only goes as far as using this for a ML-based DF method, and no theory of accuracy or resolution is given. However, simulations (using a uniform linear array and two uncorrelated signals) were presented, showing ML beating MUSIC

in estimation accuracy (in the single angle parameter) at lower SNR levels and matching the CRB well except at quite low SNRs. (It is pointed out that MUSIC tended to fail at lower SNRs because two peaks were not observed – i.e. it had failed in resolution, so the better resolution of ML allowed the better accuracy in the low SNR case.)

Stoica and Nehorai (SN90, [43]) followed their earlier paper on the accuracy of MUSIC and ML and the limit given by the CRB with a comparison of the accuracy attainable by unconditional and conditional ML (as well as a method called MODE). The conclusion is that CML (which includes methods such as Alternating Projection and the closely related IMP) is shown to be statistically less efficient than UML. However, the performances are very similar for weakly correlated sources, but are more significant for closely spaced sources at high correlation, except for high SNR and a large number of elements. (These particular conclusions are drawn with the help of figures based on the theoretical expressions evaluated for a ULA and considering only a single angle parameter.) Although the UML method seems better in this analysis, it requires high sample numbers (to give a good value for the signal covariances) and this may not be feasible, particularly in the radar case. The conclusions are the same in the longer and fuller book chapter written by the same authors, plus Ottersten and Viberg (OVSN93, [35]) for the case of large sample numbers.

Paper	Acc/Res	2 par	Errors	Notes
BM86[8]	-	×	×	Good description of ML. Proposed method uses ULA.
J88 [16]	-	×	×	ML method using ULA.
SN90[43]	A	×	×	No simulations to check theory.
OVSN93[35]	A	×	×	ML theory, some derivation of accuracy
VON95 [51]	A	×	×	Coherent signals study.
SK96 [38]	R	×	×	For single param. and ULA. Simplified CRB derived
CM97 [10]	A	×	×	*Known* signals with multipath.
MRL97[28]	(A R)	(✓)	(✓)	ML methods (incl. IMP) with coherent signals compared.
SMCR99[41]	R	(✓)	×	Limit of res'n considered in terms of target detectability.

Table 1.2 Papers on Maximum Likelihood methods

The comparison of the methods in the case of a small number of samples (but a large number of elements) leads to a different conclusion in the paper by three of these, Viberg, Ottersten and Nehorai (VON95, [51]). This states that for a large number of elements, but not necessarily a large number of samples, the accuracy is the same for both ML methods, deterministic (or conditional) and stochastic (or unconditional), and the asymptotic (large element number) error variances coincide with the deterministic CRB. A simulation example using a ULA of varying number of elements and two coherent signals showed accuracies close to the CRB for arrays as small as ten

elements.

Satish and Kashyap (SK96, [38]) introduce an original method for DF by maximum likelihood, though this is for the stochastic, or unconditional, case (i.e. where the signal parameters are source waveform statistical parameters only (in particular variance), rather than the waveform samples. This is not the IMP or Alternating Projection case). They produce some simple expressions for the measurement accuracy (expressed as usual as CRB covariance matrices) in the case of two signals (which may be partly correlated) using a ULA, and in one angle dimension. They also introduce an original criterion for the resolution, but this may only apply for the stochastic ML model (so not for a very small sample number) and for a moderately large number of sensors.

Cederval and Moses (CM97, [10]) consider the case of DF in the case of multipath, and also assume the signals are known to some degree, using large sample numbers. The known signal case can include, for example, mobile telephone signals with known preamble sequences. However we need, as taken here, relatively large sample numbers in order to match the signals.

Manikas, Ratnarajah and Lee (MRL95, [28]) published the only paper in this group considering the case of two-dimensional target directions. The paper is concerned with comparing in simulation a number of methods implementing the ML principle. These methods are outlined and compared in performance, in simulation, using a non-uniform circular array of 6 elements. This is a rather more general case than considered in the other papers studied here. However, although the algorithm descriptions implied the two-dimensional capability of the planar array, looking for targets in both azimuth and elevation, above the plane of the array, the simulations were all for targets at zero elevation – i.e. in the plane of the array, as in some of the earlier papers with planar arrays. All the simulations were for two equal power coherent sources, and generally showed that a method named ASPECT (and a variation on it A-AML) performed best in terms of accuracy, with Wax's ML (Alternating Projection) next and IMP considerably inferior. However, it is possible that the implementation of IMP may not have been optimum, in particular their equation (28) for the IMP function is incorrect. (The actual IMP function is much closer to their equation (15) for another method, DOSE, which performed better.) If this was used then this could account for the performance discrepancy. The simulations also included the case of element position errors, and all the methods performed similarly at a substantial source separation. Although the detection/resolution and accuracy lower bounds are quoted (based on earlier Manikas *et*

al papers, using differential geometry of the array, and so algorithm independent) there is no discussion of how the simulation results compare with these theoretical values. This paper does not consider the theory of the accuracy or resolution of ML.

A different and original approach to the question of resolution of superresolution methods, particularly of the ML type, was taken by Speirs, Mather, Clarke and Rees in 1999 (SMCR99, [41]). They consider that if a signal is detected close to another signal then the two signals are resolved, so the question of resolution is closely connected to the question of detection (as implied also by the approach of Manikas *et al*). (However, as Manikas *et al* [27] point out it is also possible to detect that there are two, or more, signals present without being able to resolve them – MUSIC providing a clear example). ML methods of the IMP (or Alternating Projection) type place array gain nulls on one signal in order to remove this signal and allow observation (and detection) of the other signal. The null on the first signal causes loss of gain to the second signal, and so the detectability of the signals, and hence their resolution, depends on their (absolute) signal to noise ratios. This seems a sound and attractive basis for defining the resolution threshold, and is the only paper here dealing with the resolution of ML systems, except for that of Satish and Kashyap, which is for the large sample, stochastic (unconditional) ML case, which does not include IMP, and also excepting the differential geometry approach of Manikas *et al*.

We conclude that the question of the accuracy of ML methods, in the absence of errors, has been covered fairly extensively, though only for the single parameter case (except for Manikas *et al*) and usually with linear arrays. The performance with errors has not been analysed in any of these papers (though simulated in one). However the question of resolution of ML systems has not been covered to the same degree, and the resolution of IMP is the subject of Chapter 4 below.

1.2.3 Cramér-Rao Bound and algorithm independent studies

Papers on the Cramér-Rao Bound are listed in Table 1.3. This fundamental bound is important in this work as it is shown in various studies that superresolution system accuracies (in the absence of errors, of course) approach this bound, at least asymptotically (i.e. for large samples or large numbers of array elements) in some cases. Thus we can use this bound as a good indication of potential performance even though actual systems may not quite reach it, even without significant errors. In their influential 1989 paper Stoica and Nehorai (SN89, [42]) went a considerable way in presenting a general form for this bound, stating that this did not seem to be available in the literature

at the time. However this was for the case of a single parameter (per source, for multiple sources) and Yau and Bresler (YB92, [53]) gave a form for the multiple parameter case, extending the work of Stoica and Nehorai. Lee and Jachner (LJ94, [23]) extended the bound further, to include the case where the array responses to the signals (one applicable term being the point source vectors, PSVs) differ in form from one element to another. This is in distinction to the generally taken case where the responses are identical functions of the angle parameters. This more usual case corresponds to that defined here as EPP – equal, parallel patterns – all the element responses have the same shape and orientation (the 'parallel pattern' case) and equal gains. Lee and Jachner's formulation drops this assumption, so that the element patterns may differ. Their motivation, in particular, was to take the case of polarization diverse arrays; a subject of some interest, but not the form for many practical arrays which would have similar elements with the same polarization.

Paper	Acc/ Res	2 par	Notes
SN89 [42]	A (R)	×	Derives CRB (1 parameter). Compares MUSIC and ML accuracy with CRB
YB92 [53]	A	✓	Multiple parameter CRB; similar derivation to mine.
LJ94 [23]	A	✓	CRB with multiple parameters
KM96 [26]	A R	✓	CRB derived using differential geometry of planar array. Two angle parameters
MAK97 [27]	A R	✓	Accuracy, resolution and detection using differential geometry. 8 arrays compared

Table 1.3 Papers on the Cramér-Rao Bound

Karimi and Manikas (1996, [26]) took a very original approach to array performance, analysing array configurations using differential geometry applied to a planar array (not generally regular, but taken as EPP) used for signal angle determination in two dimensions (azimuth and elevation). Starting with the basic CRB expression they use differential geometry to obtain more convenient and quite simple expressions for the variance of the errors in azimuth and elevation, for both the single signal and two signal cases (though these expressions might be derivable without this approach). Curves of these theoretical variances are given. The later paper, by Manikas, Alexiou and Karimi [27], takes the differential geometry approach rather further, defining conditions for the limit of detection and the limit of resolution, leading to expressions for corresponding thresholds, as well as a bound on the angle estimate variances at the limit of resolution. They also extend their results to a more general planar array, dropping the EPP condition and allowing the element gains to differ in any given direction. They then evaluate these expressions for a number of planar arrays to compare their performances (in theory). The ideas here are very interesting but there are no illustrating simulation

results (which would be of particular interest in detection and resolution, in particular) or comparisons with other resolution (if not detection) expressions.

1.3 AIM OF THESIS

In general it is important for theory to determine the limits of performance as well as to show how the performance depends on various factors and parameters of the data gathering and processing system. Without any such information, it may not be clear whether the performance of a practical system was anywhere near to the limit possible, or if much improvement might be gained by better design and implementation. However, to obtain expressions of considerable complexity, with the aim of defining very precisely the performance limits, is not necessarily helpful to the general user. A general aim was to try to find relatively simple expressions, by making modest approximations if necessary, which would retain the parameter dependence factors, but would be rather easier to use to estimate the performance that might be expected from a given system. By 'system' in the direction finding case, we mean, to a major extent, the structure and layout of the sensor array used, and it would be helpful to express the performance measures relatively simply in terms of the array description. A possible outcome would be a 'rule of thumb', which would give a reasonable indication of the performance to be expected, in some respect, with very little calculation, rather than precise figures for specific cases, obtained with considerable computation. This result would not be expected to be highly accurate but to be close enough for practical use, bearing in mind that often the available data regarding the situation in which the equipment is to be used may be to some extent vague, and hence it would not be appropriate to use a precise and complex formula.

More specifically, it seemed that the area of two dimensional direction finding (e.g. in azimuth and elevation) had not been widely covered. This requires arrays of two dimensions (planar) at least, and three could be used, so it would be worth covering this general case (with little extra complexity in the expressions, as it happens). Thus the accuracy of MUSIC in the two parameter case, with system errors, was taken for the major topic of this work. (The case without errors is essentially covered by the CRB, for which 2-parameter expressions had been obtained.) The resolution of Maximum Likelihood methods (without errors) had been tackled very little, and it seemed that an estimate of the resolution performance should be available in terms of the array description.

1.4 OUTLINE OF THESIS

After reviewing in this chapter the published work and identifying areas where further work could be of value, in the second chapter we present the background to the work in Chapters 3 and 4. We discuss the principles of superresolution in general terms, not limiting the application to direction finding at this stage, so making clear that the principle is applicable more widely. This leads to an outline of a range of proposed superresolution methods, which are considered briefly, for which some results in connection with adaptive array theory are required, provided in an appendix (Appendix 2B). There are a considerable number of appendices throughout the thesis, which contain work leading to results used in the main text, but whose working would interrupt the argument. These appendices are a significant part of the thesis, with results not always readily accessible elsewhere.

Following the outline of methods, with particular interest in two of them, MUSIC and IMP, these are then chosen and described in detail. For the Maximum Likelihood based method, IMP, the likelihood function is derived from a basic level, and this is simplified to give a condensed likelihood function, by maximizing over the noise power level and the signal waveform samples, which are not the prime parameters of interest. The IMP function is then derived from the condensed ML function.

Although there is no specific discussion of the Cramér-Rao bound, as this function is of considerable significance in this work it is derived in an appendix for the case of multiple parameters per source (not as well known as the simpler single parameter case) to give a rather general expression. For application at other points in the report a number of simpler and more easily applied forms are obtained for special cases.

In Chapter 3 the topic of the accuracy of MUSIC, in particular for the 2-dimensional direction finding application, as affected by system gain and amplitude errors is taken. The subject builds up from the single signal, single parameter case using a linear array to the multiple signal, two parameter case using a general array, general except that the array elements are all taken to have the same gain in any given direction. We term this the EPP case – equal, parallel pattern elements, i.e. the element gain patterns may be shaped, but they must be similar, oriented in the same direction and equally scaled. This is widely assumed to be the case in most (possibly all) the papers reviewed, though is not generally specified, and, as well as being more feasible for analysis, it is a very common case in practice. (At least, in many cases, the array is generally intended to be of EPP form). However, in one case in Chapter 3 this assumption is dropped and a

(slightly more complex) result is obtained. Although the initial case of a linear array (and necessarily a single angle parameter) has been widely analysed previously, the approach used here is the same as, or is consistent with, the later, more complex cases, and forms a good introduction, and establishes the notation required and used. The approach taken here differs from many in the prior literature, in particular in not requiring any eigenanalysis, with complex expressions using eigenvectors and eigenvalues.

Perhaps unusually for a study of a superresolution system, we include the analysis for the case of a single signal. In this case the superresolution system (whether MUSIC or IMP) is equivalent to a simple beamforming system. This gives a theoretical result (confirmed in simulation) not apparently reported previously, that the accuracy is unaffected by amplitude errors (up to moderately high levels, unlikely in practice, in a well engineered system) and is only sensitive to phase errors. When two signals are present the accuracy is affected by both forms of error. The analyses show that an important quantity in the expressions (for both accuracy and resolution, shown in Chapter 4) is the array moment matrix. This is the matrix of second moments of the array elements about the mean position (the array centroid with uniform weighting) – or it can be considered as a covariance matrix of the array positions. (For the non-EPP case we still need this moment matrix, but we also need the matrix of second moments about the mean with element weighting given by the amplitude power gains of the elements.) This moment matrix is easily formed (as the 3×3 matrix $\mathbf{R}\mathbf{R}^{\mathrm{T}}/n$), given the 3×$n$ matrix $\mathbf{R}$ of the n element positions in three dimensions. (If the origin for the element positions given in $\mathbf{R}$ is not the centroid, then we form the centroid position vector $\bar{\mathbf{r}} = \sum_k \mathbf{r}_k \big/ n$, where $\mathbf{r}_k$ is column k of $\mathbf{R}$, and subtract this from all the columns of $\mathbf{R}$ before forming the moment matrix.) Although some authors effectively obtained the second moment in the linear array case (a single scalar value for this 1D case, e.g. Seidman's CRB expression [40], eq. (12) or (13)), it was not generally recognized as such. For 2D and 3D arrays this moment matrix is more clearly important.

The CRB expressions (in which case the accuracy is limited by noise, rather than errors) are compared with the error based expressions derived for MUSIC and the latter are seen to be given by the CRB expressions with the error variances, correctly scaled, replacing the noise variance in the CRB. This is an interesting and useful result, perhaps not altogether surprising, but it is important to have this shown analytically. A side effect of this is that the CRB is expressed in terms of the moment matrix, again not

apparently a form seen previously.

The accuracy of IMP is the same as MUSIC (and simple beamforming) in the single signal case, but for two signals the more complicated function makes the analysis too complex. However, simulations showed that the performance of IMP was apparently also given by the theoretical expressions derived for MUSIC in the two signal case, which corresponds with both being close to the CRB expressions.

The resolution of IMP (or ML systems in general) is seen to have been given very little attention in the published work, and is the subject of Chapter 4. Only one of the papers (SMCR99, or [41]) reviewed tackled this subject, though the approach was promising, and formed a starting point for the resolution study, in the case of non-coherent signals. The paper did not relate the resolution performance to the array structure or parameters, which is done in this chapter. Also the method is refined to give a more accurate resolution estimate, in terms of the dependence on the strengths of the two signals (in the high signal, error-free case). The case of coherent signals was also tackled, following the IMP procedure more closely than for the coherent case. This case is of interest as it corresponds to the single sample case (or multiple samples of coherent signals in an effectively stationary system) and so to the case of IMP applied to radar. Only partial success was achieved here – it is shown that the lowest (second) order approximations used, successfully in the case of the accuracy analysis and the non-coherent resolution case, is not adequate for the case where the complex echoes are in phase or antiphase, and that perhaps a higher order (probably fourth) is needed. However, the approach seems to give a more promising result for the quadrature case.

The appendices in Chapters 2 and 4 are placed at the end of the chapters and are labelled 2A, 2B etc. but in Chapter 3 there are a considerable number of them and they are often specific to the different major sections and so are placed at the end of each section and labelled 3.2A, 3.2B, etc.. The equation numbering includes the chapter number and the major section number. In Chapter 2, for example, the third equation in section 2.2 is labelled 2.2.3, and the third equation in Appendix 2A is (2.A.3). In Appendix 3.2A, in Chapter 3, for example, the fourth equation is (3.2A.4). The figures and tables are numbered in order throughout each chapter.

Chapter 2: Multiple Parameter Estimation by Superresolution - Principles and Examples

2.1 DEFINITIONS AND BASIC PRINCIPLES

2.1.1 Example of parameter estimation; parameter and observation domains

To illustrate some general principles of parameter estimation we consider a specific example (Example 1). We suppose that we require to find the azimuth angle of some distant signal source, and we use a cylindrical parabolic dish antenna to collect and receive this signal. The antenna is scanned over an angular region, possibly the whole azimuth range of 360°, and the output power is monitored. If there is a single strong signal present the response will peak when the antenna points exactly in the signal direction and so the parameter of interest, the azimuth angle of the signal, is obtained.

We note that to estimate the parameter (signal direction, in this example) we use a probe that is sensitive to this parameter, the beam of the antenna. By scanning this probe over a range of the parameter and finding the peak response, the value is estimated. Clearly the more sensitive the probe the more accurately the peak can be found, and in this case this means the narrower the antenna beamwidth should be, in the azimuth plane. In turn, this means the larger the aperture of the parabolic dish required in the horizontal plane. We can define the full range of possible parameter values as the **parameter** or **measurement domain** and the dimension over which the data is collected as the **observation** or **data domain**. Thus the parameter domain for this example is the azimuth direction, in this case finite, extending over 360° or 2π radians, and the observation domain is the one-dimensional horizontal axis, an infinite domain, of which a finite interval, horizontally across the face of the antenna dish, is used to collect the signal. This interval is known as the (horizontal) **aperture** of the antenna, and in general the larger the aperture in the observation domain the more sensitive the response in the measurement domain.

We can estimate the directions of several such signals, in principle, if they are sufficiently separated in angle. As we scan across each one the output power rises to a peak, giving the azimuth position of that signal. Furthermore we can find a second parameter, elevation angle, for each signal, by scanning vertically, as well as horizontally, using a circular paraboloidal antenna. In this case the parameter domain is two dimensional, covering the two angles of azimuth and elevation, and there is, correspondingly, a two dimensional observation domain which can be defined as the

XY-plane, over the face of the antenna, with the direction of maximum response, the peak of the antenna beam, as the Z axis.

2.1.2 Accuracy and resolution

We now consider what may be some of the limitations on the performance of this system. In practice there will always be some noise perturbing the received data, so that the response will be distorted. This means that the peak will not be at exactly the right position, leading to an error in the measurement. Thus the strength of the signal, relative to the noise level, is important – the larger the signal, compared with noise, (i.e. the larger the signal to noise power ratio, the SNR) the smaller the noise distortion and the more accurate the measurement. Secondly, if the beam response, as a function of look direction, is made narrower the errors will be smaller at a given SNR value. This is illustrated diagrammatically in Figure 2.1, where the noise contribution, added to the target response causes the response to rise by δp with an offset of $\delta\alpha$. The value of $\delta\alpha$

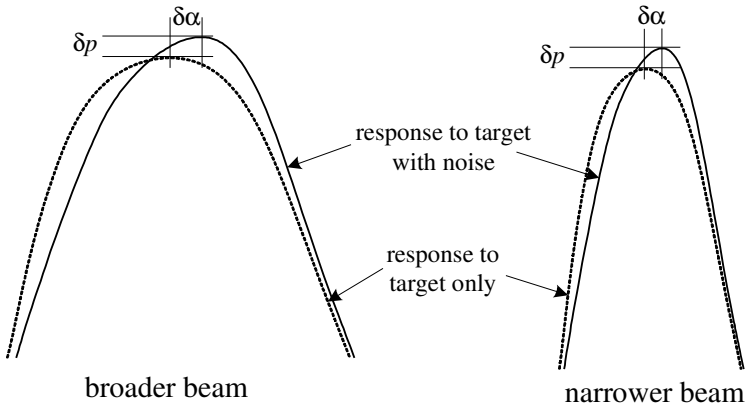


Figure 2.1 Effect of beam width on accuracy

will be smaller for a narrower beam than for a wider one, as shown. The width of the main lobe of the response function is inversely proportional to the antenna aperture, the interval over which the signal is received, so as the aperture is increased and the beamwidth becomes narrower the **accuracy** improves. (In addition, as the aperture is increased the signal energy received increases, so that the accuracy is further improved by this SNR gain).

We also note that if two signals are present and these are brought closer together, at some point the two responses will merge and only one peak will be seen. This means that the signals are no longer seen as distinct – i.e. they are not resolved. Again this situation is improved if the beam is made narrower – the signals can be brought closer before the limit of resolution is reached. We need a convenient definition for

resolution, and one is the **Rayleigh criterion**, which is that two equal power signals are considered to be at the limit of resolution when one is at the position of the first zero of the response centred on the other signal. This is very close to the condition that the two peaks merge into a single peak (at least for the case of a uniformly weighted aperture producing a sinc function beam response) so is a realistic, as well as a precise, definition. However this definition is not so useful when signals of different strengths are present. In this case the response of the weaker signal will disappear into the side of the main beam of the stronger at greater separations than that given by the Rayleigh criterion, or into the sidelobes of the stronger response at much greater separations. If the aperture is weighted, to reduce sidelobe levels, then the main lobe of the response becomes wider, reducing the resolution, but allowing more weak signals to be seen fairly close to a strong one (though outside the main beam response). We see that a definition of resolution is rather more complicated when signals of different power levels are to be considered.

We note that if a weak signal is brought within the main beam of a strong one there will be a distortion of the response, such as a shoulder on the main lobe of the response, which will indicate that there is, or may be, a second signal present. In principle more complicated processing than simple peak detection could be used to attempt to determine the strengths of the signals and their positions. This would be a use of signal processing to improve the resolution above the standard (peak detection) method and in that sense could be considered a form of **superresolution**. (Here 'super' is better interpreted as meaning 'above' rather than 'superb' – the resolution achieved by superresolution methods is above, or better than, that achieved by the standard approach.) In practice superresolution methods generally take various approaches to improving resolution, but a method such as IMP (or a closely related form, Alternating Projection) operate somewhat in this way by identifying the strong response and removing it in order to observe the weak one. IMP is discussed in detail in §2.2.3 below.

There is one other aspect of performance that may be required of the system and that is the capability of **detection** of a number of sources. In the standard case this is essentially the same as the capability of resolution – the number of signals resolved is the number detected. However, in the case of two close signals, whose responses merge, the signals have not been resolved but the extra width of the response detectable can indicate two, or possibly more, signals in this small region. If we consider that there

may be a wide dynamic range of signals present, then there may be uncertainty whether small responses are sidelobe responses to the large signals or due to small signals. Perhaps this is more a problem of detection than resolution though the distinction is not very easy to make. In a superresolution method such as MUSIC, discussed below, it is possible to be clear how many significant signals are present (within the receiver bandwidth used) but not to be able to resolve them all. However, in IMP the resolution method actually detects signals successively so in this method the number of signals resolved is also the number detected.

2.1.3 Further examples; Fourier transform relationship

It is well known that the range resolution of a radar system is proportional to the bandwidth used. In particular, for a simple pulse radar, if the range resolution (the minimum range separation of two point targets) is considered to be equal to the pulse width times $c/2$ (where c is the velocity of light, and also, of course, of the radar pulse), then the narrower the pulse the better the resolution, and the wider the bandwidth required. To improve the resolution, instead of using a narrower pulse (which would have to be correspondingly larger in amplitude in order to contain enough energy for detection) the alternative is to transmit a long pulse, with a relatively low bandwidth envelope, but with some modulation to increase the bandwidth. (The advantage of this method is that it saves having to generate inconveniently high peak power levels for transmission.) The pulse is demodulated on reception to produce the required narrow response and this is the principle of pulse compression (or spread spectrum). One form of modulation is a linear frequency sweep (chirp modulation) and another, to be considered later, is to transmit a sequence of pulses at different frequencies, a discrete form of spectrum spreading. This is Example 2, and in this case the parameter domain is range and the measurement domain is frequency. The bandwidth used is the aperture, in this context.

Another example, Example 3, is the case of finding the frequencies of sinewaves in noise; this was the original application of MUSIC. In the standard form of processing the Fourier transform of the data, giving the spectrum of the data, provides the frequencies, and in this case the parameter domain is frequency, the observation domain is time and the aperture is the time interval over which the waveform is observed for the purpose of the transform. This last example makes explicit the Fourier transform relationship between the data and the parameter estimation function. In fact this relationship is also present in the other examples. In Example 1 the far field beam

pattern of an antenna aperture (using the direction cosine coordinate, see [7, Chapter 7]) is the Fourier transform of the aperture function. In this case, with the antenna looking in one direction the antenna forms the value of the transform of the aperture illumination at a single point in the azimuth domain; scanning over the whole field gives the full transform. In Example 2 the response in the time domain is the Fourier transform of the frequency response of the target. In general, we conclude that, at least for these, and similar, parameter estimation problems the processing is essentially by Fourier transform, or is closely related to it, and the targets are detected as peaks in the transform, the parameter values being given by the positions of the peaks in the transform, or parameter, domain.

2.1.4 Discrete or sampled apertures

In the case of a parabolic dish the Fourier transform is performed passively, the shape of the dish automatically combining the data across the aperture in the correct manner. If this were to be done actively it would be necessary to apply a multiplicative complex exponential phase factor across the whole aperture, and this would be difficult to achieve in practice. However, we know that the continuous aperture can be replaced, without losing information, by a discrete aperture, if sampled at a small enough interval, and this is generally taken to be one half of the wavelength of the centre frequency of operation, if we want to form the equivalent to a continuous interval. In fact it is not necessary to take a single interval in the observation domain as the aperture, and in the discrete case arrays may well be used with the elements having irregular spacing, being distributed over an area, or volume, and with separations greater than a half wavelength. (Here we take the narrowband case, where the fractional bandwidth is small enough to allow effects of the finite bandwidth on the beam pattern to be neglected with negligible error. This narrowband approximation is valid for a wide range of problems, including most microwave radar cases). This sampling is achieved by using an array form of antenna; in this case the signal in each antenna element of the array is brought down to complex baseband and digitized so that the data is captured as a set of (synchronized) time series and all processing is subsequently carried out using digital signal processing. For example, for standard Fourier transform processing, we can now apply the complex exponential factors to the outputs of a linear array, instead of using a parabolic dish. More significantly, with the data in this digital form, from a sampled aperture, we can now apply more complex signal processing to improve the resolution over that given by the basic Fourier transform, and any such methods can be termed superresolution

methods. We assume that the superresolution methods considered here will use a sampled aperture (though not necessarily regularly, or fully, sampled) followed by digital signal processing.

2.1.5 Point source response vector and system manifold

The data on which the parameter measurement is based forms some function over the aperture used in the measurement domain. This function is the sum, or integral, of contributions from all the sources present. If we consider a single unit strength point source then the observed function over the aperture can be termed the point source function. In the case of a discrete aperture this has a finite set of values, for a given point in the parameter domain, and this set can be termed the **point source response vector** (PSV) (or source position vector, SPV, in the Manikas *et al* papers [26-29]) for that parameter value. The aim of superresolution methods in general is not to find the parameter directly and explicitly (in the form, for example, of the equation "α_k = . . ." where α_k is the azimuth angle of source k) but to find the PSV for this signal, and so, in this example, identify α_k as the generator of this PSV. In fact one superresolution method used the name PTMF, or parametric target model fitting, and this could describe superresolution more generally, as the aim is to find the set of PSVs which, suitably weighted, form, or fit, the observed data (the PSV being the model for a single point source). The full set of PSVs, for all points in the parameter domain, is termed the **manifold** of the data collection system. (In the case of an antenna array it is called the array manifold.) If the aperture is sampled at n points (e.g. in the array case, if there are n antenna elements) then the PSVs have n components, which are complex, in general, and if there are m parameters to be found (for example $m = 2$ in the case of a system determining direction in both azimuth and elevation) then the manifold is a hypersurface of m dimensions in an n-dimensional Hilbert space. In principle this forms a continuum but in practice a sampled form is used.

2.1.6 Superresolution methods

2.1.6.1 Spectral Methods

The term superresolution does not refer to a particular method or a group of methods for parameter estimation, but rather to the general principle of achieving improved resolution, compared with what might be termed the basic method (beamforming, or Fourier transform, for example), using a discrete aperture followed by digitization and digital signal processing. Within this broad definition a wide variety of methods have

been devised. However, we can divide these superresolution methods into two general groups, **spectral methods** and **maximum likelihood methods**. The spectral methods are so called because a scalar function of the data is formed over a range of parameter values, equivalent, for example, to the power spectrum over the frequency domain in Example 3 above. Peaks (or in some cases zeros) in the scalar function indicate the positions of sources or targets on the parameter axis. To be more specific we consider a superresolution method based on Ex. 1.

In Ex. 1 we considered a simple power scan, looking for the peak values, and we noted that this function is a form of Fourier transform of the incident field across the aperture. As the domain over which this function is evaluated is azimuth, rather than frequency, it is sometimes called the **spatial spectrum**. In the case of an array antenna, forming a discrete aperture, rather than the continuous aperture of a parabolic dish, the outputs of the array elements, after digitization, are weighted and summed in a process called beamforming. The weights are in fact Fourier transform factors (complex exponentials) and by varying these factors appropriately the main beam of the array is scanned over the azimuth range, without requiring any movement of the array, in contrast to the physical scanning of the parabolic dish required in the case of the continuous aperture. This particular digital signal processing method for azimuth direction finding may be termed **scanned fixed beam** (**SFB**) and gives the basic resolution capability of the array, on which superresolution methods aim to improve.

The cause of the limitation on resolution is the fact that a point source observed over a necessarily finite aperture produces a spread response in the transform – the **point spread response**. The overlapping of these responses as two sources are brought nearer together causes the responses to merge so that only one peak is seen and the sources are not resolved. If we can narrow the point spread response we should be able to improve the resolution performance. One way of achieving this is to make the system into an *adaptive* array, which minimizes the received power, subject to maintaining a given gain in the look direction. When the array is steered accurately onto a source, power is received with the specified maximum array gain, but all other sources are cancelled and so their contributions are largely removed from the response. As the steer direction moves away from the source, the power from this source falls rapidly as cancellation becomes effective for this source as well. This has the effect of narrowing the point spread response. This method, known as **scanned adaptive beam** (**SAB**) [5] is a superresolution method, with a better resolution performance than SFB. This method

was first proposed by Capon [9], and is also known as Capon's method. It was originally called Capon's MLM (Maximum Likelihood Method), but this is now considered to be an erroneous description and the method has little in common with methods now based on the maximum likelihood principle.

Other spectral methods include MEM (Maximum Entropy Method) [3], Kumaresan and Tufts method [22], Pisarenko's method [36], but generally all of these have been eclipsed by a very effective spectral method called MUSIC (*Mu*ltiple *Si*gnal *C*lassification) [39a,b,c] and variations upon it.

2.1.6.2 MUSIC

A more mathematical description of MUSIC is presented later in §2.2.2, but here we give an outline of the method. We first form the estimated system covariance matrix, used also in SFB and SAB, consisting of all the mean co- and cross-products of the element outputs over the set of data frames (or snapshots) available. We can show that if we perform an eigenanalysis of this matrix (strictly the actual covariance matrix, requiring infinite data) the eigenvalues give the array output power level when the corresponding eigenvectors are used as the array weights. We also find that if there is a number m of point sources present and we use an array of n elements, where $n > m$, then there will be $n - m$ eigenvalues which correspond to the level of receiver noise – i.e. there is no signal power present using the corresponding eigenvectors, which are thus sometimes called **noise eigenvectors**. It follows that the array gain, using any one of these vectors, is zero at the positions of all m signals. Thus if we plot the gain pattern we simply identify zeros in this pattern with signal positions, or, allowing for small errors in practice (such as the fact that only an approximate, estimated covariance matrix is available) we look for the m lowest local minima of the power pattern.

We see that, in principle, using only one noise eigenvector should be adequate, and indeed this is the basis of Pisarenko's method, but a more reliable result should be achievable by using all $n - m$ noise eigenvectors in some optimum combination. This in fact is what is done in MUSIC. We note that the gain in a particular direction using some weight vector can be interpreted as the projection of the PSV for that direction onto the weight vector. The MUSIC function at a given point is a generalization of this idea, in that it is (the square magnitude of) the projection of the PSV for this target point into the space spanned by all the $n - m$ noise eigenvectors, termed the **noise space**, (or, more precisely, the noise subspace). If the projection of the PSV for a signal onto any one of the noise eigenvectors is zero then the projection into the full noise space will

also be zero. Thus we scan the manifold PSVs and find the m lowest points, at which the function is most near to zero, and the PSVs at these points are taken to correspond to the m signals. (We could equivalently take the projection orthogonal to the noise space, using normalized PSVs, and look for peaks, ideally of value unity, as vectors orthogonal to all the noise eigenvectors must be signal vectors. This is the form described in §2.2.2 below.) Alternatively the m signal eigenvectors (the principal eigenvectors, with eigenvalues above the noise level) can be used for either projection, instead of the $n - m$ noise eigenvectors, and this is computationally more economical if $m < n/2$. In this case we either project into the signal space (spanned by the signal eigenvectors) and look for maxima, or orthogonally to it and look for minima.

The $n - m$ dimensional Hilbert space spanned by the set of noise eigenvectors is a subspace of the n-dimensional space containing the manifold vectors (which are point source vectors, PSVs) and also the array weight (or steering) vectors. Because of this use of a partition of the vector space MUSIC and related methods are often known as **subspace methods** or signal subspace methods.

2.1.6.3 Maximum Likelihood Methods

In the absence of noise or errors, the data that will be observed in a given system with a given set of source parameters can be calculated exactly (for a finite set of point sources, and without various forms of ambiguity in the manifold), and conversely, in principle, given the data the parameters can be found exactly. In practice, even if system errors are negligible, all data will be perturbed by additive noise, and if the noise statistics are known the probability of observing any set of data, **d**, given the parameters, **p**, can be obtained. This probability density expression $p(\mathbf{d};\mathbf{p})$ is a function of both the data and the parameters. If we look at this *same* function from a different point of view we see that it is also the probability (or *likelihood* $L(\mathbf{p};\mathbf{d})$) that, given the data **d** that has been actually received and observed, the parameters giving rise to it are given by **p**, and we define the **likelihood function** by $L(\mathbf{p};\mathbf{d}) = p(\mathbf{d};\mathbf{p})$. We now deduce that the *most likely* set of parameters that give rise to the observed data are those that maximize the likelihood function, explaining the term **maximum likelihood** as the name for this method.

A full maximum likelihood approach would involve evaluating, at suitably fine resolution, the likelihood function, given the observed data **d**, for all values of the parameters **p**. The total number of parameters to be explored will depend on the number of targets and the number of parameters per target – for example in the direction finding

case there may be two parameters, azimuth and elevation. Other parameters could be target range, velocity or signal polarization (requiring two parameters for its specification). We see immediately that there is a preliminary decision to be made; the number of targets present, or **model order**. This may be known from other sources, or may be estimated from the data in some cases. In the absence of this information, one solution is to take the case of one target only, and find whether this model gives an adequate fit to the data, and if not increase the number, until a good fit is found. One problem here is that eventually one may be introducing 'targets' to provide a fit to the noise present in the data, so a decision needs to be made as to how far to go in matching the model to the data; this would be based on some knowledge of the level of the perturbing noise.

A further complication is that the data modelling includes not just the point response functions of the sources, generally depending upon, and containing, the parameters of interest (direction, range, etc.), but also the signal waveforms. There are two approaches to modelling the waveforms; in the first (termed conditional or deterministic) the signal samples are considered to be fixed for all possible realizations of the perturbing noise values, and are treated as parameters of the system. In the second (termed unconditional or stochastic) the signals are considered to be random variables, with different values in each realization, and in this case the signal waveform parameters in the model are just their variances and covariances (we take the case of zero mean waveforms, in general). Although involving more parameters, the conditional model (CML) is easier to handle and is used for many ML methods, including IMP, the example taken here. In fact the maximum of the ML function with respect to the non-waveform signal parameters is found without having to explore all waveform possibilities – the direction and other parameters can be 'decoupled' from the waveform parameters. The unconditional model (UML) requires relatively large sample numbers, in order to obtain good approximations for the signal variances and covariances (the elements of the signal covariance matrix) but CML allows smaller numbers, particularly in the radar case.

As a comprehensive search for the maximum of the ML function would require a formidable computational effort, more economical methods of finding the peak are required. Several methods have been proposed, for example Alternating Projection (AP)[56], Expectation-Maximization (EM) [12,14,32], PTMF (Parametric Target Model Fitting) [1,2,34], ASPECT [29], WSF (Weighted Subspace Fitting) [49,50], MODE [44] and IMP (Incremental Multi-Parameter) [11,31]. Some of these (ASPECT, WSF and

MODE) use the Newton-Raphson (or Gauss-Newton) approach to maximizing a non-linear function of multiple variables, but this requires first and second derivatives of the PSVs, which may not be readily available. Others (AP, EM, IMP) use a sequence of scans over the parameter domain (which may be multidimensional), similar to the single scan of the spectral methods, though with a different function, to give a practical algorithm. IMP and AP are closely related, both consisting of scans estimating the target parameter sets, one at a time, and then adjusting the estimates to approach the ML peak. The details of the methods differ, and there is a case for preferring the IMP method, which is described below.

2.1.6.4 IMP

Although IMP has been described as a maximum likelihood method, it can be described in a way much more intuitively appealing to readers with a background in radar and radio, and particularly in array antennas. The relationship with the maximum likelihood principle is made in the mathematical description below.

We consider the application of IMP to Ex. 1, the case of determining target azimuth positions, using an antenna array in this case. Given a set of data from an array of n elements we first carry out a simple power scan (more accurately a signal to noise ratio scan), evaluating the power received in each azimuth direction. (The scans are carried out purely by computation, using the manifold vectors for each direction to steer the beam; there is no mechanical scanning). The peak of this response is taken as the first estimate of the position of the largest signal, though in fact this estimate is in error due to sidelobe power from other signals present. Having the approximate PSV for this signal, a second (SNR) scan is carried out (using the same data) with a gain null maintained at this azimuth position. Even though this position is not quite correct, most of the power from this target will now be excluded and the peak of the second scan is taken to give closely the position of the second strongest signal. A process of fine adjustment of these estimates, termed tweaking, is now carried out. A null is placed at the position of the second signal's estimated position and the position of the first signal is refined, now that most of the second signal's sidelobe contribution has been removed. We then put a null at the updated estimate of the first signal and improve the second signal's estimate, and so on, alternately, until the corrections are of negligible magnitude. Having these two estimates, we carry out a third scan, with nulls maintained in both these directions and look for a third signal. If a significant signal is found, another round of tweaking is performed, with two nulls inserted while correcting each

signal position, and cycling this process round the three signals. As nulls are inserted, to reject energy from signals found so far, the level of the power scan will fall, and when all significant signals have been found the scan will be at noise level. To decide whether a signal is present there must be a peak above some detection threshold. Thus we see that IMP is a detection method as well as a parameter estimation method.

Alternating Projection differs in that the number of targets m (the model order) is estimated first (for example by eigenanalysis of the system covariance matrix, or by more complex methods – such as MDL, or using Akaike's criterion). Then m scans are carried out, as for IMP, with nulls at the estimated signal positions, but without tweaking at this stage. When initial estimates have been obtained for all m targets tweaking is carried out as in IMP. Thus there are two differences between this algorithm and IMP. Firstly the model order must be known or determined first, and secondly no tweaking is carried out in the course of finding the initial estimates of target positions. This might not make any significant difference in many cases – as there is no tweaking initially these initial estimates will be in more error than at the IMP stage when the mth target is detected, so there may be more tweaking at this stage. However, in unfavourable cases the fact that no tweaking is carried out initially will mean that the nulls are not so accurately placed and the resultant poorer cancellation could lead to significant errors, and possibly erroneous peak positions, choosing a sidelobe response from a poorly cancelled large signal instead of a peak due to a small signal. A snag about using model order estimators is that they tend to overestimate, leading to imaginary targets attempting to fit the noise in the data to some extent.

2.2 TWO SUPERRESOLUTION METHODS

In this section two superresolution methods are considered in more detail. We take MUSIC as an example of the spectral methods and IMP as an approximate maximum likelihood method, but before describing these methods we also look at the point source vectors and the system manifold in more detail as these concepts are important in the descriptions of the superresolution methods.

2.2.1 Point Source Vectors and Manifold

2.2.1.1 The system manifold

Superresolution methods in general take the model of the signal environment to consist of a (small) finite number m of point sources in the parameter domain. (This domain is q-dimensional if the targets have q parameters each to be estimated.) Rather than find

the parameters directly the general approach is to find the point source vectors of the targets, where these vectors depend directly (but not generally linearly) on the parameters. Thus the approach is to search the full set of PSVs over the parameter domain to find the m vectors that best model the data. The vectors are of size $n \times 1$ (the number of array elements and the size of a single data frame) so these vectors are elements in a Hilbert space (a complex vector space with a norm or distance measure defined on it) of dimension n. If $q = 1$ then the set of vectors for all values of this single parameter will define a line (generally curved) in this space – a one-dimensional subspace. If there are two parameters then we have a surface in this space, and so on. In general the full set of (normalized) PSVs will define a q-dimensional 'hypersurface' in this space and this is known as the system manifold. We assume that $q \leq n$; it is not possible to estimate more parameters than n. As the parameter space is generally a continuum (though it might be finite, at least in some dimensions, as for example, azimuth angle, extending over 2π radians only, or elevation from $-\pi/2$ to $+\pi/2$ at most) there will be an infinity of PSVs in the manifold. Thus in general a finite stored manifold is used, of PSVs at some suitably finely sampled intervals over the parameter domain. This is generally referred to simply as the manifold, rather than the stored sampled manifold.

This manifold (the set of PSVs over the sampled domain) can be calculated, in some cases, given the parameter values and the system configuration, but in other cases the system responses, described by the PSVs, are not simply related to the parameters and not enough information is available to calculate the responses. An example of this could be the direction finding case, using an antenna array. The response of the array to a point source can be calculated for the direct paths from the point to the array elements, but there could well be various multipath contributions which may not be known, and also mutual coupling effects between elements, which may not be negligible if the elements are close enough. In this case it will be necessary to calibrate the array at suitable resolution, or possibly to use a combination of calibration and calculation (to perform interpolation) to characterize the array. It is often desirable to keep the number of manifold sample points down (for reasons of data storage, the cost of calibration, the number of DF function points to be evaluated, and so on), particularly for a multidimensional parameter domain. A suitably low parameter domain sampling rate would be a few samples (perhaps as low as three) per natural response beamwidth, with the aperture used. However, this would be far too coarse for positioning the nulls in

IMP and a capacity to determine PSVs between the stored sampled values, either by calculation, knowing the form of the measurement system, or by interpolation, is generally essential for good operation of IMP.

2.2.1.2 Examples of PSVs

A point source vector is a set of responses over the sampled aperture from a point source in the parameter domain. We are not concerned, in general, with the absolute values of the responses but with the relative responses across the aperture, and it is generally convenient to take the PSVs to be normalized (i.e. such that $\mathbf{v}^H\mathbf{v} = \|\mathbf{v}\|^2 = 1$, for a vector $\mathbf{v}$). (NB The raised suffix H indicates the Hermitian, or complex conjugate transpose of a vector or matrix, and T indicates simple transpose. In general vectors are in bold lower case type, and matrices in bold uppercase. Scalars and function names are in italics.) The responses are complex, in general, and often the magnitudes across the aperture are equal, leaving only phase variations between the responses at different aperture points. For the phases some point in the observation domain is chosen as a reference. This need not be within the aperture, in principle, but it is often convenient to take it as some mid-point or centroid position, or it could be taken at one edge. Only phase differences are physically significant, an overall phase shift makes no difference, so any reference position is acceptable.

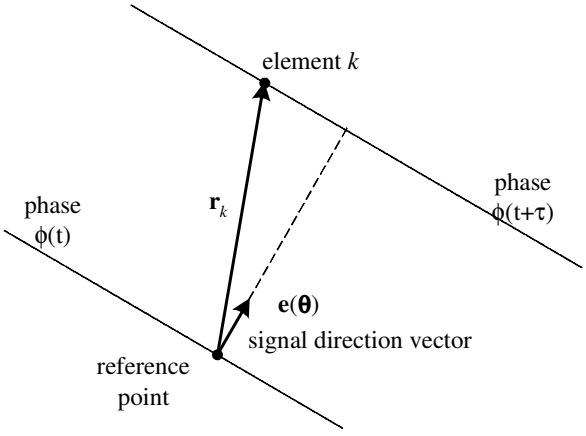


Figure 2.2 Signal phase at element *k*

(a) PSV for an antenna array used for DF

We take the case of a distant source so that the phase front is essentially flat, and from a direction given by two angle coordinates, $\boldsymbol{\theta} = [\alpha\ \varepsilon]^T$. When the phase at the reference point is $\phi(t)$ it is $\phi(t+\tau)$ at element k where τ is the time taken for the wavefront to move from the element to the reference point (Figure 2.2). The delay is given by

$$\tau_k(\boldsymbol{\theta}) = \mathbf{r}_k^T\mathbf{e}(\boldsymbol{\theta})/c. \qquad (2.2.1)$$

where $\mathbf{r}_k^T\mathbf{e}$ (the inner product of $\mathbf{r}_k$ and $\mathbf{e}$) gives the magnitude of $\mathbf{r}_k$, the element position vector, resolved along the vector $\mathbf{e}$, which is the unit vector in the source direction, and c is the velocity of propagation. If f_0 is the frequency of the signal (and

we take the narrowband case, where the bandwidth occupied by the signal is small compared with the carrier frequency) then the phase difference is $2\pi f_0\tau_k = 2\pi f_0 \mathbf{r}_k^{\mathrm{T}}\mathbf{e}(\boldsymbol{\theta})/c = 2\pi \mathbf{r}_k^{\mathrm{T}}\mathbf{e}(\boldsymbol{\theta})/\lambda_0$, where λ_0 is the wavelength of the signal. If all the amplitude responses are equal then the (normalized) PSV in this case is given by

$$\mathbf{v}(\boldsymbol{\theta})^{\mathrm{T}} = \left[e^{2\pi i \mathbf{r}_1^{\mathrm{T}}\mathbf{e}(\boldsymbol{\theta})/\lambda_0} \quad e^{2\pi i \mathbf{r}_2^{\mathrm{T}}\mathbf{e}(\boldsymbol{\theta})/\lambda_0} \quad \ldots \quad e^{2\pi i \mathbf{r}_n^{\mathrm{T}}\mathbf{e}(\boldsymbol{\theta})/\lambda_0} \right] \Big/ \sqrt{n} \,. \qquad (2.2.2)$$

If the elements have individual patterns, given by $\mathbf{a}_k(\boldsymbol{\theta})$ then the PSV is given by

$$\mathbf{v}(\boldsymbol{\theta})^{\mathrm{T}} = \left[a_1(\boldsymbol{\theta}) e^{2\pi i \mathbf{r}_1^{\mathrm{T}}\mathbf{e}(\boldsymbol{\theta})/\lambda_0} \quad a_2(\boldsymbol{\theta}) e^{2\pi i \mathbf{r}_2^{\mathrm{T}}\mathbf{e}(\boldsymbol{\theta})/\lambda_0} \quad \ldots \quad a_n(\boldsymbol{\theta}) e^{2\pi i \mathbf{r}_n^{\mathrm{T}}\mathbf{e}(\boldsymbol{\theta})/\lambda_0} \right] \Big/ \sqrt{n} a_{rms}(\boldsymbol{\theta}) \qquad (2.2.3)$$

where $a_{\mathrm{rms}}(\boldsymbol{\theta})^2 = \dfrac{1}{n}\sum_{k=1}^{n} \left| a_k(\boldsymbol{\theta}) \right|^2$ and hence $\mathbf{v}$ is normalized. If the elements all have patterns with the same shape and orientation, so that $a_1(\boldsymbol{\theta}) = a_2(\boldsymbol{\theta}) = \ \ldots\ = a_n(\boldsymbol{\theta})$ for all values of $\boldsymbol{\theta}$, then the PSVs are all of the form of (2.2.2), even though the patterns vary with source direction. This array can be described as having equal parallel pattern elements. The patterns are not only parallel but equally scaled, so we can describe the array as EPP – having equal, parallel pattern elements.

The form given here is for a general array from which the vectors for special cases, such as planar or linear arrays, and regular arrays can easily be derived. $\mathbf{r}_k$ is a 3-element Cartesian position vector in general but may be given as (x_k, y_k, z_k) or $(r_k \cos\psi_k, r_k \sin\psi_k, z_k)$ using cylindrical polar coordinates, for example, or may be expressed in spherical polar coordinates. The source direction vector is also a 3-element Cartesian vector, but is a unit vector and may be written as $(\cos\alpha_j \cos\varepsilon_j, \sin\alpha_j \cos\varepsilon_j, \sin\varepsilon_j)$ using azimuth and elevation coordinates (α,ε) or as (u_j, v_j, w_j) in direction cosine coordinates ('u-v' form). This is still only a two-dimensional angle of course as w is not independent of u and v, being given by $w = \sqrt{1-(u^2+v^2)}$. In Chapters 3 and 4 we suppose the position vectors are given in wavelength units, so λ_0 disappears in PSVs of the form of (2.2.2) or (2.2.3).

(b) PSV for range measurement

In this case the radar pulse echo delay for a target at range r is $2r/c$ so the phase shift on a carrier of frequency f_0 between transmission and reception is $4\pi f_0 r/c$. For a group of targets within one range gate there is a large common phase shift to the centre of this

range position which is of no interest and can be neglected. Thus we can consider r to represent the difference in range between the target and the centre of the range gate. For a set of pulses at frequencies f_1 to f_n the PSV is given by

$$\mathbf{v}(r)^{\mathrm{T}} = [a(f_1)e^{4\pi i f_1 r/c} \quad a(f_2)e^{4\pi i f_2 r/c} \quad \ldots \quad a(f_n)e^{4\pi i f n r/c}] \Big/ \sqrt{n} a_{rms} \quad (2.2.4)$$

Here we have taken into account the fact that the channel response might be significantly frequency sensitive. If this is not the case then we put $a(f_k) = 1$ (for $k = 1$ to n) and $a_{rms} = 1$, similarly to (2.2.2).

This is for the case of stationary targets. If the targets are all moving with the same radial velocity (for example, being echo points on an aircraft or ship) then a correction needs to be made for this movement – the same rate of change of phase being used as that that gives rise to the Doppler shift (if multiple samples in time are used). If targets are present moving at different radial velocities then we can use a 2-dimensional superresolution process, observing the phase measurements on a set of pulses distributed in both frequency and time. This is a different 2D superresolution example from the two angle DF problem. Here the parameter domain is two-dimensional (range and radial velocity), and the two dimensions of the observation domain are frequency and time.

(c) PSV for frequency estimation

A complex sinusoid, or cisoid, of frequency f sampled at times t_k has phase values given (apart from an additive constant, which is not significant) by $2\pi f t_k$ so its PSV is given by

$$\mathbf{v}(f)^{\mathrm{T}} = [e^{2\pi i f t_1/c} \quad e^{2\pi i f t_2/c} \quad \ldots \quad e^{2\pi i f t_n/c}] \Big/ \sqrt{n}\,, \quad (2.2.5)$$

a very simple form. The observation or data domain is time, sampled at times t_k (for $k =$ 1 to n) and the vector is for a point, f, in the parameter domain of frequency.

2.2.2 The MUSIC method

A single frame of data $\mathbf{y}$ consists of n samples across the aperture in the observation domain. This is related to the m signal samples $\mathbf{x}$ (as received in a reference element at the array origin) by

$$\mathbf{y} = \mathbf{A}(\boldsymbol{\Theta})\mathbf{x} + \mathbf{n} \quad (2.2.6)$$

where

$$\mathbf{A}(\boldsymbol{\Theta}) = \left[\mathbf{a}(\boldsymbol{\theta}_1) \quad \mathbf{a}(\boldsymbol{\theta}_2) \quad . \quad . \quad . \quad \mathbf{a}(\boldsymbol{\theta}_m)\right] \in \mathbb{C}^{n \times m} \quad (2.2.7)$$

contains the m signal PSVs and $\mathbf{n}$ is the set of n noise samples present with the signal data. $\boldsymbol{\Theta}$ is the full set of qm parameters and $\boldsymbol{\theta}_j$ (column j of $\boldsymbol{\Theta}$) contains the q parameters of interest for signal j. If we take a set of p frames we have

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\Theta})\mathbf{X} + \mathbf{N} \tag{2.2.8}$$

where

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_p \end{bmatrix},\ \mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \dots & \mathbf{y}_p \end{bmatrix},\ \mathbf{N} = \begin{bmatrix} \mathbf{n}_1 & \mathbf{n}_2 & \dots & \mathbf{n}_p \end{bmatrix}. \tag{2.2.9}$$

($\mathbf{X} \in \mathbb{C}^{m\times p}, \mathbf{N}, \mathbf{Y} \in \mathbb{C}^{n\times p}$). In (2.2.8) the only quantities that are known are the observed values, the elements of $\mathbf{Y}$. The quantities that are required, ultimately, are the parameters, in $\boldsymbol{\Theta}$, but we do not find them directly, but only the target PSVs, which are the m columns of $\mathbf{A}$. We are not interested in the waveforms as such, so the first step in MUSIC and many algorithms of this kind, is to form the estimated covariance matrix, which reduces the waveforms to power-like covariance values (or second order statistics). This matrix is given (though the scaling factor p is often omitted) by

$$\begin{aligned} \mathbf{R}_\mathbf{Y} &= \mathbf{Y}\mathbf{Y}^{\mathrm{H}} \big/ p = \left(\mathbf{A}(\boldsymbol{\Theta})\mathbf{X} + \mathbf{N}\right)\left(\mathbf{X}^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})^{\mathrm{H}} + \mathbf{N}^{\mathrm{H}}\right) \big/ p \\ &= \left(\mathbf{A}(\boldsymbol{\Theta})\mathbf{X}\mathbf{X}^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})^{\mathrm{H}} + \mathbf{A}(\boldsymbol{\Theta})\mathbf{X}\mathbf{N}^{\mathrm{H}} + \mathbf{N}\mathbf{X}^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})^{\mathrm{H}} + \mathbf{N}\mathbf{N}^{\mathrm{H}}\right) \big/ p. \end{aligned} \tag{2.2.10}$$

If we take the expectation values of the terms in this equation we obtain the actual covariance values. Using $\langle\cdot\rangle$ to indicated the expectation of the argument, we put

$$\langle \mathbf{Y}\mathbf{Y}^{\mathrm{H}} \rangle = p\bar{\mathbf{R}}_\mathbf{Y},\ \langle \mathbf{X}\mathbf{X}^{\mathrm{H}} \rangle = p\bar{\mathbf{R}}_\mathbf{X},\ \langle \mathbf{X}\mathbf{N}^{\mathrm{H}} \rangle = \mathbf{0},\ \langle \mathbf{N}\mathbf{N}^{\mathrm{H}} \rangle = p\psi\mathbf{I} \tag{2.2.11}$$

where $\bar{\mathbf{R}}_\mathbf{X}$ and $\bar{\mathbf{R}}_\mathbf{Y}$ are the covariance matrices of the waveforms of which $\mathbf{X}$ and $\mathbf{Y}$ are samples of length p (i.e. of sizes $m\times p$ and $n\times p$ respectively), the expectation products of the noise and signal waveforms, being uncorrelated, are zero, the noise waveforms are also independent, so the off-diagonal terms of their covariance matrix are also zero and their variances (or powers) are ψ, appearing on the main diagonal. Using (2.2.11) in (2.2.10) gives

$$\bar{\mathbf{R}}_\mathbf{Y} = \mathbf{A}(\boldsymbol{\Theta})\bar{\mathbf{R}}_\mathbf{X}\mathbf{A}(\boldsymbol{\Theta})^{\mathrm{H}} + \psi\mathbf{I}. \tag{2.2.12}$$

Now let $\mathbf{v}$ be an eigenvector of $\bar{\mathbf{R}}_\mathbf{Y}$ with corresponding eigenvalue λ (so that $\bar{\mathbf{R}}_\mathbf{Y}\mathbf{v} = \lambda\mathbf{v}$) then we have

$$\mathbf{A}\bar{\mathbf{R}}_\mathbf{X}\mathbf{A}^{\mathrm{H}}\mathbf{v} = \left(\bar{\mathbf{R}}_\mathbf{Y} - \psi\mathbf{I}\right)\mathbf{v} = (\lambda - \psi)\mathbf{v} \tag{2.2.13}$$

and we see that the eigenvectors of $\bar{\mathbf{R}}_{\mathbf{Y}}$ are also eigenvectors of $\mathbf{A}\bar{\mathbf{R}}_{\mathbf{X}}\mathbf{A}^{\mathrm{H}}$, which has eigenvalues ψ smaller. (We omit the indication of dependence on $\boldsymbol{\Theta}$ temporarily). If $\mathbf{V}$ is the full set of eigenvectors and $\boldsymbol{\Lambda}$ a diagonal $n \times n$ matrix containing the eigenvalues of $\bar{\mathbf{R}}_{\mathbf{Y}}$ then we have

$$\mathbf{A}\bar{\mathbf{R}}_{\mathbf{X}}\mathbf{A}^{\mathrm{H}}\mathbf{V} = \mathbf{V}\boldsymbol{\Lambda}' = \mathbf{V}(\boldsymbol{\Lambda} - \psi\mathbf{I}) \tag{2.2.14}$$

where $\boldsymbol{\Lambda}'$ contains the eigenvalues of $\mathbf{A}\bar{\mathbf{R}}_{\mathbf{X}}\mathbf{A}^{\mathrm{H}}$, given by $\lambda_k' = \lambda_k - \psi$ (k = 1 to n).

A matrix of the form of $\bar{\mathbf{R}}_{\mathbf{Y}}$ is Hermitian ($\bar{\mathbf{R}}_{\mathbf{Y}} = \bar{\mathbf{R}}_{\mathbf{Y}}{}^{\mathrm{H}}$) and positive definite ($\mathbf{x}^{\mathrm{H}}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\mathbf{x} = \mathbf{z}^{\mathrm{H}}\mathbf{z} > 0$ for all $\mathbf{x} \neq \mathbf{0}$, where $\mathbf{z} = \mathbf{Y}^{\mathrm{H}}\mathbf{x}$, if $\mathbf{Y}$ has full rank, which is the case here because of the independent noise waveforms) so has real, positive eigenvalues and orthogonal eigenvectors. The matrix $\mathbf{A}\bar{\mathbf{R}}_{\mathbf{X}}\mathbf{A}^{\mathrm{H}}$ is also Hermitian but $\mathbf{A}$, which is an $n \times m$ matrix is only of rank m so $\mathbf{A}\bar{\mathbf{R}}_{\mathbf{X}}\mathbf{A}^{\mathrm{H}}$ is positive semidefinite, with $n - m$ eigenvalues of zero. (If $\mathbf{x}$ is in the null space of $\mathbf{A}$, we have $\mathbf{A}^{\mathrm{H}}\mathbf{x} = \mathbf{0}$ so that we can only write $\mathbf{x}^{\mathrm{H}}\mathbf{A}\bar{\mathbf{R}}_{\mathbf{X}}\mathbf{A}^{\mathrm{H}}\mathbf{x} \geq 0$ for all $\mathbf{x}$). Thus $n - m$ eigenvalues of $\bar{\mathbf{R}}_{\mathbf{Y}}$ have value ψ and the other m have values greater than ψ. As the eigenvectors of $\bar{\mathbf{R}}_{\mathbf{Y}}$ (and also of $\mathbf{A}\bar{\mathbf{R}}_{\mathbf{X}}\mathbf{A}^{\mathrm{H}}$) are orthogonal and are taken to be normalized, we have $\mathbf{V}\mathbf{V}^{\mathrm{H}} = \mathbf{I}_n$, so (2.2.14) can be put in the form

$$\mathbf{A}\bar{\mathbf{R}}_{\mathbf{X}}\mathbf{A}^{\mathrm{H}} = \mathbf{V}\boldsymbol{\Lambda}'\mathbf{V}^{\mathrm{H}} \tag{2.2.15}$$

(on multiplying both sides on the right by $\mathbf{V}^{\mathrm{H}}$).

Now we partition $\mathbf{V}$ into the vectors with non-zero eigenvalues and those with zero eigenvalues, and $\boldsymbol{\Lambda}'$ similarly, so that

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_s & \mathbf{V}_n \end{bmatrix} \text{ and } \boldsymbol{\Lambda}' = \begin{bmatrix} \boldsymbol{\Lambda}_s' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{2.2.16}$$

and substituting these in (2.2.15) gives

$$\mathbf{A}\bar{\mathbf{R}}_{\mathbf{X}}\mathbf{A}^{\mathrm{H}} = \mathbf{V}_s\boldsymbol{\Lambda}_s'\mathbf{V}_{\mathbf{s}}{}^{\mathrm{H}}. \tag{2.2.17}$$

Now the n columns of the matrix on the left of (2.2.17) are all linear combinations of the m columns of $\mathbf{A}$, and the columns of the matrix on the right are linear combinations of the m columns of $\mathbf{V}_s$ – but these matrices are equal, so we deduce that $\mathbf{A}$ and $\mathbf{V}_s$ span the same m-dimensional space. This does not mean that $\mathbf{A} = \mathbf{V}_s$, so we have not yet

found the signal PSVs, the columns of **A**, but are now in a position to do so. We know that all the target PSVs lie in the space spanned by the columns of $\mathbf{V}_s$ (which is termed the signal space, or subspace, as distinct from the noise space, the column space of $\mathbf{V}_n$) so to find which of the manifold vectors these are we now project all the stored manifold vectors into the signal subspace and deduce that the m manifold vectors which lie most closely in this space are those corresponding to the signals present. In order to measure this 'closeness' we simply compute the square magnitude of the projected vector. This vector has maximum magnitude unity, corresponding to a unit magnitude manifold vector lying exactly in the signal subspace. The magnitude will not be exactly unity, in practice, because of various sources of error, such as (a) the discrete sampling of the manifold, (b) errors in the manifold description due to imperfect modelling (with unknown multipath or mutual coupling effects) or small errors in the experimental calibration and (c) the approximation of the estimated covariance matrix to the true one. We have also assumed that the manifold vectors are unique for each point in the parameter domain. In many cases this is true, but in some cases two or more points may give rise to the same PSV. An example of this is the case of a horizontal linear array used for azimuth DF. In this case the vector for angle α (measured from broadside) is the same as the vector for $180° - \alpha$. Another source of ambiguity is when using a regular array with element separations of more than a half wavelength. Planar arrays of the form of a regular hexagon can also have high ambiguity. However in most cases ambiguity can be avoided by suitable array design or can be accepted and resolved using collateral information.

If $\mathbf{P}_{\mathbf{V}_s}$ is the matrix which projects into the space spanned by the columns of $\mathbf{V}_s$ then the MUSIC function is given by

$$f_{\mathrm{M}}(\boldsymbol{\theta}) = \left\| \mathbf{P}_{\mathbf{V}_s} \mathbf{a}(\boldsymbol{\theta}) \right\|^2 = \mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}} \mathbf{P}_{\mathbf{V}_s}{}^{\mathrm{H}} \mathbf{P}_{\mathbf{V}_s} \mathbf{a}(\boldsymbol{\theta}) = \mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}} \mathbf{P}_{\mathbf{V}_s} \mathbf{a}(\boldsymbol{\theta}) \qquad (2.2.18)$$

where $\left\| \mathbf{P}_{\mathbf{V}_s} \mathbf{a}(\boldsymbol{\theta}) \right\|^2$ is the square magnitude of the projection of the manifold vector for parameter set $\boldsymbol{\theta}$ and we use projection matrix properties ($\mathbf{P}^{\mathrm{H}} = \mathbf{P}$, $\mathbf{P}^2 = \mathbf{P}$) to simplify the expression. The projection matrix is given by

$$\mathbf{P}_{\mathbf{V}_s} = \mathbf{V}_s (\mathbf{V}_s{}^{\mathrm{H}} \mathbf{V}_s)^{-1} \mathbf{V}_s{}^{\mathrm{H}} = \mathbf{V}_s \mathbf{V}_s{}^{\mathrm{H}} \qquad (2.2.19)$$

in this case, as the columns of $\mathbf{V}_s$ are orthonormal, so that $\mathbf{V}_s{}^{\mathrm{H}} \mathbf{V}_s = \mathbf{I}_m$. Putting this into (2.2.18) gives, finally,

$$f_{\mathrm{M}}(\boldsymbol{\theta}) = \mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}} \mathbf{V}_{\mathrm{s}} \mathbf{V}_{\mathrm{s}}^{\mathrm{H}} \mathbf{a}(\boldsymbol{\theta}) = \left\| \mathbf{V}_{\mathrm{s}}^{\mathrm{H}} \mathbf{a}(\boldsymbol{\theta}) \right\|^2 \tag{2.2.20}$$

and this is the usual form of the MUSIC function. (We note that $\mathbf{P}_{\mathbf{V}_s} \mathbf{a}(\boldsymbol{\theta}) = \mathbf{V}_s \mathbf{V}_s^{\mathrm{H}} \mathbf{a}(\boldsymbol{\theta})$ so that this projected vector is a linear combination of the columns of $\mathbf{V}_{\mathrm{s}}$ – so lies in the signal subspace, and the coefficients of these vectors are given by $\mathbf{V}_{\mathrm{s}}\mathbf{a}(\boldsymbol{\theta})$. As the columns of $\mathbf{V}_{\mathrm{s}}$ are orthogonal and of unit length, the square of the length of the projected vector is the sum of the squares of these coefficients, as given in (2.2.20)).

Thus, given a block of data $\mathbf{Y}$, and a stored sampled manifold of PSVs of form $\mathbf{a}(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a set of q parameters, the MUSIC algorithm can be given as

1) form the data covariance matrix $\mathbf{R}_{\mathbf{Y}} = \mathbf{Y}\mathbf{Y}^{\mathrm{H}}$

2) perform an eigenanalysis on $\mathbf{R}_{\mathbf{Y}}$

3) find the m eigenvalues above noise level, ψ, and select the corresponding eigenvectors, $\mathbf{V}_{\mathrm{s}}$

4) For all $\boldsymbol{\theta}$ in the domain evaluate the MUSIC function (2.2.20)

5) find the m highest peaks in this function; the PSVs at the peak positions are those corresponding to the estimated target parameter sets.

(Or, equivalently, we could use the $n - m$ eigenvectors $\mathbf{V}_n$ corresponding to the eigenvalues at noise level, and put $\mathbf{V}_n$ in (2.2.20) instead of $\mathbf{V}_s$, to project into the noise space, and then we look for the m lowest points – nulls, apart from the effect of errors).

Figure 2.3 illustrates in diagrammatic form the principle of the MUSIC search. We assume here there are just two targets present with PSVs $\mathbf{a}_1$ and $\mathbf{a}_2$. The eigenanalysis finds two eigenvalues above the threshold ψ and the two corresponding eigenvectors are $\mathbf{v}_1$ and $\mathbf{v}_2$, which are orthogonal. These define the 'signal subspace' – the vector space containing the target, or signal, PSVs. We then take each of the manifold vectors $\mathbf{a}(\boldsymbol{\theta})$ and project this into the signal space to form a vector

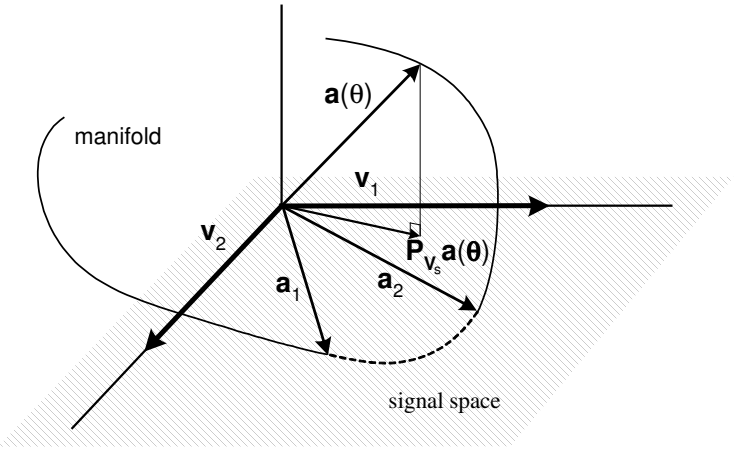
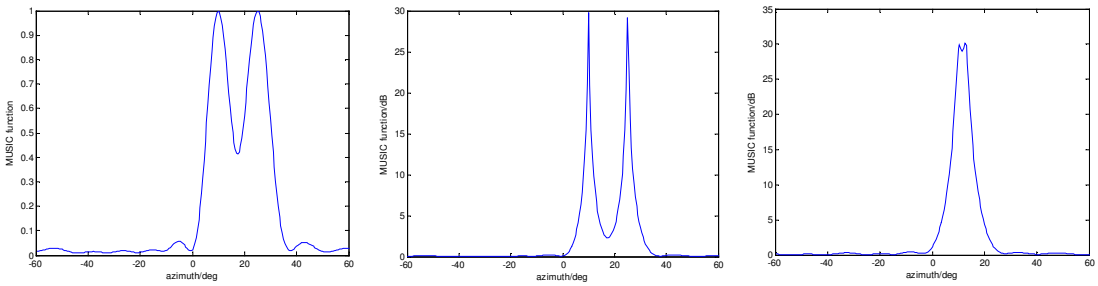


Figure 2.3 Signal space and MUSIC function

$\mathbf{P}_{\mathbf{V}_s}\mathbf{a}(\boldsymbol{\theta})$ and find its square magnitude. Only vectors lying in the space, as $\mathbf{a}_1$ and $\mathbf{a}_2$ do, will be unchanged, with magnitude unity, while all the others will only have some component, a part of the vector, in this space and will have magnitude less than unity.

As stated above, we could alternatively project all the manifold vectors into the noise subspace (represented in Figure 2.3 by the Z axis, orthogonal to the signal subspace of the XY plane). In this case the projected vector is $\mathbf{Q}_{\mathbf{V}_s}\mathbf{a}(\boldsymbol{\theta})$ where $\mathbf{Q}_{\mathbf{V}_s} = \mathbf{I} - \mathbf{P}_{\mathbf{V}_s}$ (and we can also show that $\mathbf{Q}_{\mathbf{V}_s} = \mathbf{P}_{\mathbf{V}_n} = \mathbf{V}_n\mathbf{V}_n^{\mathrm{H}}$). For the manifold vectors, such as $\mathbf{a}_1$ and $\mathbf{a}_2$, which lie in the signal space this projection is of zero length, so in this case we look for the *m* lowest minima, and the parameters characterizing the PSVs giving these minima are the parameter estimates for the *m* point targets. This is the form corresponding to the description in §2.1.6.2. The MUSIC function in this case is $\left\|\mathbf{Q}_{\mathbf{V}_s}\mathbf{a}(\boldsymbol{\theta})\right\|^2 = 1 - \left\|\mathbf{P}_{\mathbf{V}_s}\mathbf{a}(\boldsymbol{\theta})\right\|^2 = 1 - \left\|\mathbf{V}_s\mathbf{a}(\boldsymbol{\theta})\right\|^2 = \left\|\mathbf{V}_n\mathbf{a}(\boldsymbol{\theta})\right\|^2$, but often the reciprocal of this is plotted, so that we again search for peaks (maxima) rather than minima. In this case a perfect match of a manifold vector to the signal point response corresponds to an infinite spike, and often the spikes at the signal parameter positions can in practice be quite high, and quite narrow. This looks quite striking, as if spectacular performance is possible, but in fact the positions of these spikes are no more precise than those given by the much smaller and rounder peaks of (2.2.20). Nor does the narrowness of the spikes mean that the resolution is correspondingly good – as targets are moved closer together and the corresponding MUSIC spikes approach each other the dip between the peaks rises rapidly and disappears at much greater separation than the spike widths.



(a) linear form (b) reciprocal form (c) close targets (log)

Figure 2.4 MUSIC plots

These points are illustrated in Figure 2.4. Figure 2.4(a) shows the MUSIC function as defined in (2.2.20), plotting a linear, power-like function, with maximum value unity. (The signal sources are at 10° and 25°, here). The reciprocal form, as defined in the

preceding paragraph, is plotted in logarithmic (dB) form in (b), and we see the very sharp peaks at the signal positions. In (c) the targets are moved much closer (2.8° apart) and the nulls between the peaks are almost filled in. (The case taken for illustration was a 9 element, uniform linear array at 0.6 wavelength spacing. There was no perturbing noise, but small manifold errors were added to reduce the peaks to realistic levels.)

2.2.3 The IMP superresolution method

2.2.3.1 Structure of IMP

We avoid using the term 'algorithm' here because there is no single definitive form for the IMP process; the general principle outlined below can be implemented with considerable variations in detail. It seems more appropriate to consider IMP as a principle for achieving parameter estimation from which suitable algorithms can be formulated.

The IMP process basically consists of a sequence of power scans over the parameter domain, which might be multidimensional. The function evaluated is effectively a signal to noise ratio over the parameter range, with nulls inserted at the positions of all the signals found previously. Thus on the first scan no null is present and a signal position is estimated (assuming there is a peak above some detection threshold). This position is in the parameter domain, so might be in azimuth (Ex.1), azimuth and elevation, range (Ex.2), frequency (Ex.3), for example. On the second scan, with a null at this position, a second signal is estimated. After two signal positions have been estimated an iterative process of refinement of these positions is carried out, called tweaking. In this process a null is placed on one signal, removing its sidelobe power to a large extent, and the position of the other signal is found, and then a null is placed on this, improved, signal position and the position of the first signal is then found more accurately. This process continues until there is good

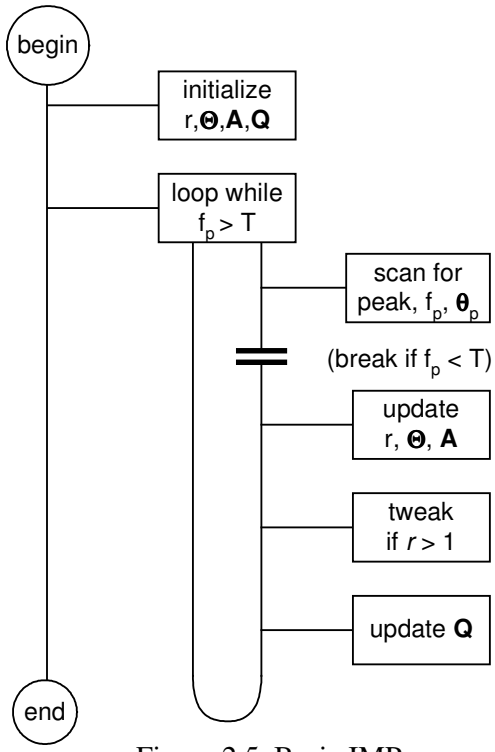


Figure 2.5 Basic IMP

convergence and the changes in both signal positions are small. For the next scan nulls are placed at both signal positions. When more signals are found, on further scans, nulls are placed on all signals except the one being tweaked, and the process cycles round all the signals found at this stage. When all significant signal contributions have been found, the scan, with nulls on all these signals, will be essentially at noise level, and there will be no peak above the detection threshold.

We outline the structure of the IMP program with the help of design structure diagrams. Figure 2.5 shows the overall form of the IMP program, given a stored sampled manifold **M**, consisting of the point source vectors (PSVs) of a set of points over the parameter space. Initially the number of signals found, r, is set to zero, the array of the set of estimated parameters $\boldsymbol{\Theta}$ is set empty, and so is the matrix of the set of signal PSVs, **A**. The projection matrix **Q**, which inserts nulls and is calculated from **A**, is initially set to the identity (of order n, the number of samples in a frame of data).

With the scan peak value f_p initially set above the detection threshold T, we enter the main loop. The IMP function is calculated using the manifold, and f_p is set to the peak value. If this exceeds the threshold the peak position is noted, and refined using a quadratic interpolation estimate. (This uses points round the peak and is based on the fact that the function is flat at the peak so the variation round the peak is, to lowest order, quadratic. For a single parameter this requires only 3 points, for two parameters at least 6, and preferably 9, points are used and for three parameters 19 are used.) The number of signals found is updated, and so is the set of signal parameter estimates $\boldsymbol{\Theta}$, and the PSV for this signal is calculated and added, as a new column, to **A**. On subsequent scans, when more than one signal has been found, the tweaking routine is carried out. Finally the projection matrix **Q**, which inserts the nulls in the signal positions, is updated, ready for the next scan.

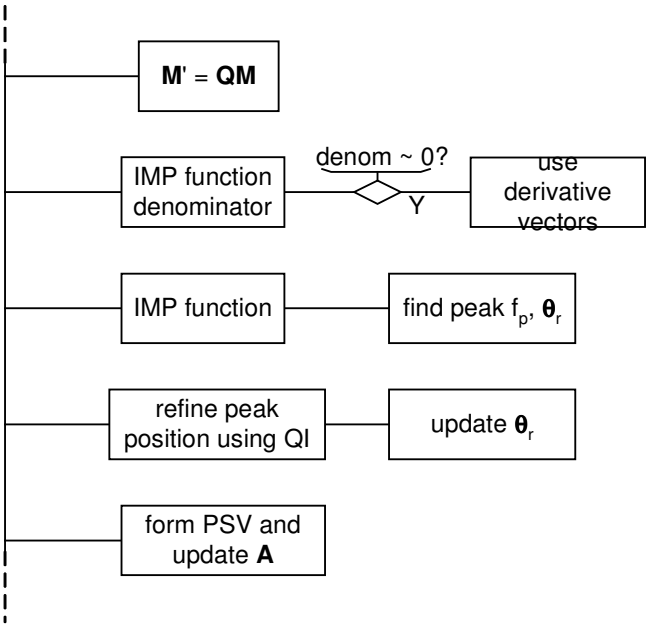


Figure 2.6 IMP scan

Figure 2.6 shows the scan routine in more detail. We start by multiplying all the manifold vectors by **Q**, which effectively means that we form the IMP response for each parameter point specified by the manifold vector, with nulls at the positions of signals found so far. If one of these nulls is at a manifold vector position then the IMP function denominator will be very close to zero, even though the function should give a valid value at this point. This value will be obtained securely if the derivative of this manifold vector (with respect to at least one parameter) is used (a form of L'Hôpital's rule). The full IMP function (given in (2.2.26) below) can now be evaluated and the peak found. This peak position estimate is refined using quadratic interpolation (QI) as described above and then **Θ** and **A** are updated.

The tweak routine is shown in Figure 2.7. *D* is the measure of the degree of convergence. It is the sum of the squares of the latest tweak corrections to the *r* points found so far. The latest tweak shift for target *k* is recorded (if significant, above some threshold *th*) as *d*(*k*), a component in an *r*-vector **d**, and *D* is the value of the sum of the

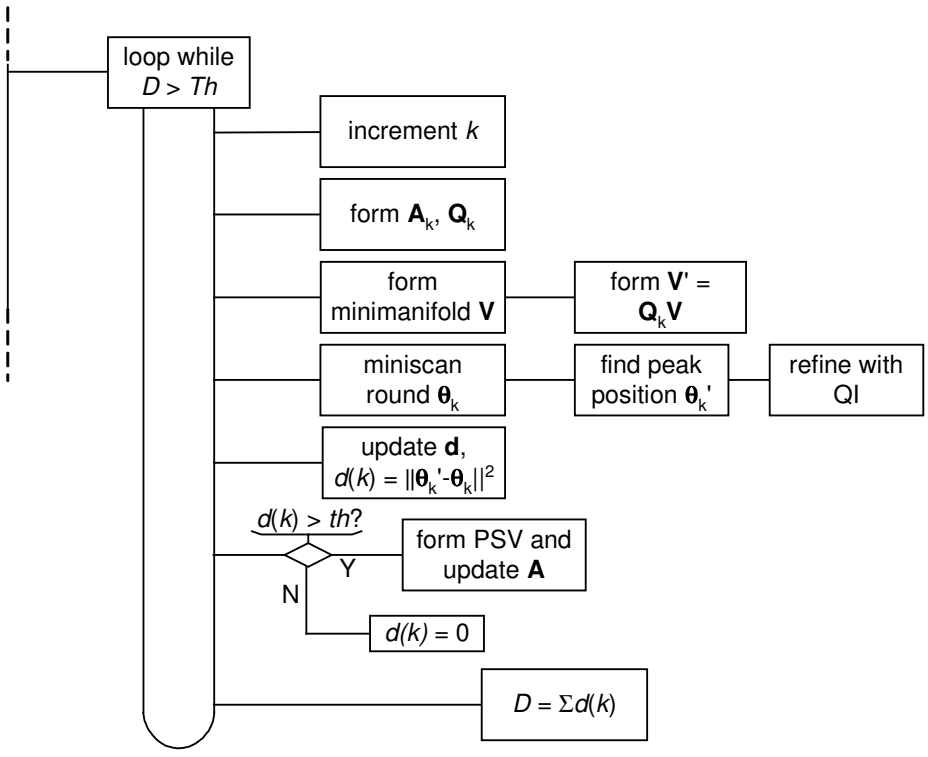


Figure 2.7 IMP tweak

components of **d**. When *D*, which is evaluated after every tweak operation, falls below some minimum threshold *Th* the tweak routine is exited.

Starting with the latest signal to be found, for each signal *k* we form $\mathbf{A}_k$ which is the set of signal PSVs **A** *excluding* that for signal *k*. We then form $\mathbf{Q}_k$, which will provide nulls

to all the signals except this signal. We form a mini-manifold $\mathbf{V}_k$, about $\boldsymbol{\theta}_k$, the position of signal k and then evaluate the IMP function over this region (much smaller than the full scan region, so more economical in computation), in a similar manner to the full IMP scan, already described. Again the peak position is corrected by quadratic interpolation, and the tweak shift, $d(k)$ is found as the (square) distance between the new peak position $\boldsymbol{\theta}_k'$ and the old, $\boldsymbol{\theta}_k$. If $d(k)$ is below a threshold th the shift is considered negligible, but otherwise $\boldsymbol{\theta}$ is updated and so is the PSV for signal k and then $\mathbf{A}$. Finally D is evaluated and the loop is left if D is small enough.

The iteration may take a long time to converge in some cases, so a maximum iteration number might be used. An accelerated convergence method could also be introduced which detects the trend of the convergence and jumps to the estimated convergence point, which, even if not completely accurate, will reduce the number of steps needed.

2.2.3.2 The IMP function

We first derive the IMP function from the physical description of a signal to noise ratio, with nulls, and then identify this function with the condensed likelihood function (maximized over the waveform values $\mathbf{X}$ and the receiver noise value ψ) found in Appendix 2A. Let a block of data $\mathbf{Y} \in \mathbb{C}^{n \times p}$ (i.e. p frames of n values over the observation aperture) be received and let this be 'beamformed' using a weight vector $\mathbf{w}$. (If the n samples are from n channels fed from outputs of the elements of an antenna array then this will indeed be a spatial beamforming process. More generally the output is an element of a transform, in some cases a simple Fourier transform.) The beam output waveform is given by

$$\mathbf{z} = \mathbf{Y}^{\mathrm{T}}\mathbf{w} \tag{2.2.21}$$

($\mathbf{z} \in \mathbb{C}^{p \times 1}$) and the energy in this output is given by

$$\|\mathbf{z}\|^2 = \mathbf{z}^{\mathrm{T}}\mathbf{z}^* = \mathbf{w}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\mathbf{w}^*. \tag{2.2.22}$$

(The asterix indicates complex conjugate.) Now let $\mathbf{a}(\boldsymbol{\theta})$ be the (normalized) point source vector for parameter vector $\boldsymbol{\theta}$. (We assume in general we are looking for q parameters for each target. In many cases $q = 1$, for example the single parameter may be azimuth direction. In the case of more general direction finding $q = 2$, where the parameters may be azimuth and elevation, or alternatively u and v, where these are direction cosine coordinates.) The weight vector which maximizes the signal to noise ratio for a target with the parameters $\boldsymbol{\theta}$ (with 'white' noise across the channels) can be

shown to be $\mathbf{a}(\boldsymbol{\theta})^*$ (see Appendix 2B). If j targets have been found so far, with parameter vectors $\boldsymbol{\theta}_1$ to $\boldsymbol{\theta}_j$ and PSVs $\mathbf{a}_1 = \mathbf{a}(\boldsymbol{\theta}_1)$ to $\mathbf{a}_j = \mathbf{a}(\boldsymbol{\theta}_j)$ then to put nulls on these point sources while steering at a point with PSV $\mathbf{a}(\boldsymbol{\theta})$ we define a matrix $\mathbf{A}_j = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad . \quad . \quad . \quad \mathbf{a}_j] \in \mathbb{C}^{n\times j}$ and use the weight vector

$$\mathbf{w} = \mathbf{Q}_{\mathbf{A}_j *}\mathbf{a}(\boldsymbol{\theta})^* \tag{2.2.23}$$

which ensures that $\mathbf{w}$ is orthogonal to $\mathbf{a}_1^*$ to $\mathbf{a}_j^*$ (i.e. $\mathbf{w}^{\mathrm{T}}\mathbf{a}_k = 0 = \mathbf{w}^{\mathrm{H}}\mathbf{a}_k{}^*$, k = 1 to j.) (It is also shown in Appendix 2B that the weight vector in (2.2.23) maximizes signal reception while meeting the condition of maintaining the nulls.) Substituting this weight vector in (2.2.22) we have

$$\text{total output energy} = \mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}_j}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}_j}\mathbf{a}(\boldsymbol{\theta}) \tag{2.2.24}$$

(using $(\mathbf{Q}_{\mathbf{A}*})^* = \mathbf{Q}_{\mathbf{A}}$ and $\mathbf{Q}_{\mathbf{A}}{}^{\mathrm{H}} = \mathbf{Q}_{\mathbf{A}}$). The noise energy is (proportional to) $\|\mathbf{w}\|^2$, for the 'white' case, with equal noise power levels in each channel, and from (2.2.23) this is

$$\|\mathbf{w}\|^2 = \mathbf{w}^{\mathrm{T}}\mathbf{w}^* = \mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}_j}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}_j}\mathbf{a}(\boldsymbol{\theta}) = \mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}_j}\mathbf{a}(\boldsymbol{\theta}) \tag{2.2.25}$$

using projection matrix properties (see (2.A.20)). Thus the IMP function, expressed as a signal-plus-noise to noise ratio when steered for a point source with parameters $\boldsymbol{\theta}$, with nulls to sources with parameters $\boldsymbol{\theta}_1$ to $\boldsymbol{\theta}_j$ is given, from (2.2.24) and (2.2.25), by

$$f(\boldsymbol{\theta}) = \frac{\mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}_j}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}_j}\mathbf{a}(\boldsymbol{\theta})}{\mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}_j}\mathbf{a}(\boldsymbol{\theta})}. \tag{2.2.26}$$

This is the function that is evaluated, over the q-dimensional domain of $\boldsymbol{\theta}$, in the IMP routine described in §2.3.1 above. If we write $\mathbf{Y}\mathbf{Y}^{\mathrm{H}} = \mathbf{R}_{\mathbf{Y}}$ and $\mathbf{b}(\boldsymbol{\theta}) = \mathbf{Q}_{\mathbf{A}_j}\mathbf{a}(\boldsymbol{\theta})$ then (2.2.26) can be written

$$f(\boldsymbol{\theta}) = \frac{\mathbf{b}^{\mathrm{H}}\mathbf{R}_{\mathbf{Y}}\mathbf{b}}{\|\mathbf{b}\|^2} \quad (\mathbf{b} = \mathbf{Q}_{\mathbf{A}_j}\mathbf{a}(\boldsymbol{\theta})) \tag{2.2.27}$$

and this is a relatively economical way of evaluating the function, when the number of frames, p, is relatively large. If p is small it is economical to evaluate $\mathbf{c} = \mathbf{b}^{\mathrm{H}}\mathbf{Y}$ and then $f = \|\mathbf{c}\|^2/\|\mathbf{b}\|^2$ and if $p = 1$ so that $\mathbf{Y} = \mathbf{y}$, then we have simply $c = \mathbf{b}^{\mathrm{H}}\mathbf{y}$ and $f = |c|^2/\|\mathbf{b}\|^2$.

A condensed likelihood function is given in Appendix 1 (eq. (2.A.28)) as $\tilde{F}(\boldsymbol{\Theta}) = \mathrm{tr}(\mathbf{P}_{\mathbf{A}(\boldsymbol{\Theta})}\mathbf{R}_{\mathbf{Y}})$, where tr(.) is the trace of the matrix argument. This function is to

be maximized over $\boldsymbol{\Theta}$ to find the most likely parameter values, i.e. the ML solution for the parameters is given by

$$\hat{\boldsymbol{\Theta}} = \arg \max_{\boldsymbol{\Theta}} \left\{ \mathrm{tr}(\mathbf{P}_{\mathbf{A}(\boldsymbol{\Theta})} \mathbf{R}_{\mathbf{Y}}) \right\} \tag{2.2.28}$$

and the problem is to search over the domain of $\boldsymbol{\Theta}$ (which is of dimension *mq*) in some economical way to find the maximizing set of values. In IMP this is achieved by a series of *m* searches over a *q*-dimensional domain, finding the set of parameters for each signal in turn.

The full (condensed) likelihood function when *m* targets are present is

$$\tilde{F}(\boldsymbol{\Theta}_m) = \mathrm{tr}(\mathbf{P}_m \mathbf{R}_{\mathbf{Y}}) \tag{2.2.29}$$

where we have introduced the notation

$$\boldsymbol{\Theta}_k = \left[\boldsymbol{\theta}_1 \quad \boldsymbol{\theta}_2 \quad . \quad . \quad . \quad \boldsymbol{\theta}_k \right] \tag{2.2.30}$$

and $$\mathbf{P}_k = \mathbf{P}_{\mathbf{A}_k} \tag{2.2.31}$$

where $$\mathbf{A}_k = \left[\mathbf{a}_1 \quad \mathbf{a}_2 \quad . \quad . \quad . \quad \mathbf{a}_k \right] = \left[\mathbf{a}(\boldsymbol{\theta}_1) \quad \mathbf{a}(\boldsymbol{\theta}_2) \quad . \quad . \quad . \quad \mathbf{a}(\boldsymbol{\theta}_k) \right] \tag{2.2.32}$$

for $k = 1$ to m. Now we put

$$\mathbf{A}_m = \left[\mathbf{A}_{m-1} \quad \mathbf{a}_m \right] \tag{2.2.33}$$

and we now require the projection matrix for a partitioned matrix. This is given by

$$\mathbf{P}_{[\mathbf{B} \;\; \mathbf{C}]} = \mathbf{P}_{\mathbf{B}} + \mathbf{P}_{\bar{\mathbf{C}}} \tag{2.2.34}$$

where $$\bar{\mathbf{C}} = \mathbf{Q}_{\mathbf{B}} \mathbf{C}, \tag{2.2.35}$$

i.e. the projection into the space spanned by the columns of **B** and **C** (where the columns of the full matrix are linearly independent) is found, first by modifying **C** to $\bar{\mathbf{C}}$ which defines (or is a basis for) a space orthogonal to that of the columns of **B**, and then summing the projections into the two spaces, which are now non-overlapping. In this case we have, from (2.2.31), (2.2.33) and (2.2.34)

$$\mathbf{P}_m = \mathbf{P}_{m-1} + \mathbf{P}_{\mathbf{b}_m} \tag{2.2.36}$$

where, from (2.2.35) and (2.2.33),

$$\mathbf{b}_m = \mathbf{Q}_{m-1} \mathbf{a}_m . \tag{2.2.37}$$

Now $$\mathbf{P}_{\mathbf{b}_m} = \mathbf{b}_m(\mathbf{b}_m{}^{\mathrm{H}}\mathbf{b}_m)^{-1}\mathbf{b}_m{}^{\mathrm{H}} = \mathbf{b}_m\mathbf{b}_m{}^{\mathrm{H}}\big/\|\mathbf{b}_m\|^2 \tag{2.2.38}$$

so now we can write, from (2.2.29), (2.2.36) and (2.2.38),

$$\tilde{F}(\boldsymbol{\Theta}_m) = \mathrm{tr}(\mathbf{P}_m\mathbf{R}_{\mathbf{Y}}) = \mathrm{tr}\left((\mathbf{P}_{m-1} + \mathbf{b}_m\mathbf{b}_m{}^{\mathrm{H}}\big/\|\mathbf{b}_m\|^2)\mathbf{R}_{\mathbf{Y}}\right) = \mathrm{tr}\left(\mathbf{P}_{m-1}\mathbf{R}_{\mathbf{Y}}\right) + \mathrm{tr}\left(\mathbf{b}_m\mathbf{b}_m{}^{\mathrm{H}}\mathbf{R}_{\mathbf{Y}}\right)\big/\|\mathbf{b}_m\|^2 .$$

Now (see (2.A.22)) $\mathrm{tr}\left(\mathbf{b}_m\mathbf{b}_m{}^{\mathrm{H}}\mathbf{R}_{\mathbf{Y}}\right) = \mathrm{tr}\left(\mathbf{b}_m{}^{\mathrm{H}}\mathbf{R}_{\mathbf{Y}}\mathbf{b}_m\right) = \mathbf{b}_m{}^{\mathrm{H}}\mathbf{R}_{\mathbf{Y}}\mathbf{b}_m$ (as this is scalar) so that,

$$\tilde{F}(\boldsymbol{\Theta}_m) = \mathrm{tr}\left(\mathbf{P}_{m-1}\mathbf{R}_{\mathbf{Y}}\right) + \mathbf{b}_m{}^{\mathrm{H}}\mathbf{R}_{\mathbf{Y}}\mathbf{b}_m\big/\|\mathbf{b}_m\|^2 . \tag{2.2.39}$$

We can perform the same expansion on $\mathrm{tr}\left(\mathbf{P}_{m-1}\mathbf{R}_{\mathbf{Y}}\right)$, and then on $\mathrm{tr}\left(\mathbf{P}_{m-2}\mathbf{R}_{\mathbf{Y}}\right)$, and so on, to obtain

$$\tilde{F}(\boldsymbol{\Theta}_m) = \mathrm{tr}\left(\mathbf{P}_1\mathbf{R}_{\mathbf{Y}}\right) + \sum_{k=2}^{m} \mathbf{b}_k{}^{\mathrm{H}}\mathbf{R}_{\mathbf{Y}}\mathbf{b}_k\big/\|\mathbf{b}_k\|^2$$

but as $\mathbf{P}_1 = \mathbf{P}_{\mathbf{a}_1} = \mathbf{a}_1\mathbf{a}_1{}^{\mathrm{H}}/\|\mathbf{a}_1\|^2$ we have $\mathrm{tr}\left(\mathbf{P}_1\mathbf{R}_{\mathbf{Y}}\right) = \mathbf{a}_1{}^{\mathrm{H}}\mathbf{R}_{\mathbf{Y}}\mathbf{a}_1\big/\|\mathbf{a}_1\|^2$, and also, as $\mathbf{Q}_0 = \mathbf{I}$, the identity (of order *n*), we have, formally, $\mathbf{b}_1 = \mathbf{Q}_0\mathbf{a}_1 = \mathbf{a}_1$ and so, finally,

$$\tilde{F}(\boldsymbol{\Theta}_m) = \sum_{k=1}^{m} \mathbf{b}_k{}^{\mathrm{H}}\mathbf{R}_{\mathbf{Y}}\mathbf{b}_k\big/\|\mathbf{b}_k\|^2 \tag{2.2.40}$$

or $$\tilde{F}(\boldsymbol{\Theta}_m) = \sum_{k=1}^{m} \mathbf{a}_k{}^{\mathrm{H}}\mathbf{Q}_{k-1}\mathbf{R}_{\mathbf{Y}}\mathbf{Q}_{k-1}\mathbf{a}_k\big/\|\mathbf{Q}_{k-1}\mathbf{a}_k\|^2 . \tag{2.2.41}$$

Thus we see that the condensed ML function for *m* targets, expressed as the trace of an *n*×*n* matrix, $\mathbf{P}_m\mathbf{R}_{\mathbf{Y}}$ can be split into a sum of *m* scalar components, which are precisely the terms used for the IMP function, as given in (2.2.26) or (2.2.27). (We note that $\|\mathbf{Q}_{k-1}\mathbf{a}_k\|^2 = \mathbf{a}_k{}^{\mathrm{H}}\mathbf{Q}_{k-1}{}^{\mathrm{H}}\mathbf{Q}_{k-1}\mathbf{a}_k = \mathbf{a}_k{}^{\mathrm{H}}\mathbf{Q}_{k-1}\mathbf{a}_k$ using projection matrix properties.)

The discussion above is based on the conditional likelihood function with non-coherent (or perhaps partially coherent targets). In the case of totally coherent targets (see Appendix 2C) we replace the *n*×*n* matrix $\mathbf{R}_{\mathbf{Y}}$ (or $\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$) by the *n*×*n* rank one matrix $\overline{\mathbf{y}}\overline{\mathbf{y}}^{\mathrm{H}}$ (where $\overline{\mathbf{y}} = \sum_k \mathbf{y}_k\big/p$) which allows the numerators in (2.2.40) and (2.2.41) to be expressed more simply (as $\left|\overline{\mathbf{y}}^{\mathrm{H}}\mathbf{b}_k\right|^2$ in (2.2.40)) but otherwise does not change the argument or the result.

APPENDIX 2A. LIKELIHOOD FUNCTION

2A.1 Multivariate normal probability function for complex noise

Let **x** be a real vector of n components, all of which are samples from normal distributions, with mean vector **m** and covariance matrix **S**, then **x** has an n-variate normal distribution, with probability density function (p.d.f.) given by

$$f(\mathbf{x}) = \left|2\pi\mathbf{S}\right|^{-1/2} \exp\left(-\tfrac{1}{2}(\mathbf{x}\text{-}\mathbf{m})^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{x}\text{-}\mathbf{m})\right). \tag{2.A.1}$$

(See [30, eq. (2.5.1)], for example. Note |.| indicates the determinant of the enclosed matrix and the superscript T indicates matrix or vector transpose). In the case of noise in n channels having zero mean (**m** = **0**), independent between channels and with equal variance ψ in each channel (so that **S** = ψ**I**) we have

$$f(\mathbf{x}) = \left|2\pi\psi\mathbf{I}\right|^{-1/2} \exp\left(-\tfrac{1}{2\psi}\mathbf{x}^{\mathrm{T}}\mathbf{x}\right) = (2\pi\psi)^{-n/2} \exp\left(-\tfrac{1}{2\psi}\mathbf{x}^{\mathrm{T}}\mathbf{x}\right), \tag{2.A.2}$$

as the determinant of a matrix k**I** is k^n, if **I** is of order n. Now let **z** = **x** + i**y** be a vector of complex noise with real and imaginary parts **x** and **y** respectively, and if we take the components of **y** to have the same statistics (i.e. x_k, y_k ~ N(0,ψ) for k = 1 to n) then **y** has the same form of p.d.f. as **x** in (2.A.2) and the p.d.f. of **z** is given by

$$f(\mathbf{z}) = f(\mathbf{x})f(\mathbf{y}) = (2\pi\psi)^{-n} \exp\left(-\tfrac{1}{2\psi}(\mathbf{x}^{\mathrm{T}}\mathbf{x} + \mathbf{y}^{\mathrm{T}}\mathbf{y})\right)$$

$$= (2\pi\psi)^{-n} \exp\left(-\tfrac{1}{2\psi}\mathbf{z}^{\mathrm{H}}\mathbf{z}\right). \tag{2.A.3}$$

(The superscript H indicates the Hermitian, or complex conjugate, transpose.) If we define the total noise variance in each channel to be ψ then the variance in each of the real and imaginary parts is ψ/2, rather than ψ, so replacing ψ above by ψ/2 we have, finally

$$f(\mathbf{z}) = (\pi\psi)^{-n} \exp\left(-\tfrac{1}{\psi}(\mathbf{z}^{\mathrm{H}}\mathbf{z})\right), \tag{2.A.4}$$

and this is the basis for the likelihood function that is normally used.

2A.2 Conditional and unconditional maximum likelihood; likelihood function

Equation (2.A.4) is not a likelihood function, but the probability density function for independent, identically (and normally) distributed (i.i.d.) complex noise in n channels, the noise having zero mean and total variance ψ (or ψ/2 for each of the real and imaginary components) in each channel. We now consider an n-channel measurement system containing i.i.d. noise **n** with the statistics above and a system response **y** related

to the system parameters $\boldsymbol{\Theta}$, and waveform data $\mathbf{x}$ by

$$\mathbf{y} = \mathbf{A}(\boldsymbol{\Theta})\mathbf{x} + \mathbf{n}. \tag{2.A.5}$$

(NB $\mathbf{x}$ and $\mathbf{y}$ here are not as defined in §2A.1 above, of course, and are, like $\mathbf{n}$, taken to be complex). We consider $\mathbf{A}$ and $\mathbf{x}$ to be fixed (in the conditional ML case, considered here, and defined below) and $\mathbf{n}$ to be a vector random variable. Clearly $\mathbf{y}$ has the same variance as $\mathbf{n}$, but has a non-zero mean vector $\mathbf{Ax}$. We see that the probability of receiving a set of values $\mathbf{y}$ is the same as having a set of noise values $\mathbf{n}$ such that (2.A.5) is satisfied. The probability density of $\mathbf{n}$ is given by (2.A.4) so replacing $\mathbf{z}$ by $\mathbf{n} = \mathbf{y} - \mathbf{Ax}$, we find that the probability of observing $\mathbf{y}$ when $\mathbf{A}$ and $\mathbf{x}$ are the system values is given by

$$f(\mathbf{y}) = (\pi\psi)^{-n} \exp\left(-\tfrac{1}{\psi}(\mathbf{y}\text{-}\mathbf{A}(\Theta)\mathbf{x})^{\mathrm{H}}(\mathbf{y}\text{-}\mathbf{A}(\Theta)\mathbf{x})\right). \tag{2.A.6}$$

This is the probability of receiving the single vector $\mathbf{y}$. If a series of vectors, $\mathbf{y}_1$ to $\mathbf{y}_p$ is received and the noise samples are independent between all these vectors, then the probability of receiving this set of vectors (which we can group into a matrix $\mathbf{Y} = [\mathbf{y}_1 \ldots \mathbf{y}_p]$) is given by

$$\begin{aligned} f(\mathbf{Y}) &= \prod_{k=1}^{p} f(\mathbf{y}_k) = \prod_{k=1}^{p} (\pi\psi)^{-n} \exp\left(-\tfrac{1}{\psi}(\mathbf{y}_k\text{-}\mathbf{A}(\Theta)\mathbf{x}_k)^{\mathrm{H}}(\mathbf{y}_k\text{-}\mathbf{A}(\Theta)\mathbf{x}_k)\right) \\ &= (\pi\psi)^{-np} \exp\left(\sum_{k=1}^{p}\left(-\tfrac{1}{\psi}(\mathbf{y}_k\text{-}\mathbf{A}(\Theta)\mathbf{x}_k)^{\mathrm{H}}(\mathbf{y}_k\text{-}\mathbf{A}(\Theta)\mathbf{x}_k)\right)\right). \end{aligned} \tag{2.A.7}$$

Here we have taken the general case where the vectors $\mathbf{x}_k$ may differ from one measurement vector, or data frame, to the next. This corresponds to the direction of arrival (DoA) problem with time varying, e.g. stochastic, signals. Each signal waveform is considered as having a set of values which are, in principle, determinable and remain constant over realizations of the random variable $\mathbf{n}$ (for example when finding expectation values). This is known as the deterministic, or conditional, maximum likelihood case. The main alternative approach is to take the signals to be Gaussian random variables (if appropriate) with different values in each ensemble realization. In this case the signal parameters to be found are just their variances (taking zero mean signals). This is known as the unconditional, or stochastic, maximum likelihood case. There is perhaps some benefit from reducing the number of parameters to be found, from p waveform samples for each signal, when p frames of data are taken, to just the single one of variance – in [43] theoretical results for the conditional and unconditional

cases are quoted, for a two signal problem, with UML somewhat better for highly correlated sources and low SNR and low array size n. The Gaussian model may not be appropriate for many waveforms in the case of direction finding of communication signals and certainly is not so in the radar case. Often in the radar application we may only want to take a single data frame ($p = 1$) and in other parameter estimation cases data with low values of p only may be available. Thus here we consider only conditional or deterministic maximum likelihood but we take two cases within this condition – those with coherent or non-coherent signals.

By coherent signals we mean signals that differ, over the whole length of the data sample, only by a constant phase difference and a (real) amplitude factor – or, combining these, by a complex amplitude factor, in fact. Radar echoes from a set of targets are all coherent as they are all versions of the same waveform – the transmitted radar pulse. (This also applies, approximately, to short-delay multipath signals.) However, we should be a little cautious here, as radar targets with relative radial motion will have different Doppler shifts, and, over a long enough observation time, will not remain significantly coherent and should be treated as the incoherent signal case. In the case of using a single frame of data there is no meaning to the coherence of the singly sampled signals and the two models give the same ML function.

(a) Non-coherent case

Equation (2.A.7) represents the probability that a set of measurement vectors $\mathbf{Y}$ will be received if the system parameters are $\boldsymbol{\Theta}$ and $\mathbf{X}$ (where $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ .\ .\ .\ \mathbf{x}_p]$), and f is considered to be a function of $\mathbf{Y}$ with $\boldsymbol{\Theta}$ and $\mathbf{X}$ as constants. If, in fact, $\boldsymbol{\Theta}$ and $\mathbf{X}$ are not known, but $\mathbf{Y}$ is known, being the observed values of p received vectors, then we can now regard the function as having arguments $\boldsymbol{\Theta}$ and $\mathbf{X}$, with $\mathbf{Y}$ as constants. In this case we consider f to be the *likelihood* L that $\boldsymbol{\Theta}$ and $\mathbf{X}$ (and ψ) are the system parameters, given that $\mathbf{Y}$ has been received. Thus we write

$$L(\psi,\boldsymbol{\Theta},\mathbf{X};\mathbf{Y}) = (\pi\psi)^{-np}\exp\left(-\tfrac{1}{\psi}\sum_{k=1}^{p}(\mathbf{y}_k-\mathbf{A}(\boldsymbol{\Theta})\mathbf{x}_k)^{\mathrm{H}}(\mathbf{y}_k-\mathbf{A}(\boldsymbol{\Theta})\mathbf{x}_k)\right) \qquad (2.A.8)$$

where $L(\psi,\boldsymbol{\Theta},\mathbf{X};\mathbf{Y})$ reads as 'L is the likelihood that the system parameters are ψ, $\boldsymbol{\Theta}$ and $\mathbf{X}$, given the data $\mathbf{Y}$', where the receiver noise level ψ is also considered unknown.

(b) Coherent case

If we take the radar case, with fixed targets, then the waveform values are indeed

constant (from one data frame to the next) and proportional to the reflection coefficients of the targets. Let these, referred to the receiver channel inputs, be given by **s**, then we replace all the $\mathbf{x}_k$ vectors by **s** to give

$$L(\psi,\boldsymbol{\Theta},\mathbf{s};\mathbf{Y}) = (\pi\psi)^{-np}\exp\left(-\tfrac{1}{\psi}\sum_{k=1}^{p}(\mathbf{y}_k-\mathbf{A}(\boldsymbol{\Theta})\mathbf{s})^{\mathrm{H}}(\mathbf{y}_k-\mathbf{A}(\boldsymbol{\Theta})\mathbf{s})\right). \qquad (2.A.9)$$

In this case we consider f to be the *likelihood* L that $\boldsymbol{\Theta}$ and **s** (and ψ) are the system parameters, given that **Y** has been received.

We can simplify the summation term in (2.A.9). Let

$$S = \sum_{k=1}^{p}\left(\mathbf{y}_k\text{-}\mathbf{A}(\boldsymbol{\Theta})\mathbf{s}\right)^{\mathrm{H}}\left(\mathbf{y}_k\text{-}\mathbf{A}(\boldsymbol{\Theta})\mathbf{s}\right) \qquad (2.A.10)$$

then
$$S = \sum_{k=1}^{p}\left(\|\mathbf{y}_k\|^2 - \mathbf{y}_k{}^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})\mathbf{s} - \mathbf{s}^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})^{\mathrm{H}}\mathbf{y}_k + \mathbf{s}^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})\mathbf{s}\right)$$

$$= p\left(\overline{\|\mathbf{y}\|^2} - \overline{\mathbf{y}}^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})\mathbf{s} - \mathbf{s}^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})^{\mathrm{H}}\overline{\mathbf{y}} + \mathbf{s}^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})\mathbf{s}\right) \qquad (2.A.11)$$

where
$$p\overline{\|\mathbf{y}\|^2} = \sum_{k=1}^{p}\|\mathbf{y}_k\|^2 \quad \text{and} \quad p\overline{\mathbf{y}} = \sum_{k=1}^{p}\mathbf{y}_k\,, \qquad (2.A.12)$$

and the overbar indicates mean values.

Rearranging (2.A.11),

$$S = p\left(\overline{\|\mathbf{y}\|^2} - \|\overline{\mathbf{y}}\|^2 + \|\overline{\mathbf{y}}\|^2 - \overline{\mathbf{y}}^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})\mathbf{s} - \mathbf{s}^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})^{\mathrm{H}}\overline{\mathbf{y}} + \mathbf{s}^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})^{\mathrm{H}}\mathbf{A}(\boldsymbol{\Theta})\mathbf{s}\right)$$

$$= p\left(\overline{\|\mathbf{y}\|^2} - \|\overline{\mathbf{y}}\|^2 + (\overline{\mathbf{y}} - \mathbf{A}(\boldsymbol{\Theta})\mathbf{s})^{\mathrm{H}}(\overline{\mathbf{y}} - \mathbf{A}(\boldsymbol{\Theta})\mathbf{s})\right) \qquad (2.A.13)$$

$$= p\left(nv(\mathbf{Y}) + \left(\overline{\mathbf{y}}\text{-}\mathbf{A}(\boldsymbol{\Theta})\mathbf{s}\right)^{\mathrm{H}}\left(\overline{\mathbf{y}}\text{-}\mathbf{A}(\boldsymbol{\Theta})\mathbf{s}\right)\right) \qquad (2.A.14)$$

where v is the mean sample variance over the n channels and is given by

$$v(\mathbf{Y}) = \left(\overline{\|\mathbf{y}\|^2} - \|\overline{\mathbf{y}}\|^2\right)/n\,. \qquad (2.A.15)$$

[In one channel the sample variance is $v_j = \sum_{k=1}^{p}|\,y_{jk}\,|^2/p - |\,\overline{y_j}\,|^2$ where $\overline{y}_j$ is the sample mean in channel j. Summing over all n channels

$$nv(\mathbf{Y}) = \sum_{j=1}^{n} v_j = \sum_{k=1}^{p}\sum_{j=1}^{n} \left| y_{jk} \right|^2 \Big/ p - \sum_{j=1}^{n} \left| \overline{y}_j \right|^2 = \sum_{k=1}^{p} \left\| \mathbf{y}_k \right\|^2 \Big/ p - \| \overline{\mathbf{y}} \|^2 = \overline{\| \mathbf{y} \|^2} - \| \overline{\mathbf{y}} \|^2 .]$$

Finally, for the radar, coherent target case, from (2.A.9), (2.A.10) and (2.A.14)

$$L(\psi, \boldsymbol{\Theta}, \mathbf{s}; \mathbf{Y}) = (\pi\psi)^{-np} \exp\left(-\tfrac{p}{\psi} \left(nv(\mathbf{Y}) + \left(\overline{\mathbf{y}} \text{-} \mathbf{A}(\boldsymbol{\Theta})\mathbf{s} \right)^{\mathrm{H}} \left(\overline{\mathbf{y}} \text{-} \mathbf{A}(\boldsymbol{\Theta})\mathbf{s} \right) \right) \right). \qquad (2.A.16)$$

2A.3 Concentrated Maximum Likelihood Functions

(a) Non-coherent signals

As the logarithm function is monotonic, with positive slope, it follows that the maximum values of the likelihood function and of the log of the likelihood function over any given range of values of the arguments occur at the same values. It is often convenient to work with the log likelihood function, given by $l = \log(L)$, so that the quadratic term involving the parameters $\boldsymbol{\Theta}$ and the sampled waveforms is brought out of the exponential. From (2.A.8) we have, taking the logarithm,

$$l(\psi, \boldsymbol{\Theta}, \mathbf{X}; \mathbf{Y}) = -np \log(\pi\psi) - \tfrac{1}{\psi} \sum_{k=1}^{p} (\mathbf{y}_k - \mathbf{A}(\boldsymbol{\Theta})\mathbf{x}_k)^{\mathrm{H}} (\mathbf{y}_k - \mathbf{A}(\boldsymbol{\Theta})\mathbf{x}_k) \qquad (2.A.17)$$

and the most likely estimates of the parameters, given the data $\mathbf{Y}$, are found as the parameter values which maximize l. We begin by maximizing with respect to $\mathbf{X}$, or rather with respect to one of its component vectors $\mathbf{x}_k$. (We consider all the components of $\mathbf{X}$ to be independent complex variables.) Maximizing the real valued function l with respect to the complex vector $\mathbf{x}_k$ requires setting the gradient of l with respect to either $\mathbf{x}_k$ or $\mathbf{x}_k^*$ to zero (see [4] for example) i.e., omitting the indication of dependence on $\boldsymbol{\Theta}$ for the moment,

$$\psi \frac{\partial l}{\partial \mathbf{x}_k^*} = -\mathbf{A}^{\mathrm{H}} \left(\mathbf{y}_k - \mathbf{A}\hat{\mathbf{x}}_k \right) = \mathbf{0}$$

or

$$\hat{\mathbf{x}}_k = \left(\mathbf{A}^{\mathrm{H}} \mathbf{A} \right)^{-1} \mathbf{A}^{\mathrm{H}} \mathbf{y}_k , \qquad (2.A.18)$$

where the caret over a variable indicates the value that maximizes the likelihood with respect to that variable and $\partial/\partial\mathbf{x} = \left[\partial/\partial x_1 \quad \partial/\partial x_2 \quad . \; . \; . \quad \partial/\partial x_n \right]^{\mathrm{T}}$. (For complex variables, as the x_k are here, the partial derivative, which treats x_k^* as an independent variable, is defined in [4]).

Substituting the maximizing values $\hat{\mathbf{x}}_k$ for $\mathbf{x}_k$, we see that the factor $\mathbf{y}_k - \mathbf{A}\mathbf{x}_k$ becomes

$$\mathbf{y}_k - \mathbf{A}\hat{\mathbf{x}}_k = \mathbf{y}_k - \mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{A})\mathbf{A}^{\mathrm{H}}\mathbf{y}_k = (\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{y}_k = \mathbf{Q}_{\mathbf{A}}\mathbf{y}_k, \tag{2.A.19}$$

where $\mathbf{P}_{\mathbf{A}}$ projects into the column space of $\mathbf{A}$ and $\mathbf{Q}_{\mathbf{A}}$ projects orthogonally to this, into the null space of $\mathbf{A}$. Substituting (2.A.18) and (2.A.19) into (2.A.17), and also using the relations

$$\mathbf{Q}^{\mathrm{H}}\mathbf{Q} = \mathbf{Q}^{\mathbf{2}} = \mathbf{Q}, \tag{2.A.20}$$

which hold for all projection matrices, we have, for the log likelihood function maximized with respect to the components of $\mathbf{X}$,

$$l(\psi, \mathbf{\Theta}, \hat{\mathbf{X}}; \mathbf{Y}) = -np\log(\pi\psi) - \tfrac{1}{\psi}\sum_{k=1}^{p} \mathbf{y}_k{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}(\mathbf{\Theta})}\mathbf{y}_k . \tag{2.A.21}$$

Now for any two matrices $\mathbf{M}$ and $\mathbf{N}$ whose product is square (i.e. such that the size of one is the same as the transpose of the other; or $\mathbf{M} \in \mathbb{C}^{p\times q}, \mathbf{N} \in \mathbb{C}^{q\times p}$) we have the relation

$$\mathrm{tr}(\mathbf{MN}) = \mathrm{tr}(\mathbf{NM}) \tag{2.A.22}$$

where tr(.) indicates the trace of the matrix argument. With $\mathbf{y}_k{}^{\mathrm{H}}$ for $\mathbf{M}$ and $\mathbf{Q}_{\mathbf{A}}\mathbf{y}_k$ for $\mathbf{N}$ (with dimensions $1\times n$ and $n\times 1$ respectively) the summation in (2.A.21) becomes

$$\begin{aligned} &\sum_{k=1}^{p} \mathbf{y}_k{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}(\mathbf{\Theta})}\mathbf{y}_k = \sum_{k=1}^{p} \mathrm{tr}(\mathbf{y}_k{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}(\mathbf{\Theta})}\mathbf{y}_k) = \sum_{k=1}^{p} \mathrm{tr}(\mathbf{Q}_{\mathbf{A}(\mathbf{\Theta})}\mathbf{y}_k\mathbf{y}_k{}^{\mathrm{H}}) \\ &= \mathrm{tr}(\mathbf{Q}_{\mathbf{A}(\mathbf{\Theta})}\sum_{k=1}^{p} \mathbf{y}_k\mathbf{y}_k{}^{\mathrm{H}}) = \mathrm{tr}(\mathbf{Q}_{\mathbf{A}(\mathbf{\Theta})}\mathbf{R}_{\mathbf{Y}}) \end{aligned} \tag{2.A.23}$$

where $\mathbf{R}_{\mathbf{Y}}$ is the estimated covariance matrix of $\mathbf{Y}$ and is given by

$$\mathbf{R}_{\mathbf{Y}} = \sum_{k=1}^{p} \mathbf{y}_k\mathbf{y}_k{}^{\mathrm{H}} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_p \end{bmatrix}\begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_p \end{bmatrix}^{\mathrm{H}} = \mathbf{Y}\mathbf{Y}^{\mathrm{H}}. \tag{2.A.24}$$

[More strictly we should define $\mathbf{R}_{\mathbf{Y}}$ as $\mathbf{Y}\mathbf{Y}^{\mathrm{H}}/p$, but omitting this scaling factor of p makes no difference to the problem of finding the peak of the likelihood function. In (2.A.23) we have used the linearity of the trace function: $\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$ so that $\mathrm{tr}(\mathbf{QA}) + \mathrm{tr}(\mathbf{QB}) = \mathrm{tr}(\mathbf{Q}(\mathbf{A} + \mathbf{B}))$ and also the fact that, for a scalar z, such as $\mathbf{y}^{\mathrm{H}}\mathbf{Q}\mathbf{y}$, $\mathrm{tr}(z) = z$.] The log likelihood function, maximized with respect to all the elements of $\mathbf{X}$ is now given, from (2.A.21) and (2.A.23) by

$$l(\psi, \mathbf{\Theta}, \hat{\mathbf{X}}; \mathbf{Y}) = -np(\log\pi + \log\psi) - \tfrac{1}{\psi}\mathrm{tr}(\mathbf{Q}_{\mathbf{A}(\mathbf{\Theta})}\mathbf{R}_{\mathbf{Y}}) \tag{2.A.25}$$

We now maximize the log likelihood function with respect to the noise variance, ψ. Differentiating,

$$\frac{\partial l}{\partial \psi} = -\frac{np}{\psi} + \frac{S}{\psi^2}$$

where S is the trace quantity. The most likely value, $\hat{\psi}$, is that for which this slope is zero, giving a stationary point, and so we have $\hat{\psi} = S/np$. Thus the likelihood function, maximized with respect to $\mathbf{X}$ and ψ is now

$$l(\hat{\psi}, \boldsymbol{\Theta}, \hat{\mathbf{X}}; \mathbf{Y}) = -np \log \pi - np \log \left(\frac{\mathrm{tr}(\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Theta})} \mathbf{R}_{\mathbf{Y}})}{np} \right) - np \,. \qquad (2.A.26)$$

The only part of the expression on the right hand side that depends on the parameters $\boldsymbol{\Theta}$ is the trace factor. This is easily seen to be positive:

$$\mathrm{tr}(\mathbf{Q}\mathbf{R}_{\mathbf{Y}}) = \mathrm{tr}(\mathbf{Q}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}) = \mathrm{tr}(\mathbf{Y}^{\mathrm{H}}\mathbf{Q}\mathbf{Y}) = \mathrm{tr}(\mathbf{Y}^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}}\mathbf{Q}\mathbf{Y}) = \mathrm{tr}(\mathbf{Z}^{\mathrm{H}}\mathbf{Z}) = \sum\nolimits_k \left\| \mathbf{z}_k \right\|^2 > 0$$

if $\mathbf{Z}$ is not $\mathbf{0}$, where $\mathbf{Z} = \mathbf{Q}\mathbf{Y}$ (and $\mathbf{z}_k$ is column k of $\mathbf{Z}$) so l is real, as it should be for a probability. Because of the negative sign we see that l is maximized when

$$F(\boldsymbol{\Theta}) = \mathrm{tr}(\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Theta})} \mathbf{R}_{\mathbf{Y}}) \qquad (2.A.27)$$

is *min*imized. Also, as $\mathbf{Q} = \mathbf{I} - \mathbf{P}$, we have

$$F(\boldsymbol{\Theta}) = \mathrm{tr}((\mathbf{I} - \mathbf{P}_{\mathbf{A}(\boldsymbol{\Theta})}) \mathbf{R}_{\mathbf{Y}}) = \mathrm{tr}(\mathbf{R}_{\mathbf{Y}}) - \mathrm{tr}(\mathbf{P}_{\mathbf{A}(\boldsymbol{\Theta})} \mathbf{R}_{\mathbf{Y}})$$

so this is when

$$\tilde{F}(\boldsymbol{\Theta}) = \mathrm{tr}(\mathbf{P}_{\mathbf{A}(\boldsymbol{\Theta})} \mathbf{R}_{\mathbf{Y}}) \qquad (2.A.28)$$

is *max*imized. (As the contribution tr($\mathbf{R}_{\mathbf{Y}}$) is independent of $\boldsymbol{\Theta}$ it can be ignored.) Thus the likelihood function to be maximized over all $\boldsymbol{\Theta}$ can be reduced to the simple form $-\mathrm{tr}(\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Theta})} \mathbf{R}_{\mathbf{Y}})$ or $\mathrm{tr}(\mathbf{P}_{\mathbf{A}(\boldsymbol{\Theta})} \mathbf{R}_{\mathbf{Y}})$.

(b) Coherent signals.

From (2.A.16) we have, in this case,

$$l(\psi, \boldsymbol{\Theta}, \mathbf{s}; \mathbf{Y}) = -np \log(\pi\psi) - \tfrac{p}{\psi} \left(nv(\mathbf{Y}) + \left(\bar{\mathbf{y}} \text{-} \mathbf{A}(\boldsymbol{\Theta})\mathbf{s} \right)^{\mathrm{H}} \left(\bar{\mathbf{y}} \text{-} \mathbf{A}(\boldsymbol{\Theta})\mathbf{s} \right) \right) \qquad (2.A.29)$$

and, similarly to the non-coherent case, we maximize over the signal values, given by $\mathbf{s}$ in this case. Again we have

$$\frac{\partial l}{\partial \mathbf{s}^*} = -\mathbf{A}^{\mathrm{H}}\left(\bar{\mathbf{y}} - \mathbf{A}\hat{\mathbf{s}}\right) = \mathbf{0}$$

or $$\hat{\mathbf{s}} = \left(\mathbf{A}^{\mathrm{H}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{H}}\bar{\mathbf{y}}\,. \tag{2.A.30}$$

Substituting in (2.A.29) and using (2.A.20) gives

$$l(\psi, \boldsymbol{\Theta}, \hat{\mathbf{s}}; \mathbf{Y}) = -np\log(\pi\psi) - \tfrac{p}{\psi}\left(nv(\mathbf{Y}) + \bar{\mathbf{y}}^H\mathbf{Q}_{\mathbf{A}(\Theta)}\bar{\mathbf{y}}\right), \tag{2.A.31}$$

then maximizing over ψ similarly gives

$$\hat{\psi} = \frac{nv(\mathbf{Y}) + \bar{\mathbf{y}}^H\mathbf{Q}_{\mathbf{A}(\Theta)}\bar{\mathbf{y}}}{n}$$

and substituting in (2.A.31) gives

$$l(\hat{\psi}, \boldsymbol{\Theta}, \hat{\mathbf{s}}; \mathbf{Y}) = -np\log(\pi) - np\log\left(v(\mathbf{Y}) + \bar{\mathbf{y}}^H\mathbf{Q}_{\mathbf{A}(\Theta)}\bar{\mathbf{y}}\big/n\right) - np\,. \tag{2.A.32}$$

This is maximized over $\boldsymbol{\Theta}$ when the term $np\log\left(v(\mathbf{Y}) + \bar{\mathbf{y}}^H\mathbf{Q}_{\mathbf{A}(\Theta)}\bar{\mathbf{y}}\big/n\right)$, which is subtracted, is minimized, and, as $v(\mathbf{Y})$ is constant, this is when

$$F(\boldsymbol{\Theta}) = \bar{\mathbf{y}}^H\mathbf{Q}_{\mathbf{A}(\Theta)}\bar{\mathbf{y}} \tag{2.A.33}$$

is minimized or when

$$\tilde{F}(\boldsymbol{\Theta}) = \bar{\mathbf{y}}^H\mathbf{P}_{\mathbf{A}(\Theta)}\bar{\mathbf{y}} \tag{2.A.34}$$

is maximized. The function F is essentially the basis of that used in the IMP (Incremental Multi-Parameter), AP (Alternating Projection) and DOSE (Direction Of Signal Estimation) algorithms, modified here for the radar, fixed target case.

We note that in the case of a single data frame ($p = 1$) we put $\mathbf{Y} = \mathbf{y}_1$ in the non-coherent case, (eq. (2.A.27)) to give

$$F(\boldsymbol{\Theta}) = \mathrm{tr}(\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Theta})}\mathbf{R}_{\mathbf{Y}}) = \mathrm{tr}(\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Theta})}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}) = \mathrm{tr}(\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Theta})}\mathbf{y}_1\mathbf{y}_1^{\mathrm{H}}) = \mathrm{tr}(\mathbf{y}_1^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Theta})}\mathbf{y}_1) = \mathbf{y}_1^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Theta})}\mathbf{y}_1$$

(as this last term is scalar, a 1×1 matrix). In the coherent case $\bar{\mathbf{y}} = \mathbf{y}_1$, as there is only one frame to average, and we see from (2.A.33) that we have the same result.

APPENDIX 2B: OPTIMUM WEIGHT VECTORS

2B.1 Maximum signal to noise ratio

We take the case where there is only one signal present in *n* channels, with point source vector $\mathbf{a}_0$, and with independent noise in each channel but with equal strength in all channels. Let the sampled signal waveform, of length *p*, be **x** then the signal data matrix is $\mathbf{X} = \mathbf{a}_0\mathbf{x}^{\mathrm{T}}$ where $\mathbf{X} \in \mathbb{C}^{n\times p}$. The noise data matrix is $\mathbf{N} \in \mathbb{C}^{n\times p}$. Let the weights applied across the channels be **w** then the 'beamformed' signal waveform is given by $\mathbf{X}^{\mathrm{T}}\mathbf{w}$ and the noise is $\mathbf{N}^{\mathrm{T}}\mathbf{w}$. The signal to noise ratio as a function of **w** is given by

$$r(\mathbf{w}) = \frac{\left\|\mathbf{X}^{\mathrm{T}}\mathbf{w}\right\|^2}{\left\|\mathbf{N}^{\mathrm{T}}\mathbf{w}\right\|^2} = \frac{\left\|\mathbf{X}^{\mathrm{H}}\mathbf{w}*\right\|^2}{\left\|\mathbf{N}^{\mathrm{H}}\mathbf{w}*\right\|^2} = \frac{\mathbf{w}^{\mathbf{T}}\mathbf{X}\mathbf{X}^{\mathrm{H}}\mathbf{w}*}{\mathbf{w}^{\mathbf{T}}\mathbf{N}\mathbf{N}^{\mathrm{H}}\mathbf{w}*} = \frac{\mathbf{w}^{\mathbf{T}}\mathbf{a}_0\mathbf{x}^{\mathrm{T}}\mathbf{x}*\mathbf{a}_0{}^{\mathrm{H}}\mathbf{w}*}{\mathbf{w}^{\mathbf{T}}\mathbf{N}\mathbf{N}^{\mathrm{H}}\mathbf{w}*}. \quad (2.B.1)$$

This is the signal to noise ratio with this particular set of signal and noise data, but we now approximate the estimated noise covariance matrix $\mathbf{N}\mathbf{N}^{\mathrm{H}}$ by the actual covariance matrix, taken to be $\psi\mathbf{I}$ as the noise is independent between channels and of the same variance ψ in all channels. Also $\mathbf{x}^{\mathrm{T}}\mathbf{x}* = \left\|\mathbf{x}\right\|^2$ and we can put $\left\|\mathbf{x}\right\|^2 / \psi = r_0$, a basic signal to noise ratio. Thus we now have

$$r(\mathbf{w}) = r_0 \frac{\mathbf{w}^{\mathbf{T}}\mathbf{a}_0\mathbf{a}_0{}^{\mathrm{H}}\mathbf{w}*}{\mathbf{w}^{\mathbf{T}}\mathbf{w}*}. \quad (2.B.2)$$

To find the maximum signal to noise ratio (SNR) we set the gradient of *r* with respect to **w** to zero, using partial complex differentiation [4]. The gradient is zero when

$$\nu\frac{\partial\delta}{\partial\mathbf{w}} = \delta\frac{\partial\nu}{\partial\mathbf{w}}$$

where ν and δ are the numerator and denominator in (2.B.2) respectively. This gives

$$(\mathbf{w}^{\mathbf{T}}\mathbf{a}_0\mathbf{a}_0{}^{\mathrm{H}}\mathbf{w}*)\mathbf{w}* = (\mathbf{w}^{\mathbf{T}}\mathbf{w}*)\mathbf{a}_0\mathbf{a}_0{}^{\mathrm{H}}\mathbf{w}*.$$

Cancelling the scalar factor $\mathbf{a}_0{}^{\mathrm{H}}\mathbf{w}*$ we have

$$\mathbf{w}* = \frac{\mathbf{w}^{\mathbf{T}}\mathbf{w}*}{\mathbf{w}^{\mathbf{T}}\mathbf{a}_0}\mathbf{a}_0 = k\mathbf{a}_0 \quad (2.B.3)$$

where *k* is a scalar factor. This can be ignored as rescaling **w** makes no difference to the SNR, as this will scale both noise and signal powers equally. Thus we see that setting **w** to $\mathbf{a}_0*$ maximizes the SNR, for the case of uniform noise. (If we put, more generally,

$\mathbf{NN}^{\mathrm{H}} = \mathbf{R}_n$, the noise covariance matrix then we find $\mathbf{w}^* = \mathbf{R}_n^{-1}\mathbf{a}_0$.)

An alternative approach is to minimize the output noise subject to maintaining a fixed response to the signal, using the method of Lagrange undetermined multipliers. In this case the noise power is $\mathbf{w}^{\mathrm{T}}\mathbf{NN}^{\mathrm{H}}\mathbf{w}^*$ or $\psi\mathbf{w}^{\mathrm{T}}\mathbf{w}^*$, replacing $\mathbf{NN}^{\mathrm{H}}$ by $\psi\mathbf{I}$, and the condition on the signal is that $\mathbf{w}^{\mathrm{T}}\mathbf{a}_0$ should be constant, K. Thus the cost function to be minimized is given by

$$H(\mathbf{w}, \mathbf{w}^*) = \psi\mathbf{w}^{\mathrm{T}}\mathbf{w}^* + \lambda(K - \mathbf{w}^{\mathrm{T}}\mathbf{a}_0) + \lambda^*(K^* - \mathbf{w}^{\mathrm{H}}\mathbf{a}_0^*) \tag{2.B.4}$$

where the last term is included to keep the cost function real. Differentiating with respect to $\mathbf{w}$ gives

$$\frac{\partial H(\mathbf{w}, \mathbf{w}^*)}{\partial \mathbf{w}} = \psi\mathbf{w}^* - \lambda\mathbf{a}_0, \tag{2.B.5}$$

which is zero when $\mathbf{w}^*$ is proportional to $\mathbf{a}_0$; the same result.

2B.2 Maximum signal to noise ratio with nulls

The second method of §A2.1 above can be extended to include a number of null points in the response by adding the conditions $\mathbf{w}^{\mathrm{T}}\mathbf{a}_k = 0$ (k = 1 to m) where the null PSVs are $\mathbf{a}_1$, $\mathbf{a}_2$, . . . , $\mathbf{a}_m$. Then the cost function becomes an extended form of that in (2.B.4) above:

$$\begin{aligned} H(\mathbf{w}, \mathbf{w}^*) &= \psi\mathbf{w}^{\mathrm{T}}\mathbf{w}^* + \lambda_0(K - \mathbf{w}^{\mathrm{T}}\mathbf{a}_0) + \lambda_0^*(K^* - \mathbf{w}^{\mathrm{H}}\mathbf{a}_0^*) - (\lambda_1\mathbf{w}^{\mathrm{T}}\mathbf{a}_1 + \lambda_1^*\mathbf{w}^{\mathrm{H}}\mathbf{a}_1^*) \\ &\quad - (\lambda_2\mathbf{w}^{\mathrm{T}}\mathbf{a}_2 + \lambda_2^*\mathbf{w}^{\mathrm{H}}\mathbf{a}_2^*) - \ \ldots\ - (\lambda_m\mathbf{w}^{\mathrm{T}}\mathbf{a}_m + \lambda_m^*\mathbf{w}^{\mathrm{H}}\mathbf{a}_m^*) \\ &= \psi\mathbf{w}^{\mathrm{T}}\mathbf{w}^* + \lambda_0(K - \mathbf{w}^{\mathrm{T}}\mathbf{a}_0) - \mathbf{w}^{\mathrm{T}}\mathbf{A}\boldsymbol{\lambda} + \lambda_0^*(K^* - \mathbf{w}^{\mathrm{H}}\mathbf{a}_0^*) - \mathbf{w}^{\mathrm{H}}\mathbf{A}^*\boldsymbol{\lambda}^*, \end{aligned} \tag{2.B.6}$$

where $\boldsymbol{\lambda} = [\lambda_1\ \lambda_2\ \ldots\ \lambda_m]^{\mathrm{T}}$ and $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_m]$. Differentiating with respect to $\mathbf{w}$,

$$\frac{\partial H(\mathbf{w}, \mathbf{w}^*)}{\partial \mathbf{w}} = \psi\mathbf{w}^* - \mathbf{a}_0\lambda_0 - \mathbf{A}\boldsymbol{\lambda}$$

so that the optimizing weight $\mathbf{w}_0$ is given by

$$\psi\mathbf{w}_0^* = \mathbf{a}_0\lambda_0 + \mathbf{A}\boldsymbol{\lambda} = \begin{bmatrix}\mathbf{a}_0 & \mathbf{A}\end{bmatrix}\begin{bmatrix}\lambda_0 \\ \boldsymbol{\lambda}\end{bmatrix}. \tag{2.B.7}$$

To find the multipliers λ_0 and $\boldsymbol{\lambda}$ we use the conditions $\mathbf{w}^{\mathrm{T}}\mathbf{a}_0 = K$ and $\mathbf{w}^{\mathrm{T}}\mathbf{A} = \mathbf{0}$, which must also be satisfied by $\mathbf{w}_0$, or

$$\mathbf{w}_0^{\mathrm{T}}\begin{bmatrix}\mathbf{a}_0 & \mathbf{A}\end{bmatrix} = \begin{bmatrix}K & \mathbf{0}^{\mathrm{T}}\end{bmatrix} \tag{2.B.8}$$

and combining (2.B.7) and (2.B.8) in the form (taking the conjugate transpose of each side)

$$\begin{bmatrix} \mathbf{a}_0{}^{\mathrm{H}} \\ \mathbf{A}^{\mathrm{H}} \end{bmatrix} \mathbf{w}_0{}^* = \begin{bmatrix} K^* \\ \mathbf{0} \end{bmatrix}$$

we have

$$\begin{bmatrix} \mathbf{a}_0{}^{\mathrm{H}} \\ \mathbf{A}^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \mathbf{a}_0 & \mathbf{A} \end{bmatrix} \begin{bmatrix} \lambda_0 \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \psi K^* \\ \mathbf{0} \end{bmatrix}.$$

Multiplying the two matrices containing PSVs we have, for the Lagrange multipliers

$$\begin{bmatrix} \lambda_0 \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_0{}^{\mathrm{H}}\mathbf{a}_0 & \mathbf{a}_0{}^{\mathrm{H}}\mathbf{A} \\ \mathbf{A}^{\mathrm{H}}\mathbf{a}_0 & \mathbf{A}^{\mathrm{H}}\mathbf{A} \end{bmatrix}^{-1} \begin{bmatrix} \psi K^* \\ \mathbf{0} \end{bmatrix}. \tag{2.B.9}$$

We now need the inverse of a partitioned square matrix, which is of the form

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{Z} & -\mathbf{Z}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{Z} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{Z}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix} \tag{2.B.10}$$

where $\mathbf{A}$ and $\mathbf{D}$ are square and non-singular and $\mathbf{Z}$ is given by $(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}$. In this case we have

$$\mathbf{Z}^{-1} = \mathbf{a}_0{}^{\mathrm{H}}\mathbf{a}_0 - \mathbf{a}_0{}^{\mathrm{H}}\mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{a}_0 = \mathbf{a}_0{}^{\mathrm{H}}(\mathbf{I} - \mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}})\mathbf{a}_0 = \mathbf{a}_0{}^{\mathrm{H}}(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{a}_0 = \mathbf{a}_0{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_0 \tag{2.B.11}$$

and this is a scalar factor. Substituting for $\mathbf{C}$ and $\mathbf{D}$ from (2.B.9), $\mathbf{Z}$ from (2.B.11), we have, from (2.B.7), (2.B.9) and (2.B.10),

$$\mathbf{w}_0{}^* = \begin{bmatrix} \mathbf{a}_0 & \mathbf{A} \end{bmatrix} \begin{bmatrix} 1 & ** \\ -(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{a}_0 & ** \end{bmatrix} \begin{bmatrix} K^* \\ \mathbf{0} \end{bmatrix} \Big/ \mathbf{a}_0{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_0$$

$$= K^*(\mathbf{a}_0 - \mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{a}_0) \Big/ \mathbf{a}_0{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_0 = k(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{a}_0 = k\mathbf{Q}_{\mathbf{A}}\mathbf{a}_0 \tag{2.B.12}$$

(where ** indicates values not needed and $k = K^* / \mathbf{a}_0{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_0$, a scalar). Thus we have $\mathbf{w}_0$ proportional to $\mathbf{Q}_{\mathbf{A}^*}\mathbf{a}_0{}^*$ as the weight vector which maximizes the SNR for a signal with PSV $\mathbf{a}_0$ while maintaining nulls to signals with PSVs given by the columns of $\mathbf{A}$.

APPENDIX 2C: MULTIDIMENSIONAL CRAMÉR-RAO BOUND

The CRB matrix, which is the covariance matrix of the parameter estimates, is given by

$$\mathbf{B} = \mathbf{F}^{-1} \tag{2.C.1}$$

where **F** is the Fisher information matrix (FIM), given by

$$\mathbf{F} = \left\langle \mathbf{v}\mathbf{v}^{\mathrm{H}} \right\rangle. \tag{2.C.2}$$

where the angular brackets indicate the expectation value of the expression contained, and **v** is the score vector, given by the derivative of the log likelihood function, l, with respect to all the parameters of interest. For the general case (not assuming the waveforms are coherent, as in the radar case with fixed targets) these are the noise variance, ψ, the mp elements of the signal waveform data, **X** ($\mathbf{X} \in \mathbb{C}^{m \times p}$) and the mq parameters contained in $\boldsymbol{\Theta}$ ($\boldsymbol{\Theta} \in \mathbb{R}^{m \times q}$), where there are q parameters for each of the m sources. Thus the score vector is

$$\mathbf{v} = \left[\frac{\partial l}{\partial \psi} \quad \frac{\partial l}{\partial \mathbf{x}_1} \quad \frac{\partial l}{\partial \mathbf{x}_1^*} \quad \cdots \quad \frac{\partial l}{\partial \mathbf{x}_p} \quad \frac{\partial l}{\partial \mathbf{x}_p^*} \quad \frac{\partial l}{\partial \boldsymbol{\theta}_1} \quad \cdots \quad \frac{\partial l}{\partial \boldsymbol{\theta}_q} \right]^{\mathrm{T}}. \tag{2.C.3}$$

Here $\mathbf{x}_j$ is column j of **X**, and contains the samples of the m signal waveforms at time sample j, and similarly $\boldsymbol{\theta}_k$ is column k of $\boldsymbol{\Theta}$, and is the set of m target parameters in dimension k e.g. azimuth, elevation or range. (**v** is a column vector of $1 + 2mp + mq$ components, and, strictly, all the derivatives with respect to vectors are column vectors and should have an indication of transpose; this is omitted to avoid making the expression too cluttered.)

[As the elements of **X** are complex and each contains two degrees of freedom, we have two parameters for each signal sample and we take these to be the sample value and its conjugate, rather than the real and imaginary parts. This choice is equivalent, as the variables of each complex conjugate pair are independent, in the sense of partial differentiation, as made clear in [4], and we can use complex differentiation, rather than separate real differentiation with respect to the real and imaginary parts. If we particularly wanted the real and imaginary parts as parameters then, putting $\mathbf{x}_1 = \mathbf{u}_1 + i\mathbf{v}_1$, where $\mathbf{u}_1$ and $\mathbf{v}_1$ are real vectors, we could replace $\frac{\partial l}{\partial \mathbf{x}_1}$ and $\frac{\partial l}{\partial \mathbf{x}_1^*}$ by $\frac{\partial l}{\partial \mathbf{u}_1}$ and $\frac{\partial l}{\partial \mathbf{v}_1}$ etc. (This is the approach taken in [42]). We can obtain the derivatives with respect to $\mathbf{x}_1 \ldots \mathbf{x}_p$ and $\mathbf{x}_1^* \ldots \mathbf{x}_p^*$ then use $\frac{\partial l}{\partial \mathbf{u}} = \frac{\partial l}{\partial \mathbf{x}} + \frac{\partial l}{\partial \mathbf{x}^*}$ and $\frac{\partial l}{\partial \mathbf{v}} = i\left(\frac{\partial l}{\partial \mathbf{x}} - \frac{\partial l}{\partial \mathbf{x}^*}\right)$. These expressions follow from $\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right)$ and $\frac{\partial f}{\partial z^*} = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right)$, where $z = x + iy$,

see [4]. However, here we are really concerned with the lower bound on the variances of the parameters *other* than the waveform amplitudes, so it is simplest just to leave the score vector in the form of (2.C.3) above.]

From Appendix 2A, equation (2.A.17) we have, for the log likelihood function,

$$l(\psi,\boldsymbol{\Theta},\mathbf{X};\mathbf{Y}) = -np\log(\pi\psi) - \tfrac{1}{\psi}\sum_{j=1}^{p}(\mathbf{y}_j - \mathbf{A}(\boldsymbol{\Theta})\mathbf{x}_j)^{\mathrm{H}}(\mathbf{y}_j - \mathbf{A}(\boldsymbol{\Theta})\mathbf{x}_j) \quad (2.C.4)$$

where, from (2.A.5), we replace $\mathbf{y}_j - \mathbf{A}(\boldsymbol{\Theta})\mathbf{x}_j$ by $\mathbf{n}_j$ when convenient, i.e. when we are not performing partial differentiation with respect to any of the variables in $\boldsymbol{\Theta}$ or $\mathbf{X}$. Thus we have

$$N = \sum_{j=1}^{p}(\mathbf{y}_j - \mathbf{A}(\boldsymbol{\Theta})\mathbf{x}_j)^{\mathrm{H}}(\mathbf{y}_j - \mathbf{A}(\boldsymbol{\Theta})\mathbf{x}_j) = \sum_{j=1}^{p}\mathbf{n}_j{}^{\mathrm{H}}\mathbf{n}_j \quad (2.C.5)$$

The components of the score vector are then given (omitting the indication of the dependence of $\mathbf{A}$ on $\boldsymbol{\Theta}$ for the moment) by

$$\frac{\partial l}{\partial \psi} = -\frac{np}{\psi} + \frac{N}{\psi^2} = -\frac{np}{\psi} + \frac{\sum_{j=1}^{p}\mathbf{n}_j{}^{\mathrm{H}}\mathbf{n}_j}{\psi^2} \quad (2.C.6)$$

$$\frac{\partial l}{\partial \mathbf{x}_j{}^*} = -\frac{1}{\psi}\frac{\partial N}{\partial \mathbf{x}_j{}^*} = \frac{1}{\psi}\mathbf{A}^{\mathrm{H}}\mathbf{n}_j \text{ and } \frac{\partial l}{\partial \mathbf{x}_j} = \frac{1}{\psi}\mathbf{A}^{\mathrm{T}}\mathbf{n}_j{}^* \quad (2.C.7)$$

$$\frac{\partial l}{\partial \boldsymbol{\theta}_k} = \frac{1}{\psi}\sum_{j=1}^{p}\left(\mathbf{X}_j{}^*\mathbf{D}_k{}^{\mathrm{H}}\mathbf{n}_j + \mathbf{X}_j\mathbf{D}_k{}^{\mathrm{T}}\mathbf{n}_j{}^*\right) \quad (2.C.8)$$

where $\mathbf{X}_j = \mathrm{diag}(\mathbf{x}_j)$, i.e. $\mathbf{X}_j$ is an $m\times m$ matrix with its non-zero elements on the principal diagonal, formed from the elements of $\mathbf{x}_j$, and $\mathbf{D}_k$ is an $n\times m$ matrix given by

$$\mathbf{D}_k = \left[\frac{\partial \mathbf{a}_1}{\partial \theta_{1k}} \quad \frac{\partial \mathbf{a}_2}{\partial \theta_{2k}} \quad \ldots \quad \frac{\partial \mathbf{a}_m}{\partial \theta_{mk}}\right]. \qquad (k = 1 \text{ to } q) \quad (2.C.9)$$

[To form (2.C.8) we note that $\frac{\partial l}{\partial \theta_{\mu k}} = -\frac{1}{\psi}\frac{\partial N}{\partial \theta_{\mu k}} = \frac{1}{\psi}\sum_{j=1}^{p}\left(\frac{\partial(\mathbf{x}_j{}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}})}{\partial \theta_{\mu k}}\mathbf{n}_j + \mathbf{n}_j{}^{\mathrm{H}}\frac{\partial(\mathbf{A}\mathbf{x}_j)}{\partial \theta_{\mu k}}\right)$

and

$$\frac{\partial(\mathbf{x}_j{}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}})}{\partial \theta_{\mu k}} = \frac{\partial}{\partial \theta_{\mu k}}(x_{1j}{}^*\mathbf{a}_1{}^{\mathrm{H}} + x_{2j}{}^*\mathbf{a}_2{}^{\mathrm{H}} + \ldots + x_{mj}{}^*\mathbf{a}_m{}^{\mathrm{H}}) = x_{\mu j}{}^*\frac{\partial \mathbf{a}_\mu{}^{\mathrm{H}}}{\partial \theta_{\mu k}} = x_{\mu j}{}^*(\mathbf{d}_{\mu k})^{\mathrm{H}}$$

where $\mathbf{d}_{\mu k}$ is column μ of $\mathbf{D}_k$, whose columns are the derivatives of the m columns of $\mathbf{A}$ with respect to parameter k for the signal represented by each column. Then we have

$$\frac{\partial(\mathbf{x}_j{}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}})}{\partial\boldsymbol{\theta}_k}=\begin{bmatrix} x_{1j}*(\mathbf{d}_{1k})^{\mathrm{H}} \\ x_{2j}*(\mathbf{d}_{2k})^{\mathrm{H}} \\ . \\ . \\ . \\ x_{mj}*(\mathbf{d}_{mk})^{\mathrm{H}} \end{bmatrix}=\mathbf{X}_j*\mathbf{D}_k{}^{\mathrm{H}}.$$

We also note that $\mathbf{n}_j{}^{\mathrm{H}}\mathbf{A}\mathbf{x}_j$ is a scalar and so can be written as its transpose, which is $\mathbf{x}_j{}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{n}_j*$, and $\mathbf{x}_j{}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$ is the conjugate of $\mathbf{x}_j{}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}$. Combining these results leads to (2.C.8).]

From (2.C.2) we see that to form the Fisher information matrix, $\mathbf{F}$, we need the expectation values of all the products of the elements of the score vector, $\mathbf{v}$. Using result R1 from Appendix E of [42], ($\langle\mathbf{n}_j{}^{\mathrm{H}}\mathbf{n}_j\mathbf{n}_h{}^{\mathrm{H}}\mathbf{n}_h\rangle = n(n+\delta_{hj})\psi^2$) on the expectation values of Gaussian noise products (we have already assumed the noise is distributed as a complex Gaussian variable in forming the likelihood function) we obtain

$$\left\langle\left(\frac{\partial l}{\partial\psi}\right)^2\right\rangle=\left\langle\left(\frac{np}{\psi}\right)^2-2\frac{np}{\psi}\frac{\sum_{j=1}^{p}\mathbf{n}_j{}^{\mathrm{H}}\mathbf{n}_j}{\psi^2}+\frac{\sum_{j=1}^{p}\sum_{h=1}^{p}\mathbf{n}_j{}^{\mathrm{H}}\mathbf{n}_j\mathbf{n}_h{}^{\mathrm{H}}\mathbf{n}_h}{\psi^4}\right\rangle$$

$$=\left(\frac{np}{\psi}\right)^2-2\frac{np}{\psi}\frac{pn\psi}{\psi^2}+\frac{p(p-1)n^2\psi^2+pn(n+1)\psi^2}{\psi^4}=\frac{np}{\psi^2}. \quad (2.C.10)$$

Using R2 from Appendix E of [42], ($\langle\mathbf{n}_h{}^{\mathrm{H}}\mathbf{n}_h\mathbf{n}_j{}^{\mathrm{T}}\rangle=\mathbf{0}$), we have

$$\left\langle\left(\frac{\partial l}{\partial\psi}\frac{\partial l}{\partial\mathbf{x}_j*}\right)^{\mathrm{T}}\right\rangle=\left\langle\left(-\frac{np}{\psi}+\frac{\sum_{h=1}^{p}\mathbf{n}_h{}^{\mathrm{H}}\mathbf{n}_h}{\psi^2}\right)\frac{\mathbf{n}_j{}^{\mathrm{T}}\mathbf{A}*}{\psi}\right\rangle=\mathbf{0}_{1\times m} \quad (2.C.11a)$$

(as $\langle\mathbf{n}_j{}^{\mathrm{T}}\rangle=\mathbf{0}$ for zero mean noise waveforms). Also, as $\partial l/\partial\psi$ is real, we have

$$\left\langle\left(\frac{\partial l}{\partial\psi}\frac{\partial l}{\partial\mathbf{x}_j}\right)^{\mathrm{H}}\right\rangle=\left\langle\left(\frac{\partial l}{\partial\psi}\frac{\partial l}{\partial\mathbf{x}_j*}\right)^{\mathrm{T}}\right\rangle=\mathbf{0}_{1\times m}. \quad (2.C.11b)$$

Similarly, with $\mathbf{X}_j*\mathbf{D}_k{}^{\mathrm{H}}$ and $\mathbf{X}_j\mathbf{D}_k{}^{\mathrm{T}}$ replacing $\mathbf{A}^{\mathrm{H}}$ and $\mathbf{A}^{\mathrm{T}}$, we find

$$\left\langle \left(\frac{\partial l}{\partial \psi} \frac{\partial l}{\partial \boldsymbol{\theta}_k} \right) \right\rangle = \mathbf{0}_{m\times 1} . \tag{2.C.12}$$

Also

$$\left\langle \frac{\partial l}{\partial \mathbf{x}_h} \frac{\partial l}{\partial \mathbf{x}_j}^{\mathrm{H}} \right\rangle = \left\langle \frac{1}{\psi^2} \mathbf{A}^{\mathrm{T}} \mathbf{n}_h * \mathbf{n}_j{}^{\mathrm{T}} \mathbf{A} * \right\rangle = \frac{1}{\psi} \mathbf{A}^{\mathrm{T}} \mathbf{A} * \delta_{hj} \tag{2.C.13a}$$

and

$$\left\langle \frac{\partial l}{\partial \mathbf{x}_h} * \frac{\partial l}{\partial \mathbf{x}_j} *^{\mathrm{H}} \right\rangle = \left\langle \frac{1}{\psi^2} \mathbf{A}^{\mathrm{H}} \mathbf{n}_h \mathbf{n}_j \mathbf{A} \right\rangle = \frac{1}{\psi} \mathbf{A}^{\mathrm{H}} \mathbf{A} \delta_{hj} \tag{2.C.13b}$$

using $\langle n_{\mu j} * n_{\nu j} \rangle = \psi \delta_{\mu\nu}$, so that $\langle \mathbf{n}_j * \mathbf{n}_j{}^{\mathrm{T}} \rangle = \langle \mathbf{n}_j \boldsymbol{n}_j{}^{H} \rangle = \psi \mathbf{I}$, and $\langle \mathbf{n}_h \mathbf{n}_j{}^{\mathrm{T}} \rangle = \langle \mathbf{n}_h * \mathbf{n}_j{}^{\mathrm{H}} \rangle = \mathbf{0}$ ($h \neq j$) as the noise in different channels is taken to be uncorrelated. Also with the reasonable assumption that $\langle \mathbf{n}_h \mathbf{n}_j{}^{\mathrm{T}} \rangle = \mathbf{0}$ we have

$$\left\langle \frac{\partial l}{\partial \mathbf{x}_h} * \frac{\partial l}{\partial \mathbf{x}_j}^{\mathrm{H}} \right\rangle = \left\langle \frac{1}{\psi^2} \mathbf{A}^{\mathrm{H}} \mathbf{n}_h \mathbf{n}_j{}^{\mathrm{T}} \mathbf{A} * \right\rangle = \mathbf{0}_{m\times m} = \left\langle \frac{\partial l}{\partial \mathbf{x}_h} \frac{\partial l}{\partial \mathbf{x}_j} *^{\mathrm{H}} \right\rangle . \tag{2.C.14}$$

From (2.C.7) and (2.C.8) (and with the results above on the expectation of noise products)

$$\left\langle \frac{\partial l}{\partial \mathbf{x}_j} \frac{\partial l}{\partial \boldsymbol{\theta}_k}^{\mathrm{H}} \right\rangle = \left\langle \frac{1}{\psi^2} \mathbf{A}^{\mathrm{T}} \mathbf{n}_h * \sum_{j=1}^{p} \left(\mathbf{n}_j{}^{\mathrm{H}} \mathbf{D}_k \mathbf{X}_j{}^{\mathrm{T}} + \mathbf{n}_j{}^{\mathrm{T}} \mathbf{D}_k * \mathbf{X}_j{}^{\mathrm{H}} \right) \right\rangle = \frac{1}{\psi} \mathbf{A}^{\mathrm{T}} \mathbf{D}_k * \mathbf{X}_j{}^{\mathrm{H}} \tag{2.C.15a}$$

and, taking the conjugate (noting that $\partial l / \partial \boldsymbol{\theta}_k$ is real)

$$\left\langle \frac{\partial l}{\partial \mathbf{x}_j} * \frac{\partial l}{\partial \boldsymbol{\theta}_k}^{\mathrm{H}} \right\rangle = \left\langle \frac{1}{\psi^2} \mathbf{A}^{\mathrm{H}} \mathbf{n}_h \sum_{j=1}^{p} \left(\mathbf{n}_j{}^{\mathrm{H}} \mathbf{D}_k \mathbf{X}_j{}^{\mathrm{T}} + \mathbf{n}_j{}^{\mathrm{T}} \mathbf{D}_k * \mathbf{X}_j{}^{\mathrm{H}} \right) \right\rangle = \frac{1}{\psi} \mathbf{A}^{\mathrm{H}} \mathbf{D}_k \mathbf{X}_j{}^{\mathrm{T}} . \tag{2.C.15b}$$

Finally,

$$\left\langle \frac{\partial l}{\partial \boldsymbol{\theta}_h} \frac{\partial l}{\partial \boldsymbol{\theta}_k}^{\mathrm{H}} \right\rangle = \left\langle \frac{1}{\psi^2} \sum_{i=1}^{p} \left(\mathbf{X}_i * \mathbf{D}_h{}^{\mathrm{H}} \mathbf{n}_i + \mathbf{X}_i \mathbf{D}_h{}^{\mathrm{T}} \mathbf{n}_i * \right) \sum_{j=1}^{p} \left(\mathbf{n}_j{}^{\mathrm{H}} \mathbf{D}_k \mathbf{X}_j{}^{\mathrm{T}} + \mathbf{n}_j{}^{\mathrm{T}} \mathbf{D}_k * \mathbf{X}_j{}^{\mathrm{H}} \right) \right\rangle$$

$$= \frac{1}{\psi} \sum_{j=1}^{p} \left(\mathbf{X}_j * \mathbf{D}_h{}^{\mathrm{H}} \mathbf{D}_k \mathbf{X}_j{}^{\mathrm{T}} + \mathbf{X}_j \mathbf{D}_h{}^{\mathrm{T}} \mathbf{D}_k * \mathbf{X}_j{}^{\mathrm{H}} \right) . \tag{2.C.16}$$

As $\mathbf{X}_j$ is diagonal we have $\mathbf{X}^{\mathrm{T}} = \mathbf{X}$ and $\mathbf{X}^{\mathrm{H}} = \mathbf{X}^*$ and we include these results in applying (2.C.15) and (2.C.16) below. From (2.C.2) and the results (2.C.10) to (2.C.16) we

obtain

$$\mathbf{F} = \frac{1}{\psi}\begin{bmatrix} {}^{np}\!/_{\psi} & \mathbf{0}_{1\times mp} & \mathbf{0}_{1\times mp} & \mathbf{0}_{1\times mq} \\ \mathbf{0}_{mp\times 1} & \mathbf{G}^{*} & \mathbf{0}_{mp\times mp} & \mathbf{H}^{*} \\ \mathbf{0}_{mp\times 1} & \mathbf{0}_{mp\times mp} & \mathbf{G} & \mathbf{H} \\ \mathbf{0}_{mq\times 1} & \mathbf{H}^{\mathrm{T}} & \mathbf{H}^{\mathrm{H}} & \mathbf{K}+\mathbf{K}^{\mathbf{H}} \end{bmatrix} \tag{2.C.17}$$

where **G**, **H** and **K** are all of the form of blocks of $m\times m$ submatrices. **G** is $p\times p$ block diagonal, with its $m\times m$ blocks given by $\mathbf{A}^{\mathrm{H}}\mathbf{A}$. (We could put $\mathbf{G} = \mathbf{A}^{\mathbf{H}}\mathbf{A}\square\mathbf{I}_{p}$, where the symbol indicates the Kronecker product, where each element in the second matrix in the product is multiplied by the first matrix.) **H** is a $p\times q$ block matrix, with the jk block being given by $\mathbf{H}_{jk} = \mathbf{A}^{\mathrm{H}}\mathbf{D}_{k}\mathbf{X}_{j}$ and **K** is a $q\times q$ block matrix, with the hk block being given by $\mathbf{K}_{hk} = \sum_{j=1}^{p}\mathbf{X}_{j}{}^{*}\mathbf{D}_{h}{}^{\mathrm{H}}\mathbf{D}_{k}\mathbf{X}_{j}$. We now need the inverse of **F** to obtain the Cramér-Rao bound covariance matrix **B** (see (2.C.1)). First we put **F** in the form

$$\mathbf{F} = \frac{1}{\psi}\begin{bmatrix} {}^{np}\!/_{\psi} & \mathbf{0}_{1\times 2mp} & \mathbf{0}_{1\times mq} \\ \mathbf{0}_{2mp\times 1} & \mathbf{U} & \mathbf{V} \\ \mathbf{0}_{mq\times 1} & \mathbf{V}^{\mathbf{H}} & \mathbf{W} \end{bmatrix} \tag{2.C.18}$$

where $\mathbf{U} = \begin{bmatrix} \mathbf{G}^{*} & \mathbf{0}_{mp\times mp} \\ \mathbf{0}_{mp\times mp} & \mathbf{G} \end{bmatrix}$, $\mathbf{V} = \begin{bmatrix} \mathbf{H}^{*} \\ \mathbf{H} \end{bmatrix}$ and $\mathbf{W} = \mathbf{K}+\mathbf{K}^{*}$. Then the CRB matrix is given by

$$\mathbf{B} = \mathbf{F}^{-1} = \psi\begin{bmatrix} {}^{\psi}\!/_{np} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \mathbf{Y} \\ \mathbf{0} & \mathbf{Y}^{\mathbf{H}} & \mathbf{Z} \end{bmatrix} \tag{2.C.19}$$

with $\begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y}^{\mathbf{H}} & \mathbf{Z} \end{bmatrix} = \begin{bmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}^{\mathbf{H}} & \mathbf{W} \end{bmatrix}^{-1}$.

Now we are not particularly interested in the values of **X**, the covariances of the complex signal amplitude samples, and even less in **Y**, the covariances of the signal amplitudes with the target parameters. We are interested in the covariances of the signal parameters, contained in **Z**, and generally only in the variances of the parameters (rather than the covariances of pairs of parameters) which are on the diagonal of **Z**. The inverse of the **U-V-W** matrix can be found by performing operations on it which reduce it to the identity, while applying the same operations to an identity matrix, to produce the

inverse. We find that

Error! Objects cannot be created from editing field codes.. (2.C.20)

Substituting for **U** and **V** we find

$$\mathbf{V^H U^{-1} V} = \begin{bmatrix} \mathbf{H^T} & \mathbf{H^H} \end{bmatrix} \begin{bmatrix} \mathbf{G}^{*-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G^{-1}} \end{bmatrix} \begin{bmatrix} \mathbf{H}^* \\ \mathbf{H} \end{bmatrix} = \mathbf{H^H G^{-1} H} + (\mathbf{H^H G^{-1} H})^* \quad (2.C.21)$$

and then also substituting for **W** in (2.C.20) we obtain

$$\mathbf{Z}^{-1} = (\mathbf{K} - \mathbf{H^H G^{-1} H}) + (\mathbf{K} - \mathbf{H^H G^{-1} H})^* = 2\mathrm{Re}(\mathbf{K} - \mathbf{H^H G^{-1} H}). \quad (2.C.22)$$

From the definition above of the blocks of **H** we have

$$\mathbf{H} = \begin{bmatrix} \mathbf{A^H D_1 X_1} & \mathbf{A^H D_2 X_1} & \cdots & \mathbf{A^H D}_q \mathbf{X_1} \\ \mathbf{A^H D_1 X}_2 & \mathbf{A^H D_2 X}_2 & \cdots & \mathbf{A^H D}_q \mathbf{X}_2 \\ \vdots & \vdots & & \vdots \\ \mathbf{A^H D_1 X}_p & \mathbf{A^H D_2 X}_p & \cdots & \mathbf{A^H D}_q \mathbf{X}_p \end{bmatrix} = \begin{bmatrix} \mathbf{A^H \Delta_1} \\ \mathbf{A^H \Delta}_2 \\ \vdots \\ \mathbf{A^H \Delta}_p \end{bmatrix} \quad (2.C.23)$$

where

$$\mathbf{\Delta}_j = \begin{bmatrix} \mathbf{D}_1 \mathbf{X}_j & \mathbf{D}_2 \mathbf{X}_j & \cdots & \mathbf{D}_q \mathbf{X}_j \end{bmatrix}. \quad (2.C.24)$$

We note that **H** is $mp \times mq$ or $p \times q$ in blocks of size $m \times m$, and $\mathbf{\Delta}_j$ is $n \times mq$ or $1 \times q$ in blocks (of size $n \times m$). Recalling that **G** is block diagonal, and is of size $p \times p$ in blocks and $\mathbf{G}^{-1}$ is similarly block diagonal with diagonal blocks $(\mathbf{A^H A})^{-1}$, we find

$$\mathbf{H^H G^{-1} H} = \begin{bmatrix} \mathbf{\Delta}_1{}^{\mathrm{H}} \mathbf{A} & \mathbf{\Delta}_2{}^{\mathrm{H}} \mathbf{A} & \cdots & \mathbf{\Delta}_p{}^{\mathrm{H}} \mathbf{A} \end{bmatrix} \begin{bmatrix} (\mathbf{A^H A})^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\mathbf{A^H A})^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (\mathbf{A^H A})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{A^H \Delta_1} \\ \mathbf{A^H \Delta}_2 \\ \vdots \\ \mathbf{A^H \Delta}_p \end{bmatrix}$$

$$= \sum_{j=1}^{p} \mathbf{\Delta}_j{}^{\mathrm{H}} \mathbf{A} (\mathbf{A^H A})^{-1} \mathbf{A^H \Delta}_j = \sum_{j=1}^{p} \mathbf{\Delta}_j{}^{\mathrm{H}} \mathbf{P_A \Delta}_j \quad (2.C.25)$$

where $\mathbf{P_A} = \mathbf{A}(\mathbf{A^H A})^{-1}\mathbf{A^H}$ projects into the column space of **A.** Now, from the definitions

of $\mathbf{K}$ and $\boldsymbol{\Delta}$, we see that $\mathbf{K} = \sum_{j=1}^{p} \boldsymbol{\Delta}_j{}^{\mathrm{H}} \boldsymbol{\Delta}_j$, so we have

$$\mathbf{K} - \mathbf{H}^{\mathrm{H}}\mathbf{G}^{-1}\mathbf{H} = \sum_{j=1}^{p} \boldsymbol{\Delta}_j{}^{\mathrm{H}} \boldsymbol{\Delta}_j - \sum_{j=1}^{p} \boldsymbol{\Delta}_j{}^{\mathrm{H}} \mathbf{P}_{\mathbf{A}} \boldsymbol{\Delta}_j = \sum_{j=1}^{p} \boldsymbol{\Delta}_j{}^{\mathrm{H}} (\mathbf{I} - \mathbf{P}_{\mathbf{A}})_{\mathbf{A}} \boldsymbol{\Delta}_j$$

$$= \sum_{j=1}^{p} \boldsymbol{\Delta}_j{}^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \boldsymbol{\Delta}_j , \tag{2.C.26}$$

where $\mathbf{Q_A} = \mathbf{I} - \mathbf{P_A}$ projects into the null space of $\mathbf{A}$. From (2.C.19), (2.C.22) and (2.C.26) we see that the CRB matrix for the q parameters of each of the m sources (mq parameters in total) is given by

$$\mathbf{B}(\boldsymbol{\Theta}) = \psi \mathbf{Z} = \frac{\psi}{2} \left(\mathrm{Re}(\sum_{j=1}^{p} \boldsymbol{\Delta}_j{}^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \boldsymbol{\Delta}_j) \right)^{-1} . \tag{2.C.27}$$

A problem with this expression is that the mp signal samples appear in it (samples of each of the m waveforms, appearing in the p $\boldsymbol{\Delta}_j$ matrices). From the definition of $\boldsymbol{\Delta}_j$ in (2.C.24) we see that $\boldsymbol{\Delta}_j{}^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \boldsymbol{\Delta}_j$ is of size $q \times q$ in blocks (of size $m \times m$) and so also is $\sum_{j=1}^{p} \boldsymbol{\Delta}_j{}^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \boldsymbol{\Delta}_j$. Now block hk of this matrix is given by

$$\left(\sum_{j=1}^{p} \boldsymbol{\Delta}_j{}^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \boldsymbol{\Delta}_j \right)_{hk} = \sum_{j=1}^{p} \mathbf{X}_j * \mathbf{D}_h{}^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \mathbf{D}_k \mathbf{X}_j = \sum_{j=1}^{p} \mathbf{X}_j * \mathbf{E} \mathbf{X}_j \tag{2.C.28}$$

where we have defined $\mathbf{D}_h{}^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \mathbf{D}_k$ as $\mathbf{E}$ temporarily for clarity. Now we consider the ab element of this $m \times m$ matrix, recalling that $\mathbf{X}_j$ is diagonal, given by diag($\mathbf{x}_j$) where $\mathbf{x}_j$ is column j of the $m \times p$ signal waveform matrix $\mathbf{X}$. This element is given by

$$\left(\sum_{j=1}^{p} \mathbf{X}_j * \mathbf{E} \mathbf{X}_j \right)_{ab} = \sum_{j=1}^{p} x_{aj} * (\mathbf{E})_{ab} x_{bj} = \sum_{j=1}^{p} (\mathbf{E})_{ab} x_{aj} * x_{bj} = (\mathbf{E})_{ab} (p \tilde{\mathbf{S}}^*)_{ab} \tag{2.C.29}$$

where $\tilde{\mathbf{S}}$ is the estimated covariance matrix of the signal waveforms and is given by

$$\tilde{\mathbf{S}} = \mathbf{X}\mathbf{X}^{\mathrm{H}} / p = \sum_{j=1}^{p} \mathbf{x}_j \mathbf{x}_j{}^{\mathrm{H}} \Big/ p . \tag{2.C.30}$$

In the limit of large p, the number of frames taken, $\tilde{\mathbf{S}}$ becomes the actual covariance matrix, $\mathbf{S}$ and in the case of uncorrelated signals $\mathbf{S}$ becomes diagonal, with the signal powers on the diagonal. We note in (2.C.29) that we have the element by element (or Hadamard) product of two matrices, which we denote by $\odot$ so that we now have

$\sum_{j=1}^{p} \mathbf{X}_j * \mathbf{E}\mathbf{X}_j = p\mathbf{E} \odot \tilde{\mathbf{S}}*$ and so

$$\left(\sum_{j=1}^{p} \mathbf{\Delta}_j{}^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \mathbf{\Delta}_j \right)_{hk} = p\mathbf{D}_h{}^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \mathbf{D}_k \odot \tilde{\mathbf{S}}* . \tag{2.C.31}$$

As every block in $\sum_{j=1}^{p} \mathbf{\Delta}_j{}^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \mathbf{\Delta}_j$ forms a Hadamard product with $\tilde{\mathbf{S}}*$ we can put, formally,

$$\mathbf{B}(\mathbf{\Theta}) = \frac{\psi}{2p} \left(\mathrm{Re}[\bar{\mathbf{D}}^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \bar{\mathbf{D}} \odot (\tilde{\mathbf{S}} * \square \mathbf{1}_q \mathbf{1}_q{}^{\mathrm{T}})] \right)^{-1}, \tag{2.C.32}$$

where $\mathbf{1}_q \mathbf{1}_q{}^{\mathrm{T}}$ is a $q \times q$ matrix of ones ($\mathbf{1}_q$ is a vector of q ones) and $\tilde{\mathbf{S}} * \square \mathbf{1}_q \mathbf{1}_q{}^{\mathrm{T}}$ is a $mq \times mq$ matrix with $\tilde{\mathbf{S}}*$ replacing each unit element. We have also put $\bar{\mathbf{D}} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_2 & . & . & . & \mathbf{D}_q \end{bmatrix}$, which is of size $n \times mq$. (NB: We note that, as $\tilde{\mathbf{S}}$ is Hermitian, we can put $\tilde{\mathbf{S}}^{\mathrm{T}}$ instead of $\tilde{\mathbf{S}}*$ if preferred.)

APPENDIX 2D: SPECIAL CASES OF THE CRB

The expressions (2.C.27) and (2.C.32) given in Appendix C for the Cramér-Rao bound for the source parameter estimates (not including the signal waveform sample values or the receiver noise variance value) in the general case of m sources with q parameters each, are difficult to take further in a general theoretical form. For example, we cannot obtain simple expressions for the diagonal values, which, being the variances of the parameter estimates, are the main values of interest, because these are only found after performing an inverse of a matrix of substantial size - $mq \times mq$. If the various system parameters in this matrix (ψ, p, $\mathbf{A}$, $\mathbf{D}_1$ to $\mathbf{D}_q$ and $\tilde{\mathbf{S}}$, or $\mathbf{S}$) are known or estimated then these values could be substituted and the CRB matrix evaluated numerically quite easily, for a given case, but this is not a general solution. However, we can obtain some useful forms for the particular cases below.

2D.1 One source, one parameter

In this case we have $m = 1$ and $q = 1$, so that the matrix of point source vectors can be written, as $m = 1$, $\mathbf{A} = \mathbf{a} \in \mathbb{C}^{n \times 1}$ and, as $q = 1$, there is only one derivative of **A, so** $\bar{\mathbf{D}} = \mathbf{D} = \mathbf{d} \in \mathbb{C}^{n \times 1}$. Also $\tilde{\mathbf{S}}$ is of size 1×1 so can be written $\tilde{s}$, where $\tilde{s}$ is the mean square value of the source waveform and, in the limit of large p, approaches the power

level of this source. With these definitions we note that $\bar{\mathbf{D}}^{\mathrm{H}}\mathbf{Q}_{\mathrm{A}}\bar{\mathbf{D}} = \mathbf{d}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}$ is a real scalar, and so is $\tilde{\mathbf{S}} * \square \mathbf{1}_q \mathbf{1}_q{}^{\mathrm{T}} = \tilde{s}* = \tilde{s}$ and we have, in this case,

$$b(\theta) = \frac{\psi}{2p\tilde{s}\mathbf{d}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}}. \tag{2.D.1}$$

(b is a scalar value, the lower bound on the variance on the estimate of the single parameter, θ). We note that $\tilde{s}/\psi$ is an estimate of the single channel signal to noise ratio and that $p\tilde{s}/\psi$ is the integrated estimated signal to noise ratio, being improved by taking p frames of data.

2D.2 One source, two parameters

With two parameters but only one source we have $\bar{\mathbf{D}} = \begin{bmatrix}\mathbf{D_1} & \mathbf{D_2}\end{bmatrix} = \begin{bmatrix}\mathbf{d}_1 & \mathbf{d}_2\end{bmatrix}$, where $\mathbf{d}_r = \dfrac{\partial \mathbf{a}}{\partial \theta_r}$, and (2.C.32) can be written, using $\bar{\mathbf{D}}^{\mathrm{H}}\mathbf{Q}_{\mathrm{A}}\bar{\mathbf{D}} = \begin{bmatrix}\mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_1 & \mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_2 \\ \mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_1 & \mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_2\end{bmatrix}$,

$$\mathbf{B}(\boldsymbol{\theta}) = \frac{\psi}{2p}\left(\mathrm{Re}\begin{bmatrix}\mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_1\tilde{s} & \mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_2\tilde{s} \\ \mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_1\tilde{s} & \mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_2\tilde{s}\end{bmatrix}\right)^{-1}.$$

$$= \frac{\psi}{2p\tilde{s}}\left(\begin{bmatrix}\mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_1 & \mathrm{Re}(\mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_2) \\ \mathrm{Re}(\mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_1) & \mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_2\end{bmatrix}\right)^{-1} \tag{2.D.2}$$

If we put $k_{12} = \mathrm{Re}(\mathbf{d}_1{}^{\mathrm{H}}\ \mathbf{Q}_{\mathbf{a}}\mathbf{d}_2) = \mathrm{Re}(\mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_1)$ (as $\mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_1 = (\mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_2)^{\mathrm{H}} = (\mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_2)*$) and $k_{rr} = \mathbf{d}_r{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{d}_r$ (r = 1,2) and invert the 2×2 matrix, we obtain

$$\mathbf{B}(\boldsymbol{\theta}) = \frac{\psi}{2p\tilde{s}}\left(\begin{bmatrix}k_{11} & k_{12} \\ k_{12} & k_{22}\end{bmatrix}\right)^{-1} = \frac{\psi}{2p\tilde{s}}\begin{bmatrix}k_{22} & -k_{12} \\ -k_{12} & k_{11}\end{bmatrix}\Big/\det$$

where det is the determinant of the matrix and $\tilde{s}$ is the mean square value of the signal samples (or power estimate) as given in §2D.1. If we now put

$$\det = \left(k_{11}k_{22} - k_{12}{}^2\right) = k_{11}k_{22}(1 - k_{12}{}^2/k_{11}k_{22}) = k_{11}k_{22}(1-\lambda)$$

then we have

$$\mathbf{B}(\boldsymbol{\theta}) = \frac{\psi}{2p\tilde{s}}\begin{bmatrix}k_{11}{}^{-1} & -\lambda k_{12}{}^{-1} \\ -\lambda k_{12}{}^{-1} & k_{22}{}^{-1}\end{bmatrix}\Big/(1-\lambda)$$

$$= \frac{\psi}{2p\tilde{s}} \begin{bmatrix} \left(\mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_\mathbf{a}\mathbf{d}_1\right)^{-1} & -\lambda\left(\mathrm{Re}(\mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_\mathbf{a}\mathbf{d}_2)\right)^{-1} \\ -\lambda\left(\mathrm{Re}(\mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_\mathbf{a}\mathbf{d}_2)\right)^{-1} & \left(\mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_\mathbf{a}\mathbf{d}_2\right)^{-1} \end{bmatrix} \Big/ (1-\lambda) \qquad (2.D.3)$$

where $\lambda = k_{12}{}^2 / k_{11}k_{22} = (\mathrm{Re}(\mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_\mathbf{a}\mathbf{d}_2))^2 / \mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_\mathbf{a}\mathbf{d}_1\mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_\mathbf{a}\mathbf{d}_2$. (2.D.4)

We see, from the diagonal terms, that the lower bounds on the variances of the estimates of the two parameters, θ_1 and θ_2, are the same as for the single parameters separately, given in (2.D.1), except for the factor $(1 - \lambda)^{-1}$, which increases the values ($0 \le \lambda < 1$, from the Schwarz inequality). In fact if, as can be the case, $\mathbf{d}_1$ and $\mathbf{d}_2$ are orthogonal to $\mathbf{a}$ then applying $\mathbf{Q}_\mathrm{a}$ will leave them unchanged, and if $\mathbf{d}_1$ and $\mathbf{d}_2$ are also orthogonal to each other then k_{12} and λ are zero, in which case the bounds are the same as in the single parameter case. (We certainly have $\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{d}_\mu) = 0$, though not necessarily $\mathbf{a}^{\mathrm{H}}\mathbf{d}_\mu = 0$, i.e. the imaginary part of $\mathbf{a}^{\mathrm{H}}\mathbf{d}_\mu$ is not zero, in general.) This expression shows that, depending on the precise geometry of the data gathering system, there can be some interaction when estimating two (or more) parameters, but this interaction, if the parameters are sufficiently independent, can be small.

[NB $\mathbf{d}_1$ and $\mathbf{d}_2$ here differ from $\mathbf{d}_1$ and $\mathbf{d}_2$ in §2D.3 below: here they represent $\mathbf{d}_{11}$ and $\mathbf{d}_{12}$, the derivatives of the (only) target vector $\mathbf{a}_1$ (where the first suffix 1 is dropped) while in the section below they represent $\mathbf{d}_{11}$ and $\mathbf{d}_{21}$, the derivatives of the two source vectors with respect to the (only) parameter (and in this case the second suffix 1 is dropped)].

2D.3 Two sources, one parameter

This is the case of particular interest in relation to resolving two sources in one parameter. In this case $m = 2$ and $q = 1$, so that $\mathbf{A} = \begin{bmatrix}\mathbf{a_1} & \mathbf{a_2}\end{bmatrix} \in \mathbb{C}^{n\times 2}$ and $\bar{\mathbf{D}} = \mathbf{D} = \begin{bmatrix}\mathbf{d_1} & \mathbf{d_2}\end{bmatrix} \in \mathbb{C}^{n\times 2}$, where $\mathbf{d}_\mu = \dfrac{\partial \mathbf{a}_\mu}{\partial\theta}$. (Note the comment on $\mathbf{d}_1$ and $\mathbf{d}_2$ at the end of the previous section.) The covariance matrix is again 2×2 ($mq \times mq$) and so is the covariance matrix of the source waveforms, $\tilde{\mathbf{S}}$, so, also replacing $\mathbf{Q_a}$ by $\mathbf{Q_A}$, we have, instead of (2.D.2),

$$\mathbf{B}(\boldsymbol{\theta}) = \frac{\psi}{2p}\left(\mathrm{Re}\begin{bmatrix} \mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_\mathbf{A}\mathbf{d}_1\tilde{s}_{11} & \mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_\mathbf{A}\mathbf{d}_2\tilde{s}_{12} \\ \mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_\mathbf{A}\mathbf{d}_1\tilde{s}_{21} & \mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_\mathbf{A}\mathbf{d}_2\tilde{s}_{22} \end{bmatrix}\right)^{-1} . \qquad (2.D.5)$$

Now $\tilde{s}_{21} = \tilde{s}_{12}{}^*$ so $\mathbf{d}_2{}^{\mathrm{H}}\mathbf{Q}_\mathbf{A}\mathbf{d}_1\tilde{s}_{21} = (\mathbf{d}_1{}^{\mathrm{H}}\mathbf{Q}_\mathbf{A}\mathbf{d}_2\tilde{s}_{12})^*$, and performing the matrix inverse in a similar way to that above we obtain

$$\mathbf{B}(\boldsymbol{\theta}) = \frac{\psi}{2p}\begin{bmatrix} k_{11}^{-1} & -\lambda k_{12}^{-1} \\ -\lambda k_{12}^{-1} & k_{22}^{-1} \end{bmatrix} \Big/ (1-\lambda)$$

$$= \frac{\psi}{2p}\begin{bmatrix} \left(\mathbf{d}_1^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_1\tilde{s}_{11}\right)^{-1} & -\lambda\left(\mathrm{Re}(\mathbf{d}_1^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_2\tilde{s}_{12})\right)^{-1} \\ -\lambda\left(\mathrm{Re}(\mathbf{d}_1^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_2\tilde{s}_{12})\right)^{-1} & \left(\mathbf{d}_2^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_2\tilde{s}_{22}\right)^{-1} \end{bmatrix} \Big/ (1-\lambda) \qquad (2.D.6)$$

where here $k_{jk} = \mathrm{Re}(\mathbf{d}_j^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_k\tilde{s}_{jk})$ and λ is $k_{12}^2/k_{11}k_{22}$ as before (see (2.D.4)), with the modified k definitions. (Note that as $\tilde{s}_{rr}$ is real, $\mathbf{d}_r^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_r\tilde{s}_{rr}$ is also real.)

2D.4 Multiple non-coherent sources, two parameters, high integration

As there are just two parameters we have $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_m]$ and $\bar{\mathbf{D}} = \begin{bmatrix}\mathbf{D}_1 & \mathbf{D}_2\end{bmatrix}$, where $\mathbf{D}_k = \begin{bmatrix}\mathbf{d}_{1k} & \mathbf{d}_{2k} & . & . & \mathbf{d}_{mk}\end{bmatrix}$ for parameter k (k = 1,2) and $\mathbf{d}_{rk} = \dfrac{\partial \mathbf{a}_r}{\partial \theta_{rk}}$, for source r and parameter k. From (2.C.32) we have, in this case

$$\mathbf{B}(\boldsymbol{\Theta}) = \frac{\psi}{2p}\left(\mathrm{Re}\left\{\begin{bmatrix}\mathbf{D}_1^{\mathrm{H}} \\ \mathbf{D}_2^{\mathrm{H}}\end{bmatrix}\mathbf{Q}_{\mathbf{A}}\begin{bmatrix}\mathbf{D}_1 & \mathbf{D}_2\end{bmatrix} \odot (\mathbf{S} \square \begin{bmatrix}1 & 1 \\ 1 & 1\end{bmatrix})\right\}\right)^{-1}. \qquad (2.D.7)$$

or

$$\mathbf{B}(\boldsymbol{\Theta}) = \frac{\psi}{2p}\mathrm{Re}\left(\begin{bmatrix}\mathbf{D}_1^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{D}_1 \odot \mathbf{S} & \mathbf{D}_1^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{D}_2 \odot \mathbf{S} \\ \mathbf{D}_2^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{D}_1 \odot \mathbf{S} & \mathbf{D}_2^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{D}_2 \odot \mathbf{S}\end{bmatrix}\right)^{-1} \qquad (2.D.8)$$

(We note $\mathbf{S}$ is real and of size $m \times m$, $\mathbf{D}_k$ is of size $n \times m$ and $\mathbf{Q}_{\mathbf{A}}$ is $n \times n$.) We see that the matrix inside the brackets is of 2×2 block form with blocks of the form $\mathbf{D}_j^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{D}_k \odot \mathbf{S}$ where the symbol indicates element by element multiplication.

In the case of high integration we approximate $\tilde{\mathbf{S}}$, given in (2.C.30) as an average over a finite number of samples, p, by the actual signal covariance matrix. This gives the signal powers on the diagonal and, if we assume the signals are independent, the off-diagonal covariance values are zero. Thus we have, for the case of m sources, $\mathbf{S} = \begin{bmatrix} s_1 & 0 & . & . & 0 \\ 0 & s_2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & . & . & . & s_m \end{bmatrix}$, where s_r is the power of signal r (at the point where the noise level is ψ). As $\mathbf{S}$ is diagonal (and real), it follows that the block is diagonal also,

with diagonal element r given by $\mathbf{d}_{rj}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{rk}s_r$. Thus we have

$$\mathbf{B}(\boldsymbol{\Theta}) = \frac{\psi}{2p}\left(\begin{bmatrix}\mathbf{W} & \mathbf{X}\\ \mathbf{Y} & \mathbf{Z}\end{bmatrix}\right)^{-1} \tag{2.D.8}$$

where $\mathbf{W} = \mathrm{diag}(\mathbf{w})$, etc., and

$$\begin{aligned} &w_r = \mathbf{d}_{r1}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r1}s_r, \quad x_r = \mathbf{d}_{r1}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r2}s_r, \; z_r = \mathbf{d}_{r2}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r2}s_r \\ &y_r = \mathrm{Re}(\mathbf{d}_{r2}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r1})s_r = \mathrm{Re}\left((\mathbf{d}_{r1}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r2})^*\right)s_r = x_r \end{aligned}. \tag{2.D.9}$$

The 2×2 matrix of m×m blocks in (2.D.8) can be inverted to give

$$\mathbf{B}(\boldsymbol{\Theta}) = \frac{\psi}{2p}\begin{bmatrix}\left(\mathbf{W}-\mathbf{X}\mathbf{Z}^{-1}\mathbf{Y}\right)^{-1} & * \\ * & \left(\mathbf{Z}-\mathbf{Y}\mathbf{W}^{-1}\mathbf{X}\right)^{-1}\end{bmatrix} \tag{2.D.10}$$

where we ignore the off-diagonal blocks as we are only really interested in the variances of the parameters (rather than the covariances of parameters, both of parameters, such as azimuth and elevation, of the same source (e.g. of α_1 and ε_1) and of different sources (e.g. of α_1 and α_2 or α_1 and ε_2)), which are to be found on the diagonals of the diagonal blocks. As we noted above, all the blocks are themselves diagonal which makes their inverses simply to find – we just replace the diagonal elements by their reciprocals – and their products are simply the element-by-element products of the diagonal elements. Thus the rth diagonal element of the first diagonal block in (2.D.10) is given by $\dfrac{1}{w_r - x_r y_r / z_r} = \dfrac{1/w_r}{1 - x_r y_r / w_r z_r}$ and the result for the second block is $\dfrac{1}{z_r - x_r y_r / w_r} = \dfrac{1/z_r}{1 - x_r y_r / w_r z_r}$. Thus we can put

$$\mathbf{B}(\boldsymbol{\Theta}) = \frac{\psi}{2p}\begin{bmatrix}diag(\mathbf{u}) & * \\ * & diag(\mathbf{v})\end{bmatrix} \tag{2.D.11}$$

where $\mathbf{u}$ is an m-vector containing the CRB bounds on the variances of the estimates of the first parameter for the m sources and $\mathbf{v}$ contains those for the second parameter. For source r we have

$$u_r = \left(\mathbf{d}_{r1}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r1}s_r\right)^{-1}\Big/(1-\mu_r) \quad \text{and} \quad v_r = \left(\mathbf{d}_{r2}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r2}s_r\right)^{-1}\Big/(1-\mu_r), \tag{2.D.12}$$

using (2.D.9) , where

$$\mu_r = \frac{\left(\mathrm{Re}(\mathbf{d}_{r1}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r2})\right)^2}{\mathbf{d}_{r1}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r1}\mathbf{d}_{r2}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r2}}, \tag{2.D.13}$$

and μ_r can be seen to be less than unity by the Schwarz inequality.

This is a result which seems intuitively satisfactory. Without the factor $1-\mu_r$, u_r and v_r are the values expected from signals without any interaction; the results correspond to those of the single target case. The factor $(1-\mu_r)^{-1}$ increases the bounds on the variances, as might be expected due to the mutual interference of the signals.

If we reorder the parameters, instead of $\alpha_1, \alpha_2, \ldots \alpha_m, \varepsilon_1, \varepsilon_2, \ldots \varepsilon_m$, into the form $\alpha_1, \varepsilon_1, \alpha_2, \varepsilon_2, \ldots \alpha_m, \varepsilon_m$, then we can put (2.D.7) in the form

$$\mathbf{B}(\boldsymbol{\Theta}) = \frac{\psi}{2p}\left(\mathrm{Re}\{ \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} & \ldots & \mathbf{M}_{1m} \\ \mathbf{M}_{21} & \mathbf{M}_{22} & \ldots & \mathbf{M}_{2m} \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \mathbf{M}_{m1} & \mathbf{M}_{m2} & \ldots & \mathbf{M}_{mm} \end{bmatrix} \odot \left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \square \mathbf{S} \right) \} \right)^{-1}$$

$$= \frac{\psi}{2p}\left(\mathrm{Re} \begin{bmatrix} \mathbf{M}_{11}\tilde{s}_{11} & \mathbf{M}_{12}\tilde{s}_{12} & \ldots & \mathbf{M}_{1m}\tilde{s}_{1m} \\ \mathbf{M}_{21}\tilde{s}_{21} & \mathbf{M}_{22}\tilde{s}_{22} & \ldots & \mathbf{M}_{2m}\tilde{s}_{2m} \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \mathbf{M}_{m1}\tilde{s}_{m1} & \mathbf{M}_{m2}\tilde{s}_{m2} & \ldots & \mathbf{M}_{mm}\tilde{s}_{mm} \end{bmatrix} \right)^{-1} \tag{2.D.14}$$

where the 2×2 blocks $\mathbf{M}_{rs}$ are given by

$$\mathbf{M}_{rs} = \begin{bmatrix} \mathbf{d}_{r1}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{s1} & \mathbf{d}_{r1}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{s2} \\ \mathbf{d}_{r2}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{s1} & \mathbf{d}_{r2}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{s2} \end{bmatrix} \tag{2.D.15}$$

and

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \square \mathbf{S} = \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\tilde{s}_{11} & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\tilde{s}_{12} & \ldots & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\tilde{s}_{1m} \\ \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\tilde{s}_{21} & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\tilde{s}_{22} & \ldots & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\tilde{s}_{2m} \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\tilde{s}_{m1} & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\tilde{s}_{m2} & \ldots & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\tilde{s}_{mm} \end{bmatrix}.$$

In the case of high integration, with noncoherent signals **S** is diagonal and we replace $\tilde{s}_{rr}$ by the power level s_r, which is real. In this case

$$\mathbf{B}(\boldsymbol{\Theta}) = \frac{\psi}{2p}\left(\mathrm{Re}\begin{bmatrix} \mathbf{M}_{11}s_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{22}s_2 & \cdots & \mathbf{0} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}_{mm}s_m \end{bmatrix} \right)^{-1}$$

$$= \frac{\psi}{2p}\begin{bmatrix} \mathbf{M}_{11}^{-1}/s_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{22}^{-1}/s_2 & \cdots & \mathbf{0} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}_{mm}^{-1}/s_m \end{bmatrix} \qquad (2.D.16)$$

with a slight change in notation so that we have, now,

$$\mathbf{M}_{rr} = \begin{bmatrix} \mathbf{d}_{r1}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r1} & \mathrm{Re}(\mathbf{d}_{r1}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r2}) \\ \mathrm{Re}(\mathbf{d}_{r2}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r1}) & \mathbf{d}_{r2}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_{r2} \end{bmatrix}. \qquad (2.D.17)$$

Performing the inverse of $\mathbf{M}_{rr}$ (using $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} 1/a & -\mu/c \\ -\mu/b & 1/d \end{bmatrix} \Big/ (1-\mu)$, with $\mu = bc/ad$) we see that the diagonal components of the 2-parameter CRB covariance matrix for target *r* are the same as given in (2.D.12) and (2.D.13).

2D.5 Two non-coherent sources, two angle parameters, high integration

In this case we have parameters α and ε (q = 2) for two targets (m = 2), or 4 parameters in total. We put $\mathbf{a}_{1\alpha}$ for $\mathbf{d}_{11}$ and $\mathbf{a}_{1\varepsilon}$ for $\mathbf{d}_{12}$, and similarly for $\mathbf{d}_{2k}$ (recalling that $\mathbf{d}_{rk} = \frac{\partial \mathbf{a}_r}{\partial \theta_{rk}}$, putting $\theta_1 = \alpha$, $\theta_2 = \varepsilon$, and using the notation of Chapter 3: $\mathbf{a}_{r\alpha} = \frac{\partial \mathbf{a}_r}{\partial \alpha_r}$, $\mathbf{a}_{r\varepsilon} = \frac{\partial \mathbf{a}_r}{\partial \varepsilon_r}$). With this notation in (2.D.15), (2.D.12) and (2.D.13) we find that the CRB for the parameters of target r is given by

$$\mathbf{B}(\alpha_r, \varepsilon_r) = \frac{\psi}{2ps_r}\begin{bmatrix} \mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\alpha} & \mathrm{Re}(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\varepsilon}) \\ \mathrm{Re}(\mathbf{a}_{r\varepsilon}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\alpha}) & \mathbf{a}_{r\varepsilon}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\varepsilon} \end{bmatrix}^{-1} \qquad (2.D.18)$$

or, more conveniently, for target *r*, the variances of the estimated parameters are

$$\mathrm{var}(\alpha_r) = \frac{\psi}{2ps_r}\frac{\left(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\alpha}\right)^{-1}}{(1-\mu_r)} \text{ and } \mathrm{var}(\varepsilon_r) = \frac{\psi}{2ps_r}\frac{\left(\mathbf{a}_{r\varepsilon}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\varepsilon}\right)^{-1}}{(1-\mu_r)} \quad (2.D.19)$$

where

$$\mu_r = \frac{\left(\mathrm{Re}(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\varepsilon})\right)^2}{\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\alpha}\mathbf{a}_{r\varepsilon}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\varepsilon}}. \qquad (r \in \{1,2\}) \quad (2.D.20)$$

Chapter 3: Study of DF accuracy with element gain errors

3.1 INTRODUCTION

We consider here the effect of receiver channel errors on the performance of an array used for direction finding on one or more targets. Although superresolution systems use more complex processing when multiple signals are present, these methods are essentially extensions of the beamforming method, and in the single signal case, considered in Sections 3.2 and 3.3 of this chapter, reduce to simple beamforming. For example, in this case MUSIC gives a single principal eigenvector, which is (apart from corruption by noise) the point source response vector (PSV) for the target. In this case the MUSIC scan is as for simple beamforming. The vector is, of course, the *actual* target PSV, which will generally differ slightly from the stored value from the sampled manifold, whether this manifold vector is obtained by calculation from the array geometry, by calibration of the array, or by a combination of these methods (such as interpolated calibration). In the case of IMP the first scan also is a simple beamforming scan, and this will find the signal when there is only one present. For linear arrays, considered in Section 3.2, the scan is only over one dimension, but the argument here is general and applies for more complex systems, for example planar arrays with two-dimensional scans, in azimuth and elevation, as in Sections 3.3 and 3.4. In Section 3.4 we obtain a rather general expression for the DF errors on multiple targets which is then developed into a more practical form for the specific case of two targets.

The difference between the stored set of values for the target PSV and the actual values is the error vector, and here it is assumed that some estimated (or statistical) value of this error is available, for the vector components. We ignore the effect of receiver noise on the performance (effectively taking a very high integrated S/N value) and consider that the only limitation on the performance is that due to the complex gain errors. We take a statistical approach, modelling the errors in phase and amplitude in each element gain as samples from independent zero mean distributions of given variances. We first obtain an explicit expression for the error in the angle estimate due to a specific set of errors in the channel gains (i.e. in the stored value of the PSV), then find the standard deviation of the angle error estimates in terms of the standard deviation of the gain errors. Generally we take the ideal array to be of identical elements with equal parallel patterns (i.e. all with the same gain in any given direction, though the patterns need not be omnidirectional). However, in §3.3.4 we show how the theory can be made to include the case of more general element patterns, with different gains to the targets.

The simulation results given in §3.3.5 confirm these expressions.

It is important to distinguish between the meaning, or significance, of the statistical results obtained for the case of channel errors in contrast to the case where the measurement errors are caused by a limited signal to noise ratio. In the latter case the angle estimation errors will be different every time estimates are made of target angular positions (even in stationary conditions) – as the noise perturbing the measurements will be different each time. In the case where channel errors are the only significant limitation on the accuracy, as a given array will have a certain fixed set of errors the estimation errors will be the same each time. However, this does not help the user, as he does not know what the estimation errors actually are or what the channel errors are. (If they were known then the PSVs would be corrected and the errors would be eliminated.) We take as the model here that the user knows, or has a good estimate of, the variance (or equivalently its square root, the standard deviation) of the channel errors. Then we relate the statistics (the variance, or s.d.) of the angle estimation errors to the statistics of the channel errors. These statistics in effect relate to a large ensemble of identical arrays except for their randomly chosen channel errors. Then we can say that a given array is one sample of this ensemble and the error from this array, although in principle deterministic, given the actual channel errors, can only be described in statistical form, as the channel errors are unknown up to a given variance. (With this statistical error model we are not including the case of failure of an element or of a signal channel. Clearly this would be a valid subject for study but would require a rather different approach. In practice the case might be tackled by detecting the failure and proceeding to use the remaining array of $n - 1$ elements, with PSVs of length $n - 1$, excluding the responses of the failed element.)

If $\mathbf{a}(\theta)$ is the normalized PSV for angle θ and $\mathbf{a}(\phi)$ is the normalized vector for the target, in direction ϕ, then the beamforming function in the single target case is given (within a constant scaling factor) by

$$f(\theta) = \left|\mathbf{a}(\theta)^{\mathrm{H}}\mathbf{a}(\phi)\right|^2. \tag{3.1.1}$$

(NB: θ and ϕ may be considered 2-dimensional, in general, specified in azimuth α and elevation ε.)

By the Schwarz inequality for inner products we have $\left|\mathbf{a}(\theta)^{\mathrm{H}}\mathbf{a}(\phi)\right|^2 \le \left\|\mathbf{a}(\theta)\right\|^2 \left\|\mathbf{a}(\phi)\right\|^2$, with equality only when $\mathbf{a}(\theta) = \mathbf{a}(\phi)$, which, unless there are exact ambiguities (i.e. $\mathbf{a}(\theta_1)$

= $\mathbf{a}(\theta_2)$ for some pairs (θ_1,θ_2) where $\theta_1 \neq \theta_2$), implies $\theta = \phi$. Thus f is a maximum when $\theta = \phi$, as required, and the maximum value is 1, using normalized vectors. This is for the case of no errors, where the set of manifold vectors is exact for all signal directions (or at least for all the directions for which the manifold vectors are specified). In practice this will not be the case and there will be a difference between the target PSV for the target direction ϕ and the stored manifold vector for this direction. It is convenient to consider the target PSV to be in 'error' with respect to the manifold vector, though the converse is more accurate, but the effect is equivalent. Here we determine an expression for the error in the angle estimate which results from a small vector error $\Delta\mathbf{a}$ between the stored manifold vector and the target PSV. We suppose that the components of $\Delta\mathbf{a}$ are taken from distributions of given variance and zero mean. The error model for the components of the element gains is given in more detail in Appendix 3.1A.

In the simulations, quoted in §3.2.4, §3.3.3, §3.3.5 and §3.4.6 below, which were used to confirm the theoretical expressions, the channel errors were added to the single target vector (or vectors, in §3.4), rather than to the array manifold vectors (or point response vectors, PSVs) used to evaluate the beamforming function. This is equivalent to *not* having an error on the target vector but having the same additive error vector (or rather, its negative, which would have the same assumed statistics of zero mean and given standard deviation for each component) on all the manifold PSVs used. In turn, this is equivalent to assuming the errors are independent of the source direction that the PSVs correspond to. This will be the case if the errors arise in the channels between the elements and the analogue to digital convertors, at the start of the signal processing, which is indeed a primary source of error. Even if there are any significant look direction dependent errors (i.e. errors in the specification of the element antenna gain as a function of direction) these are likely to be slowly varying across the small angular range used to find the function peak, so again the approximation of equal errors over this range should be reasonable. If the manifold used was obtained entirely by calibration at each position (in practice an unlikely method because of its difficulty and cost) then there might be significant errors which were random from one PSV to the next, and the model taken for the simulations would not be valid in this case. However, if this form of calibration were used, it might be considered that such fluctuations would not occur in the true values and some smoothing would be appropriate over some range of angle. The main case of errors dependent on look direction is when there is

significant multipath between sources and the array. In this case reflections from relatively distant points effectively increase the aperture and produce a fast varying gain variation. This source of error is not modelled in the analysis here.

The theory here is developed beginning with the simple case of a regular linear array of equal, parallel pattern elements, which only gives angle estimation in one dimension. In principle the most complex case (of a general array with non-equal elements, estimating directions of several signals in two dimensions) could be taken, and results for the simpler cases derived from the more general case. However, the development of the theory from the simplest case gives the simpler results appropriate to these cases without having to simplify a more complex expression, and clarifies the principles applied and the methods used. Thus in Section 3.2 we analyse the case of linear arrays (including non-uniform arrays in §3.2.3) used for one-dimensional angle estimation, then in Section 3.3 we tackle the case of non-linear arrays (in practice generally planar arrays though in fact the analysis covers the less common case of volume arrays) used to find targets in two angle dimensions. In Subsection 3.3.4 we take the case of arrays whose elements do not have equal gain in any given target direction, as previously assumed, to obtain the most general result so far. This is still for the case of a single target, but in Section 3.4 we extend the study to the case of multiple targets (though returning to equal parallel patterns). In §3.4.3 we obtain a rather general expression for the estimation errors resulting from a specific set of channel errors. In §3.4.4 we expand this expression for the case of only two targets, and in §3.4.5 we obtain statistical results for the two target case. This follows the approach of Sections 3.2 and 3.3 where we first find an expression for the particular estimation errors resulting from a given set of channel errors, and then use this result to obtain the statistical results.

The simulations show that the theory, for the error model taken (see Appendix 3.1A), does in fact give the statistical error performance correctly, with excellent agreement, in general. This means that, having confidence in the theory, for the cases satisfying the conditions assumed, we can plot the statistical error performance (the azimuth and elevation error variances) for the given array for all directions of interest. This is much more general and also much more efficient than attempting to find the performance of an array by large numbers of simulations.

Appendix 3.1A Channel gain error model

We model the channel errors as fluctuations in amplitude relative gain and in phase, so the magnitudes of the errors are proportional to the actual channel gain. We take the gain and phase errors as samples from independent distributions of zero means and given variances. For element k, we can consider the error to be the sum of two orthogonal components, Δa_{kp}, parallel with a_k, and Δa_{kq} orthogonal to a_k. The magnitudes of these two components are considered to be proportional to the magnitude of the channel gain, g_k. We assume that the components are taken from zero mean distributions and we assume these error components are independent. In general both forms of error are present, but in fact it is shown below for the single target case (in §3.2 and §3.3) that the amplitude fluctuations do not contribute to the angle estimation errors (at least for small errors, for which the approach taken here is applicable), and this has been confirmed in simulation. Thus for this case we need only a value for the variance of the orthogonal component Δa_{kq}, which is related to the variance of the phase angle ϕ, and this is a value which is likely to be known, or which can be estimated reasonably well, in practice. In the two signal case it is found (§3.4.5) that amplitude errors do contribute to the angle estimation errors and this is confirmed in the simulation results of §3.4.6.

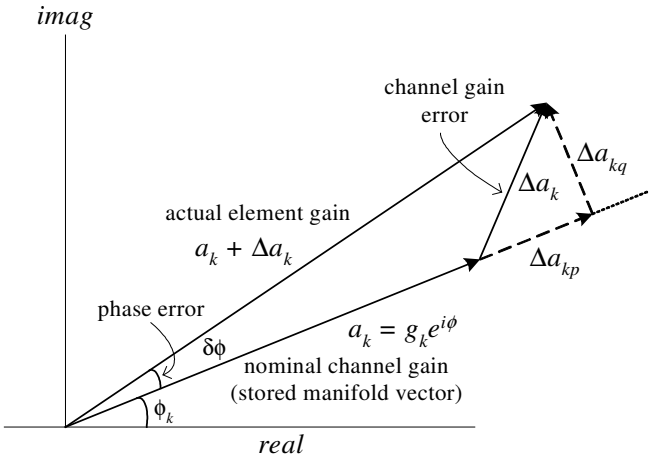


Figure 3.1 Model for element errors

3.2 LINEAR ARRAYS

3.2.1 Angle error estimate

In this case, with a linear array (of equal parallel pattern elements), we can only carry out DF in one angle dimension, with the target direction on a cone of directions whose axis is that of the array. In this section, therefore, we let α represent the target angle (usually measured from the plane normal to the array – the complement of the cone half-angle) and let $\Delta\alpha$ be the error in the DF estimation of this angle α.

The target point response vector is given by the manifold vector for the signal direction

$\mathbf{a}(\alpha)$ plus the error vector $\Delta\mathbf{a}$**.** The vector used for the scan function in (3.1.1) is normalized so let this signal vector be

$$\tilde{\mathbf{a}}(\alpha) = h(\mathbf{a}(\alpha) + \Delta\mathbf{a}), \tag{3.2.1}$$

where h is a normalizing factor. To simplify the appearance of the equations below we may put **a, d** and **s** for $\mathbf{a}(\alpha)$, $\mathbf{d}(\alpha)$ and $\mathbf{s}(\alpha)$. We take the case where $\|\Delta\mathbf{a}\|^2 << \|\mathbf{a}\|^2 = 1$, in general. h is given by

$$\|\tilde{\mathbf{a}}(\alpha)\|^2 = \ 1 = h^2\|\mathbf{a}(\alpha) + \Delta\mathbf{a}\|^2 = h^2(1 + 2\mathrm{Re}(\mathbf{a}(\alpha)^{\mathrm{H}}\Delta\mathbf{a}) + \|\Delta\mathbf{a}\|^2)$$

or
$$h^2 = 1/(1 + 2\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a}) + \|\Delta\mathbf{a}\|^2). \tag{3.2.2}$$

We now find an expression for the beamformed function f in the region of α, at $\alpha + \Delta\alpha$, in order to find the value of $\Delta\alpha$ giving the peak of f. This value is then the angle error in estimating the target position. As $\Delta\mathbf{a}$, the perturbation in $\mathbf{a}$, is small we assume $\Delta\alpha$, the perturbation in α, is also small, and we find an approximation for f to second order only, giving a quadratic function in $\Delta\alpha$ near the true target direction, α.

Let
$$\mathbf{a}(\alpha + \Delta\alpha) = k(\mathbf{a}(\alpha) + \Delta\alpha\mathbf{d}(\alpha) + \tfrac{1}{2}\Delta\alpha^2\mathbf{s}(\alpha) + \ \ldots\ ) \tag{3.2.3}$$

taking the second order Taylor expansion, where

$$\mathbf{d}(\alpha) = \frac{d\mathbf{a}(\alpha)}{d\alpha},\ \mathbf{s}(\alpha) = \frac{d\mathbf{d}(\alpha)}{d\alpha} = \frac{d^2\mathbf{a}(\alpha)}{d\alpha^2} \tag{3.2.4}$$

and k is a normalizing factor. Similarly to the expression for h, we have for k, up to second order in $\Delta\alpha$,

$$k^{-2} = \left\|\mathbf{a}(\alpha) + \Delta\alpha\mathbf{d}(\alpha) + \tfrac{1}{2}\Delta\alpha^2\mathbf{s}(\alpha) + \ \ldots\ \right\|^2$$

$$= 1 + 2\Delta\alpha\,\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{d}) + \Delta\alpha^2\left\|\mathbf{d}\right\|^2 + \Delta\alpha^2\,\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{s}) + \ \ldots \tag{3.2.5}$$

However, from the definition of the array gain vector used here (see (3.2.11) or (3.2.20) below), it follows that $\mathbf{a}(\alpha)$ is normalized for all values of α, and we can deduce that the expansion of $\mathbf{a}(\alpha + \Delta\alpha)$ is also normalized and so we have $k = 1$ (exactly, without approximations. In fact we can see that $k = 1$ at least to second order using the results (3.2A.2) and (3.2A.6) from Appendix 3.2A in (3.2.5).)

The beamforming function for direction $\alpha + \Delta\alpha$ is given, to second order, by

$$f(\alpha + \Delta\alpha) = \left|\mathbf{a}(\alpha + \Delta\alpha)^{\mathrm{H}}\tilde{\mathbf{a}}(\alpha)\right|^2 = h^2k^2\left|(\mathbf{a}(\alpha) + \Delta\alpha\mathbf{d}(\alpha) + \tfrac{1}{2}\Delta\alpha^2\mathbf{s}(\alpha))^{\mathrm{H}}(\mathbf{a}(\alpha) + \Delta\mathbf{a})\right|^2$$

$$= h^2 \left| 1 + \mathbf{a}^{\mathrm{H}}\Delta\mathbf{a} + \Delta\alpha\mathbf{d}^{\mathrm{H}}\mathbf{a} + \Delta\alpha\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a} + \tfrac{1}{2}\Delta\alpha^2\mathbf{s}^{\mathrm{H}}\mathbf{a} \right|^2, \qquad (3.2.6)$$

taking the magnitude of $\Delta\mathbf{a}$ to be small so that we neglect $\Delta\alpha^2\Delta\mathbf{a}$ terms as well as $\Delta\alpha^3$ terms (and higher). (We have used $\mathbf{a}^{\mathrm{H}}\mathbf{a} = 1$, $k = 1$ and we note that h is independent of $\Delta\alpha$.) Expanding (3.2.6) to second order gives

$$\begin{aligned} f(\alpha + \Delta\alpha) = h^2 (1 + 2\,\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a}) + \left|\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a}\right|^2 + 2\Delta\alpha\,\mathrm{Re}(\mathbf{d}^{\mathrm{H}}\mathbf{a}) + 2\Delta\alpha\,\mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}) \\ + 2\Delta\alpha\,\mathrm{Re}((\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a})(\mathbf{a}^{\mathrm{H}}\mathbf{d}) + \Delta\alpha^2 \left|\mathbf{d}^{\mathrm{H}}\mathbf{a}\right|^2 + \Delta\alpha^2\,\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{s})). \end{aligned} \qquad (3.2.7)$$

(Note that the terms $2\Delta\alpha\,\mathrm{Re}((\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a})(\Delta\mathbf{a}^{\mathrm{H}}\mathbf{d}))$, $2\Delta\alpha^2\,\mathrm{Re}((\mathbf{a}^{\mathrm{H}}\mathbf{d})(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}))$, $\Delta\alpha^2\,\mathrm{Re}((\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a})(\mathbf{a}^{\mathrm{H}}\mathbf{s}))$ and $\Delta\alpha^3\,\mathrm{Re}((\mathbf{a}^{\mathrm{H}}\mathbf{d})(\mathbf{s}^{\mathrm{H}}\mathbf{a}))$ are all of third order if $\Delta\alpha$ and $\|\Delta\mathbf{a}\|$ are small, so these terms have been neglected, as have the fourth order terms. We have also used $(\mathbf{u}^{\mathrm{H}}\mathbf{v})^* = \mathbf{u}^{\mathrm{T}}\mathbf{v}^* = \mathbf{v}^{\mathrm{H}}\mathbf{u}$ where $\mathbf{u}$ and $\mathbf{v}$ may be $\mathbf{a}$, $\Delta\mathbf{a}$, $\mathbf{d}$ or $\mathbf{s}$.)

To a second order approximation f is a quadratic function of $\Delta\alpha$ near α. If we put

$$f(\alpha + \Delta\alpha) = A + 2B\Delta\alpha + C\Delta\alpha^2 + \ldots,$$

then (on differentiating with respect to $\Delta\alpha$) the maximum of f occurs when $2B + 2C\Delta\alpha = 0$, or

$$\Delta\alpha = -\frac{B}{C} = -\frac{\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{d}) + \mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}) + \mathrm{Re}((\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a})(\mathbf{a}^{\mathrm{H}}\mathbf{d}))}{\left|\mathbf{a}^{\mathrm{H}}\mathbf{d}\right|^2 + \mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{s})}. \qquad (3.2.8)$$

Now $\mathbf{a}(\alpha)^{\mathrm{H}}\mathbf{a}(\alpha) = 1$ for all α (see (3.2.11) or (3.2.20) for the components of $\mathbf{a}$) so, on differentiating with respect to α we have $\mathbf{d}^{\mathrm{H}}\mathbf{a} + \mathbf{a}^{\mathrm{H}}\mathbf{d} = 0$, i.e. $2\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{d}) = 0$ (as also seen in (3.2A.2)) so finally we have

$$\Delta\alpha = -\frac{\mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}) + \mathrm{Re}((\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a})(\mathbf{a}^{\mathrm{H}}\mathbf{d}))}{\left|\mathbf{a}^{\mathrm{H}}\mathbf{d}\right|^2 + \mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{s})} = \frac{\mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}) + \mathrm{Re}((\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a})(\mathbf{a}^{\mathrm{H}}\mathbf{d}))}{\|\mathbf{d}\|^2 - \left|\mathbf{a}^{\mathrm{H}}\mathbf{d}\right|^2}, \qquad (3.2.9)$$

using (3.2A.6).

Now it is clear that this result, (the error in the peak position of the beamforming function due to the error $\Delta\mathbf{a}$ in the array vectors), cannot depend on the reference point, or origin, used to specify the array element positions, so we could conveniently use the array centroid, the mean of the element vector positions, as this reference. In this case, with the mean position $\bar{\mathbf{r}}$ set to zero, it follows that $\mathbf{a}^{\mathrm{H}}\mathbf{d} = 0$ (see the discussion in Appendix 3.2A, following equation (3.2A.2)) and so (3.2.9) takes the simpler form

$$\Delta\alpha = \frac{\mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a})}{\|\mathbf{d}\|^2}. \qquad (3.2.10)$$

The results (3.2.9) and (3.2.10) are general (for an array of equal, parallel pattern elements) for the case of a single parameter α, and are not limited to a linear array at this point.

3.2.2 Statistical result for the uniform linear array

Let the array consist of n elements at a regular spacing of d wavelengths, and take the centroid of the array, which is the midpoint of the array in this case, as the reference point. (The centroid is considered further in Appendix 3.2B, in particular when the element gains differ.) The array positions are then given by $-(n\text{-}1)d/2$, $-(n\text{-}3)d/2$, . . ., $(n\text{-}3)d/2$, $(n\text{-}1)d/2$. On the assumption of equal parallel patterns (i.e. all the elements have the same gain in any given direction, apart from the phase factor due to position), the normalized manifold vector for direction α (measured from the array normal) is

$$\mathbf{a}(\alpha) = [\exp(-(n-1)i\phi/2) \quad \exp(-(n-3)i\phi/2) \quad . \; . \; . \quad \exp((n-1)i\phi/2)]^{\mathrm{T}} \big/ \sqrt{n} \qquad (3.2.11)$$

where $\phi(\alpha) = 2\pi d\sin\alpha$.

The derivative of **a** is given by

$$\mathbf{d}(\alpha) = 2\pi i d\cos\alpha[-(n-1)a_1(\alpha)/2 \quad -(n-3)a_2(\alpha)/2 \quad . \; . \; . \quad (n-1)a_n(\alpha)/2] \qquad (3.2.12)$$

where the kth element of **a** is given by $a_k = \exp(i(k - (n+1)/2)\phi)/\sqrt{n}$. (We have omitted the indication of dependence on α to make the expressions clearer. This will generally be the case subsequently; it should be understood that **a**, **d** and many other variables are assumed to have their values at the signal angle α.) We note that, using $|a_k|^2 = 1/n$,

$$\mathbf{a}^{\mathrm{H}}\mathbf{d} = 2\pi i d\cos\alpha(-(n\text{-}1)/2\text{-}(n\text{-}3)/2 \; . \; . \; . \; +(n\text{-}3)/2+(n\text{-}1)/2])/n = 0. \qquad (3.2.13)$$

Also

$$\mathbf{d}^{\mathrm{H}}\mathbf{d} = (2\pi d\cos\alpha)^2(((n\text{-}1)/2)^2+((n\text{-}3)/2)^2+ \; . \; . \; . \; +((n\text{-}3)/2)^2+((n\text{-}1)/2)^2))/n$$

$$= (2\pi\cos\alpha)^2 M \qquad (3.2.14)$$

where

$$M = \frac{1}{n}\sum_{m=-(n-1)/2}^{(n-1)/2} m^2 d^2 = \frac{(n^2-1)d^2}{12}. \qquad (3.2.15)$$

as shown in Appendix 3.2C. M is the second moment of the array about its centroid.

We now consider the variance of the scalar Re($\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}$). Let $\mathbf{c}$ be the real vector

$$\mathbf{c} = d[-(n-1)/2\ -(n-3)/2\ \ldots\ (n-1)/2]^{\mathrm{T}} \tag{3.2.16}$$

so that, from (3.2.12) we have $d_k = 2\pi i \cos\alpha c_k a_k$. Then we can put

$$u = \mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}) = \mathrm{Re}(-2\pi i \cos\alpha \sum_{k=1}^{n} c_k a_k * \Delta a_k)$$

$$= 2\pi\cos\alpha \sum_{k=1}^{n} c_k\ \mathrm{Im}(a_k * \Delta a_k) = 2\pi\cos\alpha \mathbf{c}^{\mathrm{T}}\mathbf{y}$$

where $y_k = \mathrm{Im}(a_k * \Delta a_k)$. We take the errors Δa_k to be from a zero mean distribution so it follows that the expectation value of y_k is zero and so also is that of u. The variance of u is thus the expectation value of u^2, i.e.

$$\mathrm{var}(u) = \left\langle u^2 \right\rangle = \left(2\pi\cos\alpha\right)^2 \left\langle \mathbf{c}^{\mathrm{T}}\mathbf{y}\mathbf{y}^{\mathrm{T}}\mathbf{c} \right\rangle = \left(2\pi\cos\alpha\right)^2 \mathbf{c}^{\mathrm{T}} \left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle \mathbf{c}$$

(using $\mathbf{c}^{\mathrm{T}}\mathbf{y} = \mathbf{y}^{\mathrm{T}}\mathbf{c}$). From Appendix 3.2D (equation (3.2D.4)) we see that $\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle = \sigma_\phi{}^2\mathbf{I}/n^2$ so that

$$\mathrm{var}(u) = \left(2\pi\cos\alpha\right)^2 (\sigma_\phi{}^2/n^2)\mathbf{c}^{\mathrm{T}}\mathbf{c} = \left(2\pi\cos\alpha\right)^2 \sigma_\phi{}^2 M/n$$

where we have used, from (3.2.15) and (3.2.16) above,

$$\mathbf{c}^{\mathrm{T}}\mathbf{c} = \sum_{k=1}^{n} c_k{}^2 = nM$$

and $\sigma_\phi{}^2$ is the variance of the phase errors.

Finally, taking the square root to obtain the standard deviation, and dividing by $\|\mathbf{d}\|^2$, given above, the standard deviation of the error β is

$$\mathrm{s.d.}(\Delta\alpha) = \frac{2\pi\cos\alpha\sigma_\phi\sqrt{M/n}}{\left(2\pi\cos\alpha\right)^2 M} = \frac{\sigma_\phi}{2\pi\cos\alpha\sqrt{nM}} \tag{3.2.17}$$

or, in the case of a regular linear array with n elements at separation d wavelengths,

$$\mathrm{s.d.}(\Delta\alpha) = \frac{\sigma_\phi}{2\pi d\cos\alpha\sqrt{n(n^2-1)/12}} = \frac{\sigma_\phi}{\pi d\cos\alpha\sqrt{n(n^2-1)/3}}. \tag{3.2.18}$$

3.2.3 Result for more general linear array

In this section we express (3.2.10) using the form of a second order moment matrix

describing the array geometry. This is for the case where the centroid is taken for the reference point for the array element positions – i.e. for the case where $\bar{\mathbf{r}} = 0$. If we take the more general case we obtain the same result, as shown in Appendix 3.2E.

We take the case of an array of *n* equal, parallel pattern antenna elements – i.e. an array where all the elements have the same antenna gain in any given direction, though the element patterns need not be omnidirectional, and the gain may vary with direction. Let the normalized manifold vector for direction α be given by

$$\mathbf{a}(\alpha) = \left[a_1(\alpha) \quad a_2(\alpha) \quad . \quad . \quad . \quad a_n(\alpha)\right]^{\mathrm{T}} \tag{3.2.19}$$

where the components of **a** are given by

$$a_k(\alpha) = \exp\left(2\pi i \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}(\alpha)\right)/\sqrt{n}\,. \tag{3.2.20}$$

The factor $1/\sqrt{n}$ ensures **a** is normalized, (i.e. such that $\|\mathbf{a}\|^2 = \sum_k |a_k|^2 = 1$) and $\mathbf{e}(\alpha)$ is the unit vector in direction α. $\mathbf{r}_k = [x_k \quad y_k \quad z_k]^{\mathrm{T}}$ is the position of element *k*, in wavelength units, from some reference point taken here as the geometric centroid of the array. (We use Σ_k here to mean the sum from $k = 1$ to $k = n$.)

The elements of **d**, the first derivative of **a** with respect to α, are given by

$$d_k(\alpha) = \frac{da_k(\alpha)}{d\alpha} = 2\pi i \mathbf{r}_k{}^{\mathrm{T}} \dot{\mathbf{e}}(\alpha) a_k(\alpha) \tag{3.2.21}$$

where $\dot{\mathbf{e}}$ is the first derivative of **e** with respect to α.

From (3.2.21) we have

$$\|\mathbf{d}\|^2 = \sum_k |d_k|^2 = \sum_k \left(2\pi \mathbf{r}_k{}^{\mathrm{T}} \dot{\mathbf{e}}\right)^2 |a_k|^2 = 4\pi^2 \sum_k \dot{\mathbf{e}}^{\mathrm{T}} \mathbf{r}_k \mathbf{r}_k{}^{\mathrm{T}} \dot{\mathbf{e}}/n\,.$$

$$= 4\pi^2 \dot{\mathbf{e}}^{\mathrm{T}} \left(\sum_k \mathbf{r}_k \mathbf{r}_k{}^{\mathrm{T}}/n\right) \dot{\mathbf{e}} \tag{3.2.22}$$

using $|a_k|^2 = 1/n$ and $\mathbf{r}_k{}^{\mathrm{T}} \dot{\mathbf{e}} = \dot{\mathbf{e}}^{\mathrm{T}} \mathbf{r}_k$.

Now we note that $\mathbf{r}_k \mathbf{r}_k{}^{\mathrm{T}}$ is a 3×3 real matrix given by

$$\mathbf{r}_k \mathbf{r}_k{}^{\mathrm{T}} = \begin{bmatrix} x_k{}^2 & x_k y_k & x_k z_k \\ x_k y_k & y_k{}^2 & y_k z_k \\ x_k z_k & y_k z_k & z_k{}^2 \end{bmatrix}$$

and so

$$\sum_k \mathbf{r}_k \mathbf{r}_k^{\mathrm{T}} / n = \begin{bmatrix} \overline{x^2} & \overline{xy} & \overline{xz} \\ \overline{xy} & \overline{y_k^{\,2}} & \overline{yz} \\ \overline{xz} & \overline{yz} & \overline{z^2} \end{bmatrix} = \mathbf{M} \tag{3.2.23}$$

We can put

$$\mathbf{M} = \begin{bmatrix} M_{xx} & M_{xy} & M_{xz} \\ M_{xy} & M_{yy} & M_{yz} \\ M_{xz} & M_{yz} & M_{zz} \end{bmatrix}$$

where

$$M_{uu} = \sum u_k^{\,2} / n \text{ and } M_{uv} = \sum u_k v_k / n \tag{3.2.24}$$

and u, v are any pair of x, y and z. As the element coordinates are with respect to the mean positions, in this case, we see that $\mathbf{M}$ is a real 3×3 matrix of second moments about the mean; in effect it is a covariance matrix of the array element positions.

[In the frequently encountered case of a planar array, with x and y axes in the plane, all the z components are zero and we have

$$\mathbf{M} = \begin{bmatrix} M_{xx} & M_{xy} & 0 \\ M_{xy} & M_{yy} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

In the case of a linear array, with the elements along the x axis, the only non-zero element of $\mathbf{M}$ is M_{xx} and we simply have $\dot{\mathbf{e}}(\alpha)^{\mathrm{T}} \mathbf{M} \dot{\mathbf{e}}(\alpha) = M_{xx} \dot{e}_1(\alpha)^2$.]

Using (3.2.23) the denominator of (3.2.10), as given in (3.2.22), becomes

$$\left\| \mathbf{d} \right\|^2 = 4\pi^2 \dot{\mathbf{e}}^{\mathrm{T}} \mathbf{M} \dot{\mathbf{e}}. \tag{3.2.25}$$

For the variance of the error $\Delta\alpha$ we first express the numerator of (3.2.10) in terms of the moment matrix. We have, for the components of the numerator, from (3.2.21),

$$d_k * \Delta\alpha_k = -2\pi i \mathbf{r}_k^{\mathrm{T}} \dot{\mathbf{e}} a_k * \Delta\alpha_k$$

and hence

$$\mathrm{Re}(d_k * \Delta\alpha_k) = 2\pi \mathbf{r}_k^{\mathrm{T}} \dot{\mathbf{e}}\, \mathrm{Im}(a_k * \Delta\alpha_k) = 2\pi \mathbf{r}_k^{\mathrm{T}} \dot{\mathbf{e}} y_k, \tag{3.2.26}$$

where $y_k = \mathrm{Im}(a_k * \Delta\alpha_k)$ as before.

As in §3.2.2 the variance of $u = \mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}) = \mathrm{Re}(\sum_k d_k * \Delta\alpha_k) = 2\pi\sum_k c_k y_k = 2\pi\mathbf{c}^{\mathrm{T}}\mathbf{y}$ is given by

$$\mathrm{var}(u) = \left\langle u^2 \right\rangle = 4\pi^2 \left\langle \mathbf{c}^{\mathrm{T}}\mathbf{y}\mathbf{y}^{\mathrm{T}}\mathbf{c} \right\rangle = 4\pi^2 \mathbf{c}^{\mathrm{T}} \left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle \mathbf{c}$$

where $c_k = \mathbf{r}_k{}^{\mathrm{T}}\dot{\mathbf{e}}$ in this case, and so $\mathbf{c} = \mathbf{R}^{\mathrm{T}}\dot{\mathbf{e}}$. As before, from Appendix 3.2D, $\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle = \sigma_\phi{}^2\mathbf{I}/n^2$ so that

$$\mathrm{var}(u) = 4\pi^2(\sigma_\phi{}^2/n^2)\mathbf{c}^{\mathrm{T}}\mathbf{c} \tag{3.2.27}$$

Now we have, using the definition of **M** in (3.2.23),

$$\mathbf{c}^{\mathrm{T}}\mathbf{c} = \sum_k c_k{}^2 = \sum_k \dot{\mathbf{e}}^{\mathrm{T}}\mathbf{r}_k\mathbf{r}_k{}^{\mathrm{T}}\dot{\mathbf{e}} = n\dot{\mathbf{e}}^{\mathrm{T}}\mathbf{M}\dot{\mathbf{e}} \tag{3.2.28}$$

so using (3.2.28) in (3.2.27) we obtain

$$\mathrm{var}(\mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a})) = 4\pi^2\sigma_\phi{}^2\dot{\mathbf{e}}^{\mathrm{T}}\mathbf{M}\dot{\mathbf{e}}/n\,. \tag{3.2.29}$$

Finally, from (3.2.10), we have for the standard deviation of the error, using (3.2.25) and (3.2.29),

$$s.d.\ \Delta\alpha = \frac{\sqrt{\mathrm{var}(\mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}))}}{\|\mathbf{d}\|^2} = \frac{\sigma_\phi}{2\pi\sqrt{n\dot{\mathbf{e}}^{\mathrm{T}}\mathbf{M}\dot{\mathbf{e}}}}\,. \tag{3.2.30}$$

[We can express **M** in an alternative, and convenient form. We have, from the definition in (3.2.23),

$$n\mathbf{M} = \sum_k \mathbf{r}_k\mathbf{r}_k{}^{\mathrm{T}} = \mathbf{r}_1\mathbf{r}_1{}^{\mathrm{T}} + \mathbf{r}_2\mathbf{r}_2{}^{\mathrm{T}} + \ \ldots\ \mathbf{r}_n\mathbf{r}_n{}^{\mathrm{T}} = \begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_n\end{bmatrix}\begin{bmatrix}\mathbf{r}_1{}^{\mathrm{T}} \\ \mathbf{r}_2{}^{\mathrm{T}} \\ \vdots \\ \mathbf{r}_n{}^{\mathrm{T}}\end{bmatrix} = \mathbf{R}\mathbf{R}^{\mathrm{T}} \tag{3.2.31}$$

where

$$\mathbf{R} = \begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_n\end{bmatrix}, \tag{3.2.32}$$

and we see that **R** is the 3×*n* matrix of the element positions. It follows from $\mathbf{c} = \mathbf{R}^{\mathrm{T}}\dot{\mathbf{e}}$, given above, that we could also derive (3.2.28) in the alternative form

$$\mathbf{c}^{\mathrm{T}}\mathbf{c} = \|\mathbf{c}\|^2 = \left\|\mathbf{R}^{\mathrm{T}}\dot{\mathbf{e}}\right\|^2 = \dot{\mathbf{e}}^{\mathrm{T}}\mathbf{R}\mathbf{R}^{\mathrm{T}}\dot{\mathbf{e}} = n\dot{\mathbf{e}}^{\mathrm{T}}\mathbf{M}\dot{\mathbf{e}}\,.] \tag{3.2.33}$$

3.2.4 Simulation results

In this section we present some simulation results, confirming the theoretical expression above for small values of the phase error variance σ_ϕ, and in fact for higher values than might have been expected, and showing at what point the theoretical expression fails and why. We take a uniform linear array of n elements at element separation 0.5 wavelengths, and a target at 20º from the array normal. For the simulation results the standard deviation of the errors was determined over a large number of trials (generally 1000) and for each trial a new set of element gain errors was taken from a zero mean normal distribution with variance σ_ϕ^2 (for both amplitude and phase components, as discussed in Appendix 3.1A, though in fact only phase errors affect the performance in the single signal case, both theoretically and as confirmed by simulation). For each trial the signal, centred at 20º for these cases, was jittered about this figure (over a range of 3º) to avoid any bias due to a particular choice. To find the peak of the beamformed function the function was plotted relatively coarsely (at 0.45º - 400 points over a 180º interval) and quadratic interpolation was used to give an accurate estimate of the peak position. For all the plots the blue curves show the simulation results and the red curves the theoretical results using the expression in (3.2.18) above. (In these figures the red traces often disappear, being over-written by the blue traces when the simulation results are very close to the theoretical.)

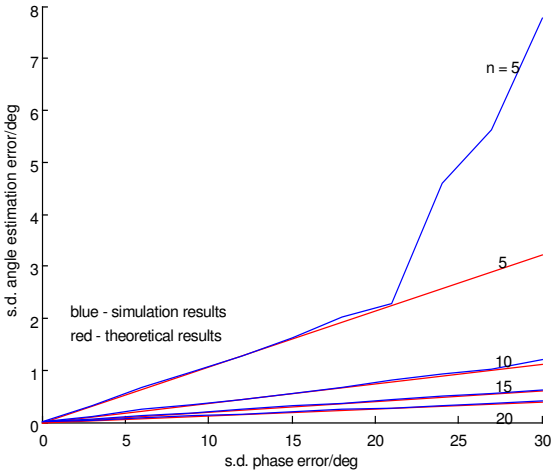


Figure 3.2 DF error as a function of element phase error

Figure 3.2 shows the statistical dependence of the DF errors on the phase errors, with the number of array elements as the parameter for a set of results. As expected the errors are linearly related for small errors, and even for what may be considered moderately large errors (above 10º rms), but when large enough the errors suddenly increase in an erratic way. However, we note that higher errors can be tolerated before this happens if there are more elements in the array. This is also shown in Figure 3.3, where the plotting variable is the number of elements, with the standard deviation of the phase error as parameter. In this case the dependence is not linear (but close to $n^{-3/2}$). This

shows more clearly the effect of increasing n on reducing the break away from the theoretical curves. We note that at 20º rms error the simulation results match closely the theoretical ones for a value of n as low as 5, but at 30º the erratic results occur at n as high as 10.

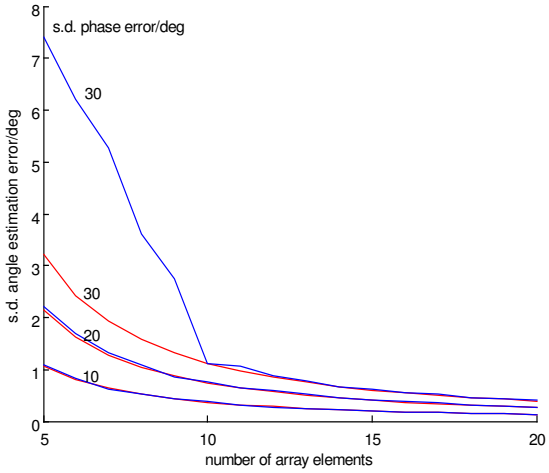
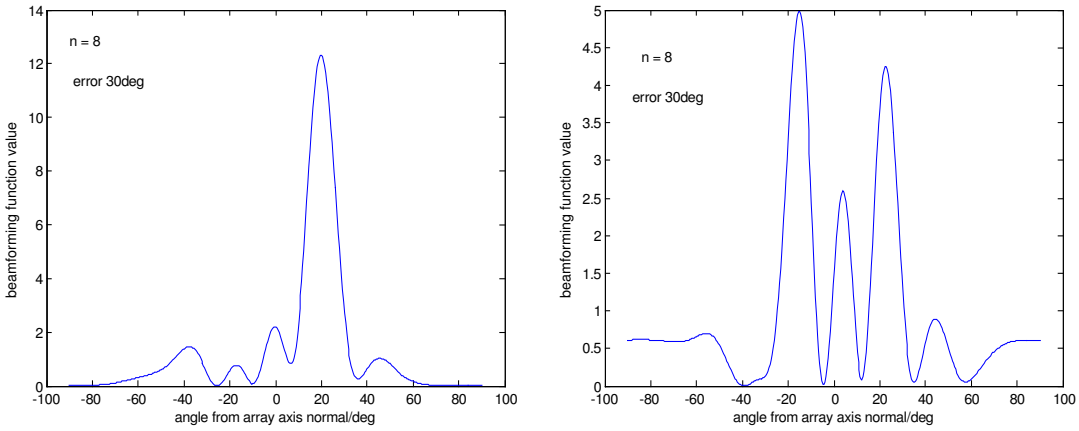


Figure 3.3 DF error as a function of number of elements

The cause of the increased DF error values is illustrated in Figure 3.4. In this case we have an array of eight elements and rms element phase errors of 30 degrees. The array beamwidth in this case is about 15.2º (from Appendix 3.2F) and errors in the position of the main beam should normally be well within this value (Figure 3.4(a)). However when the element errors are large enough the pattern can be distorted to the extent that a sidelobe exceeds the main lobe. In this case the error is not due to a perturbation in the position of the main lobe (Figure 3.4(a)) but to the choice of the wrong lobe (Figure 3.4(b)). This is termed here a gross error; the error in the position of the main lobe being a fine error. Gross errors, of course, are many times the magnitude of fine errors and greatly modify the rms DF error result.



(a) fine error (b) gross error

Figure 3.4 Fine and gross errors (signal at 20 deg.)

However, at the levels of element error taken here, they are relatively infrequent; for example for 10 elements with 30º rms errors, the incidence is only about 0.1%, for 8 elements it is 0.4%, and even for 5 elements it is only 2.1%.

We note, with the scaling used in the simulation, the peak height in the absence of errors is *n*, or 8, in this case. We see that considerable distortion results from errors at the high levels taken here – 30degrees rms in phase. In Figure 3.4(a) we see how the errors have augmented the main lobe, which is now of magnitude 12, while in Figure 3.4(b) the errors have reduced the main lobe (near 20º) to only 4.3, and below the augmented sidelobe near -20º, resulting in a gross error of nearly 40º. This error mechanism, causing gross errors, also explains why increasing *n* reduces the likelihood of gross errors. At higher values of *n* the ratio of the peak to sidelobe levels increases, so requiring more distortion to cause a sidelobe to exceed the main lobe.

The above results show that equation (3.2.18), which gives the s.d. for a linear array, is confirmed by simulations. For a more general array we need to use equation (3.2.30), where **M** is the second moment of the array element positions about the mean and $\dot{\mathbf{e}}$ is the derivative of the target direction vector with respect to the angle α from the array axis. The result, using an array of seven elements in the form of a regular heptagon, and scanning in the plane of the array, is shown in Figure 3.5. In this case the size of the array is the parameter for the set of plots. The elements of the basic array were on a circle of one wavelength radius. The array was also scaled, as indicated in the figure, by factors of two and three. Again the agreement between theory and simulation is very good. As the radius of the array increases the accuracy improves (the error s.d. falling as the reciprocal of the radius) as might be expected. However, we note that the linear relationship breaks down first with the largest array. This is because, as the array element separations increase the probability of near ambiguities – the appearance of other PSVs closely matching the actual target PSV – increases, and with errors present the probability of one of these matching the target PSV more closely than the PSV for the target direction (also corrupted by errors) increases as well.

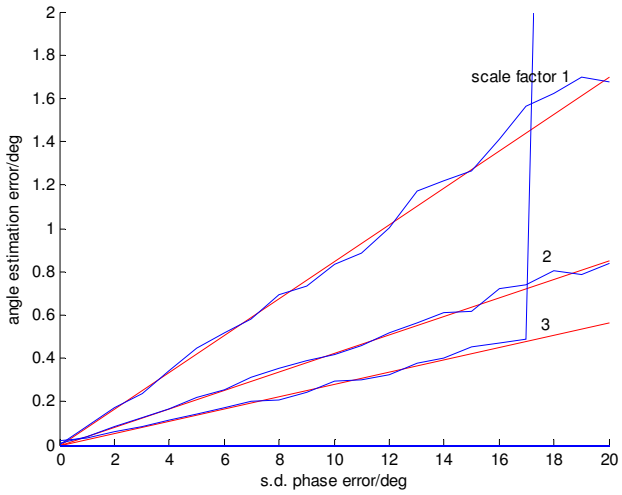


Figure 3.5 DF error for regular seven element circular array

3.2.5 Comparison with Cramér-Rao bound

The Cramer-Rao bound for a single parameter, α, of a single target, is given (from Chapter 2, equation (**2.**D.1)) by

$$\mathrm{var}(\alpha) = \frac{\psi_n}{2pa^2\mathbf{d}(\alpha)^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}(\alpha)}\mathbf{d}(\alpha)} \tag{3.2.34}$$

where ψ_n is the variance of the complex noise perturbing the measurement of α, a^2 is the mean square signal strength in an element of unit gain, p is the number of samples used (the degree of integration), $\mathbf{a}(\alpha)$ is the target PSV, $\mathbf{d}(\alpha)$ is its derivative and $\mathbf{Q}$ is the matrix projecting orthogonally into the space orthogonal to $\mathbf{a}$. Now we have

$$\mathbf{d}(\alpha)^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}(\alpha)}\mathbf{d}(\alpha) = \mathbf{d}(\alpha)^{\mathrm{H}}\left(\mathbf{I} - \mathbf{a}(\alpha)\mathbf{a}(\alpha)^{\mathrm{H}}\right)\mathbf{d}(\alpha) = \left\|\mathbf{d}(\alpha)\right\|^2 - \left|\mathbf{a}(\alpha)^{\mathrm{H}}\mathbf{d}(\alpha)\right|^2 = \left\|\mathbf{d}(\alpha)\right\|^2.$$

if we take $\mathbf{a}^{\mathrm{H}}\mathbf{d} = 0$, as above (§3.2.1). Now the derivative vectors $\mathbf{d}$ as used in the analysis of the effect of errors are based on vectors $\mathbf{a}$ which are normalized so that each component has amplitude $1/\sqrt{n}$ (in the equal, parallel pattern case) and this factor also applies to $\mathbf{d}$. In the CRB expression we take the elements to have unit gain (or else we have to redefine a, the signal amplitude) and in this case the factor $1/\sqrt{n}$ is not present in $\mathbf{a}$, or in its derivative $\mathbf{d}$, and so $\|\mathbf{d}\|^2$ in this case is a factor n higher and we replace $\left\|\mathbf{d}(\alpha)\right\|^2$ by $n\left\|\mathbf{d}(\alpha)\right\|^2$, where this $\mathbf{d}$ is as used in the error study. Thus we put (using (3.2.25))

$$\mathrm{var}(\alpha)_n = \frac{\psi_n}{2npa^2\left\|\mathbf{d}(\alpha)\right\|^2} = \frac{\psi_n}{8\pi^2 npa^2\dot{\mathbf{e}}^{\mathrm{T}}\mathbf{M}\dot{\mathbf{e}}}. \tag{3.2.35}$$

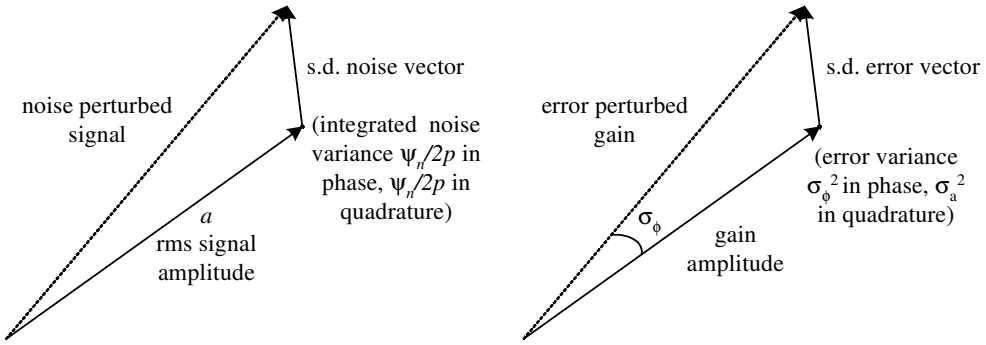
where $\mathrm{var}(\alpha)_n$ means the variance in the DF estimate of α when limited only by receiver noise. The variance in α due to element errors alone ($\mathrm{var}(\alpha)_e$) is given, from (3.2.30), as

$$\mathrm{var}(\alpha)_e = \frac{\sigma_\phi{}^2}{4\pi^2 n\dot{\mathbf{e}}^{\mathrm{T}}\mathbf{M}\dot{\mathbf{e}}}. \tag{3.2.36}$$

(In the linear array case (not necessarily regular) the results are given, from (3.2.14), by replacing $\dot{\mathbf{e}}^{\mathrm{T}}\mathbf{M}\dot{\mathbf{e}}$ by $\cos\alpha^2 M$.)

There is a close correspondence between these two results which can be explained in the sense that the noise perturbation on the signal in an element can be seen to be equivalent to the error perturbation on the gain of the element. We note that ψ_n is the variance of the complex noise in a channel, but integration over p signal samples effectively reduces

the noise variance, relative to the signal level, by a factor of p. Thus we see that errors with variance ψ_ϕ, in the case of high S/N, limit the DF performance to the same level as an integrated signal to noise ratio of $a^2/(\psi_n/2p)$ in the case of negligible errors. The



(a) signal perturbed by noise (b) gain perturbed by error

Figure 3.6 Noise and error perturbations

factor 2 may be accounted for by considering that ψ includes both in-phase and quadrature noise. If we assume that only the quadrature component is significant, as shown in the error study, then the effective noise level is $\psi/2$. The effects of these two forms of error is illustrated in Figure 3.6.

We see that the variances of the two forms of error giving the same DF accuracy are related by

$$\text{(integrated S/N)} = pa^2/\psi_n = 1/2\sigma_\phi^2. \qquad (3.2.37)$$

As an example, if we take the case of a signal at a level of 20dB in an element, with integration over 10 samples, then the integrated signal to noise ratio is 1000, which is equivalent to a gain error variance of $\psi_\phi = 1/2000$ rad^2, or a standard deviation phase error of $1/\sqrt{2000}$ radians or about 1.3 degrees. Alternatively, as another example, for a s.d. phase error of 5° the equivalent integrated signal to noise ratio is $1/2(5\pi/180)^2 = 65$, or about 18dB, for single

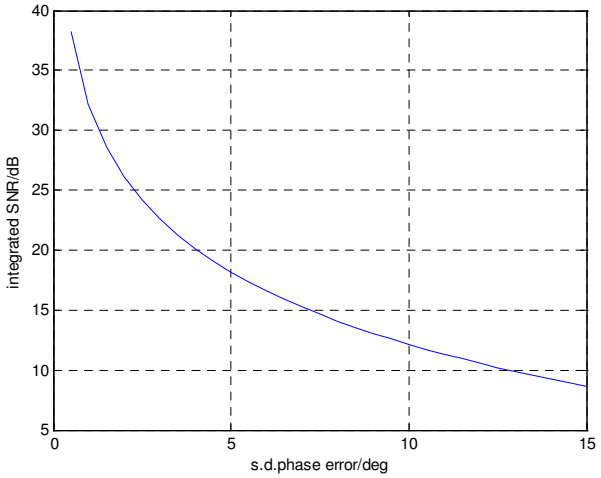


Figure 3.7 Equivalent integrated SNR limit

sample processing, or 0dB using 65 samples. Expressing this differently, we could say that, however strong the signal might be, channel phase errors with 5º s.d. will limit the accuracy to that of an error free system with an integrated SNR of 18dB. The relation between integrated signal to noise ratio and phase error variance is given in Figure 3.7.

Although it has not been proved, it seems reasonable to estimate that the parameter error variance will be the sum of the variances of the errors due to noise and array errors on their own, noting that these sources of error are independent.

Appendix 3.2A: Auxiliary results for linear array

The elements of the manifold vector **a** and its first derivative **d** are given in (3.2.20) and (3.2.21) and so we have (as $|a_k|^2 = 1/n$),

$$\mathbf{a}^{\mathrm{H}}\mathbf{d} = \sum\nolimits_k a_k{}^* d_k = \sum\nolimits_k 2\pi i \mathbf{r}_k{}^{\mathrm{T}}\dot{\mathbf{e}}/n \tag{3.2A.1}$$

and we see that (as $\mathbf{r}_k$ and $\dot{\mathbf{e}}$ are real)

$$\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{d}) = 0\,. \tag{3.2A.2}$$

(In fact we note that we can put (3.2A.1) in the form

$$\mathbf{a}^{\mathrm{H}}\mathbf{d} = 2\pi i\left(\sum\nolimits_k \mathbf{r}_k{}^{\mathrm{T}}/n\right)\dot{\mathbf{e}} = 2\pi i\bar{\mathbf{r}}^{\mathrm{T}}\dot{\mathbf{e}}$$

where $\bar{\mathbf{r}}$ is the mean element position vector, or the unweighted centroid. If this point is used as the reference for the element position specification then $\bar{\mathbf{r}} = \mathbf{0}$ and so $\mathbf{a}^{\mathrm{H}}\mathbf{d} = 0$.)

The elements of **s**, the second derivative, are given from (3.2.21) by

$$s_k(\alpha) = \frac{dd_k(\alpha)}{d\alpha} = \frac{d^2 a_k(\alpha)}{d\alpha^2} = 2\pi i\mathbf{r}_k{}^{\mathrm{T}}\dot{\mathbf{e}}(\alpha)d_k(\alpha) + 2\pi i\mathbf{r}_k{}^{\mathrm{T}}\ddot{\mathbf{e}}(\alpha)a_k(\alpha)$$

$$= -\left(2\pi\mathbf{r}_k{}^{\mathrm{T}}\dot{\mathbf{e}}(\alpha)\right)^2 a_k(\alpha) + 2\pi i\mathbf{r}_k{}^{\mathrm{T}}\ddot{\mathbf{e}}(\alpha)a_k(\alpha)\,, \tag{3.2A.3}$$

where $\dot{\mathbf{e}}$ and $\ddot{\mathbf{e}}$ are the first and second derivatives of **e** with respect to α. From (3.2.20) and (3.2A.3)

$$\mathbf{a}^{\mathrm{H}}\mathbf{s} = \sum\nolimits_k a_k{}^* s_k = \sum\nolimits_k \left(-\left(2\pi\mathbf{r}_k{}^{\mathrm{T}}\dot{\mathbf{e}}\right)^2 + 2\pi i\mathbf{r}_k{}^{\mathrm{T}}\ddot{\mathbf{e}}\right)/n\,. \tag{3.2A.4}$$

and, with (3.2.22) in the form

$$\|\mathbf{d}\|^2 = \sum|d_k|^2 = \sum\nolimits_k \left(2\pi\mathbf{r}_k{}^{\mathrm{T}}\dot{\mathbf{e}}\right)^2/n \tag{3.2A.5}$$

we see that

$$\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{s}) = -\|\mathbf{d}\|^2 . \qquad (3.2A.6)$$

Appendix 3.2B: Array centroid for linear array

Let a linear array have elements at positions x_1 . . . x_n measured from some reference point on the axis of the array. Then the mean position of the elements is given by

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k .$$

This is the array centroid, in the coordinate system with origin the reference point. If we take the centroid as the new reference point, so that the elements are now at $x_k' = x_k - \bar{x}$, then the mean of the positions in this system is at zero, the new origin. This is the point about which the second moment is required, given in Appendix 3.2C. For a regular linear array the centroid is clearly the midpoint of the array.

In the case of unequal element gains what is required is what may be defined as the gain centroid, or 'centre of gain'. This is the centroid of the array where each element position is weighted by the element power gain. (This is equivalent to the centre of mass of the array, when the points of the array have a mass numerically equal to the power gain.) For an array of equal parallel pattern elements (all with the same gain in any given direction, but not necessarily with omnidirectional patterns) the gain centroid is the same as the simple geometrical centroid given above.

We note that if we take the gain centroid for the array reference position then we have the result $\mathbf{a}^{\mathrm{H}}\mathbf{d} = 0$, even when the element gains differ. In this case let

$$\mathbf{a}(\alpha) = [a_1 \quad a_2 \quad \ldots \quad a_n]^{\mathrm{T}} \text{ with } a_k(\alpha) = g_k \exp(2\pi i x_k \sin\alpha)$$

then

$$\mathbf{d}(\alpha) = [d_1 \quad d_2 \quad \ldots \quad d_n]^{\mathrm{T}}$$

with $d_k(\alpha) = 2\pi i x_k \cos\alpha g_k \exp(2\pi i x_k \sin\alpha)$.

(Here we consider g_k to be independent of α, or only slowly varying. We define g_k to be the element gain, (complex, in general) without the phase factor due to the displacement of the element from the array centroid position.) Forming $\mathbf{a}^{\mathrm{H}}\mathbf{d}$ we have

$$\mathbf{a}^{\mathrm{H}}\mathbf{d} = \sum_k a_k^* d_k = 2\pi i \cos\alpha \sum_k |g_k|^2 x_k .$$

By definition, $\sum_k |g_k|^2 x_k = 0$, when the x-values are measured from the gain centroid, and so $\mathbf{a}^{\mathrm{H}}\mathbf{d} = 0$, as required.

Appendix 3.2C: Second moment for uniform linear array

The positions of the elements of a regular linear array, with n elements at separation d are given by

$$\mathbf{x} = [-(n-1)d/2 \quad -(n-3)d/2 \ldots \ (n-3)d/2 \quad (n-1)d/2]^{\mathrm{T}}$$

where the positions are measured from the centroid. The second moment about the centroid is given by M where

$$nM = \mathbf{x}^{\mathrm{T}}\mathbf{x} = \sum_{k=1}^{n} x_k^2 = \left((n-1)d/2\right)^2 + \left((n-3)d/2\right)^2 + \ \ldots + \left((n-3)d/2\right)^2 + \left((n-1)d/2\right)^2 .$$

If n is odd we can put $n = 2m + 1$ so that

$$nM/d^2 = m^2 + (m-1)^2 + \ \ldots + 1 + 0 + 1 + \ldots \ + (m-1)^2 + m^2 .$$

$$= 2\sum_{k=1}^{m} k^2 = \frac{2m(m+1)(2m+1)}{6} = \frac{(n-1)((n+1)/2)n}{6} = \frac{n(n^2-1)}{12} .$$

If n is even we put $n = 2m$ to give

$$4nM = 2d^2(1^2 + 3^2 + \ \ldots \ + (2m-1)^2)$$

or $$2nM/d^2 = 1^2 + 2^2 + 3^2 + \ \ldots \ + (2m-1)^2 + (2m)^2 - \left(2^2 + 4^2 + \ldots \ + (2m)^2\right)$$

$$= \frac{2m(2m+1)(4m+1)}{6} - 4\frac{m(m+1)(2m+1)}{6} = \frac{2m(2m+1)(4m+1-2(m+1))}{6}$$

$$= \frac{2m(2m+1)(2m-1)}{6} = \frac{n(n^2-1)}{6} .$$

In both cases we have $M = \dfrac{(n^2-1)d^2}{12}$ and this is the result used in (3.2.15) for the gain second moment.

Appendix 3.2D: Variance of *y* variables

Let the amplitude of the kth component of the target PSV with error be given by $a_k(1 + u_k)$ where u_k is the fractional amplitude error and is a sample from a zero mean

distribution with variance σ_a^2. Then we can put $\Delta a_{kp} = a_k u_k$, where Δa_{kp} is as in Figure 3.1. Also let v_k be the phase error, from a zero mean distribution with variance σ_ϕ^2, then we have for the error Δa_{kq} in Figure 3.1, $\Delta a_{kq} = i a_k v_k$ (very closely, for small phase errors). The factor i is required to make Δa_{kq} orthogonal to a_k. Thus the total error is given by

$$\Delta a_k = (u_k + i v_k) a_k .$$

Then defining z_k by $z_k = a_k * \Delta a_k$, we see that

$$z_k = (u_k + i v_k)\left|a_k\right|^2 = (u_k + i v_k)\left|g_k\right|^2 \qquad (3.2D.1)$$

where $|a_k| = |g_k|$, using the definition given in Appendix 3.2B in the case of differing element gains. Thus we see that

$$y_k = \mathrm{Im}(a_k * \Delta a_k) = v_k \left|g_k\right|^2$$

so it has variance given by

$$\left\langle y_k^{\,2} \right\rangle = \left\langle v_k^{\,2} \right\rangle \left|g_k\right|^4 = \sigma_\phi^{\,2} \left|g_k\right|^4 .$$

For the value of $\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle$ used in §3.2.2 we note that this is an $n \times n$ matrix with elements

$$\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle_{jk} = \left\langle y_j y_k \right\rangle = \left\langle \mathrm{Im}(a_j * \Delta a_j)\ \ \mathrm{Im}(a_k * \Delta a_k) \right\rangle$$

$$= \left\langle \left(\mathrm{Re}(a_j)\,\mathrm{Im}(\Delta a_j) - \mathrm{Im}(a_j)\,\mathrm{Re}(\Delta a_j)\right)\left(\mathrm{Re}(a_k)\,\mathrm{Im}(\Delta a_k) - \mathrm{Im}(a_k)\,\mathrm{Re}(\Delta a_k)\right) \right\rangle$$

$$= \mathrm{Re}(a_j)\,\mathrm{Re}(a_k)\left\langle \mathrm{Im}(\Delta a_j)\,\mathrm{Im}(\Delta a_k) \right\rangle + \ \ldots\ - \ \ldots\ - \ \ldots + \ldots$$

Assuming the errors between channels are statistically independent, all terms of the form $\left\langle \mathrm{Im}(\Delta a_j)\,\mathrm{Im}(\Delta a_k) \right\rangle$, $\left\langle \mathrm{Re}(\Delta a_j)\,\mathrm{Re}(\Delta a_k) \right\rangle$ (for $j \neq k$) and $\left\langle \mathrm{Im}(\Delta a_j)\,\mathrm{Re}(\Delta a_k) \right\rangle$ (for all j and k) are zero and so all the off-diagonal terms $\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle_{jk}$ will be zero. Thus $\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle$ is a diagonal matrix with values on the principal diagonal given by $\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle_{kk} = \left\langle y_k^{\,2} \right\rangle = \mathrm{var}(y_k) = \left|g_k\right|^4 \sigma_\phi^{\,2}$ as given above, so we have

$$\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle = \sigma_\phi^{\,2}\, diag(\left|g_1\right|^4\ \ \left|g_2\right|^4\ \ \ldots\ \ \left|g_n\right|^4) . \qquad (3.2D.2)$$

If we put $\mathbf{G} = \mathrm{diag}(|g_1|^2\ \ |g_2|^2\ \ \ldots\ |g_n|^2)$ (as used in §3.3.4) then we have

$$\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle = \sigma_{\phi}^{2}\mathbf{G}^{2}. \tag{3.2D.3}$$

In the case of elements with equal, parallel patterns (equal gains in any given direction) we have $|g_k| = 1/\sqrt{n}$ for all elements so that, in this case we have $|g_k|^4 = 1/n^2$, and so

$$\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle = \left(\sigma_{\phi}^{2}/n^{2}\right)\mathbf{I}. \tag{3.2D.4}$$

where **I** here is the $n \times n$ identity matrix.

In Appendix 4D below (eq. (3.4D.4)) we also require $x_k = \mathrm{Re}(a_k^* \Delta a_k) = \mathrm{Re}(z_k)$ and we see from (3.2D.1) that this is given by $u_k|g_k|^2$ and so its variance is

or σ_a^2/n^2 in the case of equal, parallel pattern elements.

Appendix 3.2E: Derivation with general origin for array position

Here we derive the results for the numerator and denominator of (2.9) for the case of an array defined with respect to an origin not assumed to be the array centroid.

We note that $\sum_k \mathbf{r}_k/n = \bar{\mathbf{r}}$, the mean position of the array elements, or the centroid of the array, and so (3.2A.1) can be written as

$$\mathbf{a}^{\mathrm{H}}\mathbf{d} = 2\pi i \bar{\mathbf{r}}^{\mathrm{T}}\dot{\mathbf{e}} \tag{3.2E.1}$$

and we can put

$$\left|\mathbf{a}^{\mathrm{H}}\mathbf{d}\right|^{2} = 4\pi^{2}(\bar{\mathbf{r}}^{\mathrm{T}}\dot{\mathbf{e}})^{2} = 4\pi^{2}\dot{\mathbf{e}}^{\mathrm{T}}\bar{\mathbf{r}}\bar{\mathbf{r}}^{\mathrm{T}}\dot{\mathbf{e}}. \tag{3.2E.2}$$

Thus, using (3.2A.5) and (3.2E.2) the denominator of (3.2.9) can be put in the form

$$\|\mathbf{d}\|^{2} - \left|\mathbf{a}^{\mathrm{H}}\mathbf{d}\right|^{2} = 4\pi^{2}\dot{\mathbf{e}}^{\mathrm{T}}\left\{\left(\sum_k \mathbf{r}_k\mathbf{r}_k^{\mathrm{T}}/n\right) - \bar{\mathbf{r}}\bar{\mathbf{r}}^{\mathrm{T}}\right\}\dot{\mathbf{e}} = 4\pi^{2}\dot{\mathbf{e}}^{\mathrm{T}}\mathbf{M}\dot{\mathbf{e}} \tag{3.2E.3}$$

where

$$\mathbf{M} = \left(\sum_k \mathbf{r}_k\mathbf{r}_k^{\mathrm{T}}/n\right) - \bar{\mathbf{r}}\bar{\mathbf{r}}^{\mathrm{T}}. \tag{3.2E.4}$$

As in §3.2 above, **M** is a real 3×3 array second moment matrix given by

$$\mathbf{M} = \begin{bmatrix} M_{xx} & M_{xy} & M_{xz} \\ M_{xy} & M_{yy} & M_{yz} \\ M_{xz} & M_{yz} & M_{zz} \end{bmatrix} \tag{3.2E.5}$$

but in this case, where we have not taken $\bar{\mathbf{r}}$ to be zero, we have

$$M_{uu} = \overline{(u-\bar{u})^2} = \overline{u^2} - \bar{u}^2 = \sum u_k^{\,2}/n - \left(\sum u_k/n\right)^2$$

$$M_{uv} = \overline{(u-\bar{u})(v-\bar{v})} = \overline{uv} - \bar{u}\bar{v} = \sum u_k v_k/n - \left(\sum u_k/n\right)\left(\sum v_k/n\right)$$

where u and v represent x, y or z. As before, the elements of $\mathbf{M}$ are second moments about the mean x, y and z positions of the elements, and $\mathbf{M}$ can also be regarded as the covariance matrix of the element positions. We note that (3.2E.4), with $\mathbf{M}$ defined for the general case, is the same as in (3.2.23) for the particular case where $\bar{\mathbf{r}} = 0$.

We now consider the variance of the numerator of (3.2.9): $\mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}) + \mathrm{Re}((\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a})(\mathbf{a}^{\mathrm{H}}\mathbf{d}))$. From (3.2.26) we have

$$\mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}) = \mathrm{Re}\left(\sum_k d_k * \Delta a_k\right) = \sum_k 2\pi \mathbf{r}_k^{\mathrm{T}} \dot{\mathbf{e}} y_k \tag{3.2E.6}$$

Also we have $\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a} = \sum_k a_k * \Delta a_k$ so, using (3.2E.1),

$$\mathrm{Re}\left((\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a})(\mathbf{a}^{\mathrm{H}}\mathbf{d})\right) = \mathrm{Re}\left(2\pi i \bar{\mathbf{r}}^{\mathrm{T}}\dot{\mathbf{e}}\sum_k a_k * \Delta a_k\right) = -2\pi \bar{\mathbf{r}}^{\mathrm{T}}\dot{\mathbf{e}}\,\mathrm{Im}\left(\sum_k a_k * \Delta a_k\right),$$

$$= -2\pi \bar{\mathbf{r}}^{\mathrm{T}}\dot{\mathbf{e}}\sum_k y_k \tag{3.2E.7}$$

where $y_k = \mathrm{Im}(a_k * \Delta a_k)$ as in §2.2. Combining (3.2E.6) and (3.2E.7) we have, for the numerator,

$$\mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}) + \mathrm{Re}\left((\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a})(\mathbf{a}^{\mathrm{H}}\mathbf{d})\right) = \sum_k 2\pi(\mathbf{r}_k - \bar{\mathbf{r}})^{\mathrm{T}}\dot{\mathbf{e}} y_k \ . \tag{3.2E.8}$$

The variance of the numerator of (3.2.9), in (3.2E.8) above, is as in (3.2.27) and (3.2.28) with $\sum_k (\mathbf{r}_k - \bar{\mathbf{r}})(\mathbf{r}_k - \bar{\mathbf{r}})^{\mathrm{T}}$ replacing $\sum_k \mathbf{r}_k \mathbf{r}_k^{\mathrm{T}}$. However this new term is simply another form of $\mathbf{M}$ as given in (3.2E.4) above, so we have

$$\mathrm{var}\left(\mathrm{Re}(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}) + \mathrm{Re}((\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a})(\mathbf{a}^{\mathrm{H}}\mathbf{d}))\right) = 4\pi^2\sigma_\phi^{\,2}\dot{\mathbf{e}}^{\mathrm{T}}\mathbf{M}\dot{\mathbf{e}}/n$$

as in (3.2.29) for the $\bar{\mathbf{r}} = 0$ case.

Appendix 3.2F: Beamwidth of regular linear array

The gain of a regular linear array in direction α when steered in direction α_0 is

$$g(\alpha) = \sum_{k=-(n-1)/2}^{(n-1)/2} \exp\left(ik(\phi(\alpha) - \phi_0)\right)$$

where $\phi(\alpha) = 2\pi d\sin\alpha$ and $\phi_0 = 2\pi d\sin\alpha_0$. The inter-element spacing is d wavelengths. Now if

$$y = \sum_{k=-(n-1)/2}^{(n-1)/2} \exp(2ix) = \exp\left(-i(n-1)x\right) + \exp\left(-i(n-3)x\right) + \ \ldots\ \exp\left(i(n-1)x\right),$$

then $$\exp(ix)\, y = \exp\left(-i(n-2)x\right) + \exp\left(-i(n-4)x\right) + \ \ldots\ \exp(inx)$$

and $$\exp(-ix)\, y = \exp(-inx) + \exp\left(-i(n-2)x\right) + \ \ldots\ \exp\left(i(n-2)x\right)$$

Subtracting these equations gives

$$(\exp(ix) - \exp(-ix))\, y = \exp(inx) - \exp(-inx)$$

so $$2i\sin(x)\, y = 2i\sin(nx)$$

and $$y = \frac{\sin(nx)}{\sin(x)}.$$

In this case $x = (\phi(\alpha) - \phi_0)/2 = \pi d(\sin\alpha - \sin\alpha_0)$ and so

$$g(\alpha) = \frac{\sin\left(n\pi d(\sin\alpha - \sin\alpha_0)\right)}{\sin\left(\pi d(\sin\alpha - \sin\alpha_0)\right)}.$$

This has its maximum value (of n) when $d(\sin\alpha - \sin\alpha_0)$ is an integer in general, but the lobe of interest is when this integer is zero, i.e. when $\alpha = \alpha_0$ (or when $\alpha = 180^\circ - \alpha_0$, which is the lobe symmetric about the axis of the array). In this case the first zeros about the main lobe are when

$$nd(\sin\alpha - \sin\alpha_0) = \pm 1$$

i.e. $$\sin\alpha = \sin\alpha_0 \pm 1/nd.$$

If we put $\sin\alpha = \sin(\alpha_0 + \delta\alpha) = \sin\alpha_0\cos\delta\alpha + \cos\alpha_0\sin\delta\alpha \approx \sin\alpha_0 + \cos\alpha_0\delta\alpha$,

taking small angle approximations, then

$$\cos\alpha_0\delta\alpha \approx \pm 1/nd.$$

If we define the beamwidth as half the interval between the first zeros then the beamwidth is close to $1/nd\cos\alpha_0$.

3.3. PLANAR AND VOLUME ARRAYS

3.3.1 2D Angle error estimate

Although planar or volume arrays could be used to estimate signal direction in only one angle dimension, this would be rather under-using the array as the potential is available for finding the full 2-dimensional angle of the signal direction. Thus here we not only assume the array is more than 1-dimensional (i.e. not a linear array) but that we perform 2-dimensional signal direction estimation. We follow the derivation of the error variance in §3.2 above, for the measurement of the single angle α, but now expanded to the case of two angles, α and ε. In the next subsection, instead of a single error variance (or its square root, the standard deviation, as in §3.2.2) we obtain a covariance matrix, as for the expression for the Cramér-Rao bound in more than one parameter. Initially we take the case of equal, parallel pattern elements, as in §3.2.

Let the normalized manifold vector for direction (α,ε) be $\mathbf{a}(\alpha,\varepsilon)$, and the actual signal point response vector for this direction be $\tilde{\mathbf{a}}(\alpha,\varepsilon) = h(\mathbf{a}(\alpha,\varepsilon) + \Delta\mathbf{a})$ where $\Delta\mathbf{a}$ is the error added to $\mathbf{a}$ and h is a normalizing factor. Let the estimated signal direction, obtained by maximizing the scan function f over the manifold, be $\mathbf{a}(\alpha + \Delta\alpha,\varepsilon + \Delta\varepsilon)$ so that $\Delta\alpha$ and $\Delta\varepsilon$ are the errors in azimuth and elevation, respectively. Expanding $\mathbf{a}(\alpha + \Delta\alpha,\varepsilon + \Delta\varepsilon)$ about $\mathbf{a}(\alpha,\varepsilon)$, assuming $\Delta\alpha$ and $\Delta\varepsilon$ are small, we have, to second order,

$$\mathbf{a}(\alpha+\Delta\alpha,\varepsilon+\Delta\varepsilon) = k\left(\mathbf{a}(\alpha,\varepsilon) + \mathbf{a}_\alpha(\alpha,\varepsilon)\Delta\alpha + \mathbf{a}_\varepsilon(\alpha,\varepsilon)\Delta\varepsilon + \right.$$

$$\left. + \tfrac{1}{2}\mathbf{a}_{\alpha\alpha}(\alpha,\varepsilon)\Delta\alpha^2 + \mathbf{a}_{\alpha\varepsilon}(\alpha,\varepsilon)\Delta\alpha\Delta\varepsilon + \tfrac{1}{2}\mathbf{a}_{\varepsilon\varepsilon}(\alpha,\varepsilon)\Delta\varepsilon^2 + \ldots\right). \qquad (3.3.1)$$

As before we note that $|\mathbf{a}(\alpha,\varepsilon)|$ is unity for all α and ε, so the expansion will also be so, and hence the normalizing factor k will be unity. We use the notation $\mathbf{a}_\alpha$ and $\mathbf{a}_{\alpha\alpha}$ to indicate first and second derivatives with respect to α, and similarly for ε. The scan function is given by

$$f(\alpha+\Delta\alpha,\varepsilon+\Delta\varepsilon) = \left|\mathbf{a}(\alpha+\Delta\alpha,\varepsilon+\Delta\varepsilon)^{\mathrm{H}}\,\tilde{\mathbf{a}}(\alpha,\varepsilon)\right|^2$$

$$= h^2\left|(\mathbf{a} + \mathbf{a}_\alpha\Delta\alpha + \mathbf{a}_\varepsilon\Delta\varepsilon + \tfrac{1}{2}\mathbf{a}_{\alpha\alpha}\Delta\alpha^2 + \mathbf{a}_{\alpha\varepsilon}\Delta\alpha\Delta\varepsilon + \tfrac{1}{2}\mathbf{a}_{\varepsilon\varepsilon}\Delta\varepsilon^2 + \ldots)^{\mathrm{H}}(\mathbf{a} + \Delta\mathbf{a})\right|^2 \qquad (3.3.2)$$

where, for clarity, we drop the indication that the vector $\mathbf{a}$ and its derivatives are for the direction (α,ε) and this is assumed to be the case subsequently. Expanding (3.3.2) to second order, and considering the magnitude of $\Delta\mathbf{a}$ to be small, as well as $\Delta\alpha$ and $\Delta\varepsilon$

(so we neglect any terms including both $\Delta\mathbf{a}$ and $\Delta\alpha^2$, $\Delta\alpha\Delta\varepsilon$ or $\Delta\varepsilon^2$, as being of third order) we have, to this accuracy, (putting $\mathbf{a}^{\mathrm{H}}\mathbf{a} = 1$),

$$f(\alpha+\Delta\alpha,\varepsilon+\Delta\varepsilon) = h^2\left|1+\mathbf{a}_\alpha{}^{\mathrm{H}}\mathbf{a}\Delta\alpha+\mathbf{a}_\varepsilon{}^{\mathrm{H}}\mathbf{a}\Delta\varepsilon+\tfrac{1}{2}\mathbf{a}_{\alpha\alpha}{}^{\mathrm{H}}\mathbf{a}\Delta\alpha^2+\mathbf{a}_{\alpha\varepsilon}{}^{\mathrm{H}}\mathbf{a}\Delta\alpha\Delta\varepsilon+\tfrac{1}{2}\mathbf{a}_{\varepsilon\varepsilon}{}^{\mathrm{H}}\mathbf{a}\Delta\varepsilon^2+\right.$$

$$\left.\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a}+\mathbf{a}_\alpha{}^{\mathrm{H}}\Delta\mathbf{a}\Delta\alpha+\mathbf{a}_\varepsilon{}^{\mathrm{H}}\Delta\mathbf{a}\Delta\varepsilon\right|^2. \qquad (3.3.3)$$

We take the array centroid as the reference for the array element positions, as the accuracy of measurement must be independent of the reference point, as shown before, and in this case we have $\mathbf{a}_\alpha{}^{\mathrm{H}}\mathbf{a} = 0$ (see (3.3A.4)) and $\mathbf{a}_\varepsilon{}^{\mathrm{H}}\mathbf{a} = 0$. Then (3.3.3) becomes, taking terms up to second order only,

$$f(\alpha+\Delta\alpha,\varepsilon+\Delta\varepsilon) = h^2\,|1+\tfrac{1}{2}\mathbf{a}_{\alpha\alpha}{}^{\mathrm{H}}\mathbf{a}\Delta\alpha^2+\mathbf{a}_{\alpha\varepsilon}{}^{\mathrm{H}}\mathbf{a}\Delta\alpha\Delta\varepsilon+\tfrac{1}{2}\mathbf{a}_{\varepsilon\varepsilon}{}^{\mathrm{H}}\mathbf{a}\Delta\varepsilon^2+$$

$$+\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a}+\mathbf{a}_\alpha{}^{\mathrm{H}}\Delta\mathbf{a}\Delta\alpha+\mathbf{a}_\varepsilon{}^{\mathrm{H}}\Delta\mathbf{a}\Delta\varepsilon\,|^2$$

$$= h^2\left(1+\mathrm{Re}(\mathbf{a}_{\alpha\alpha}{}^{\mathrm{H}}\mathbf{a})\Delta\alpha^2+2\,\mathrm{Re}(\mathbf{a}_{\alpha\varepsilon}{}^{\mathrm{H}}\mathbf{a})\Delta\alpha\Delta\varepsilon+\mathrm{Re}(\mathbf{a}_{\varepsilon\varepsilon}{}^{\mathrm{H}}\mathbf{a})\Delta\varepsilon^2+\right.$$

$$\left.+2\,\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a})+\left|\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a}\right|^2+2\,\mathrm{Re}(\mathbf{a}_\alpha{}^{\mathrm{H}}\Delta\mathbf{a})\Delta\alpha+2\,\mathrm{Re}(\mathbf{a}_\varepsilon{}^{\mathrm{H}}\Delta\mathbf{a})\Delta\varepsilon\right). \qquad (3.3.4)$$

Putting this in the form

$$f(\alpha+\Delta\alpha,\varepsilon+\Delta\varepsilon) = A+2B\Delta\alpha+2C\Delta\varepsilon+D\Delta\alpha^2+2E\Delta\alpha\Delta\varepsilon+F\Delta\varepsilon^2+\ \ldots \qquad (3.3.5)$$

then finding the peak with respect to both variables we have, from $\partial f/\partial\Delta\alpha = 0$,

$$2B+2D\Delta\alpha+2E\Delta\varepsilon = 0$$

and from $\partial f/\partial\Delta\varepsilon = 0$,

$$2C+2E\Delta\alpha+2F\Delta\varepsilon = 0\,.$$

Combining these two linear equations in a vector-matrix equation we have

$$\begin{bmatrix} D & E \\ E & F \end{bmatrix}\begin{bmatrix} \Delta\alpha \\ \Delta\varepsilon \end{bmatrix} = -\begin{bmatrix} B \\ C \end{bmatrix} \qquad (3.3.6)$$

or

$$\begin{bmatrix} \Delta\alpha \\ \Delta\varepsilon \end{bmatrix} = \mathbf{U}^{-1}\mathbf{u} \qquad (3.3.7)$$

where

$$-\mathbf{U} = \begin{bmatrix} D & E \\ E & F \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(\mathbf{a}_{\alpha\alpha}{}^{\mathrm{H}}\mathbf{a}) & \mathrm{Re}(\mathbf{a}_{\alpha\varepsilon}{}^{\mathrm{H}}\mathbf{a}) \\ \mathrm{Re}(\mathbf{a}_{\alpha\varepsilon}{}^{\mathrm{H}}\mathbf{a}) & \mathrm{Re}(\mathbf{a}_{\varepsilon\varepsilon}{}^{\mathrm{H}}\mathbf{a}) \end{bmatrix} \tag{3.3.8}$$

and

$$\mathbf{u} = \begin{bmatrix} B \\ C \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(\mathbf{a}_{\alpha}{}^{\mathrm{H}}\Delta\mathbf{a}) \\ \mathrm{Re}(\mathbf{a}_{\varepsilon}{}^{\mathrm{H}}\Delta\mathbf{a}) \end{bmatrix} \tag{3.3.9}$$

on substituting for B, C, D, E and F from (3.3.4) and (3.3.5). Using the results in Appendix 3.3A (equations (3.3A.7) etc.) we have

$$\mathbf{U} = 4\pi^2 \begin{bmatrix} \mathbf{e}_{\alpha}{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_{\alpha} & \mathbf{e}_{\alpha}{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_{\varepsilon} \\ \mathbf{e}_{\varepsilon}{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_{\alpha} & \mathbf{e}_{\varepsilon}{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_{\varepsilon} \end{bmatrix} = 4\pi^2 \begin{bmatrix} \mathbf{e}_{\alpha}{}^{\mathrm{T}} \\ \mathbf{e}_{\varepsilon}{}^{\mathrm{T}} \end{bmatrix} \mathbf{M} \begin{bmatrix} \mathbf{e}_{\alpha} & \mathbf{e}_{\varepsilon} \end{bmatrix} = 4\pi^2 \mathbf{E}^{\mathrm{T}}\mathbf{M}\mathbf{E} \tag{3.3.10}$$

if we put $\mathbf{E} = \begin{bmatrix} \mathbf{e}_{\alpha} & \mathbf{e}_{\varepsilon} \end{bmatrix}$, the matrix of the two partial derivatives of $\mathbf{e}$ with respect to the angle variables. Also, from (3.3A.8),

$$\mathbf{u} = 2\pi \begin{bmatrix} \mathbf{e}_{\alpha}{}^{\mathrm{T}}\mathbf{R}\mathbf{y} \\ \mathbf{e}_{\varepsilon}{}^{\mathrm{T}}\mathbf{R}\mathbf{y} \end{bmatrix} = 2\pi \begin{bmatrix} \mathbf{e}_{\alpha}{}^{\mathrm{T}} \\ \mathbf{e}_{\varepsilon}{}^{\mathrm{T}} \end{bmatrix} \mathbf{R}\mathbf{y} = 2\pi \mathbf{E}^{\mathrm{T}}\mathbf{R}\mathbf{y} \,. \tag{3.3.11}$$

3.3.2 2D Angle error covariance matrix

The covariance matrix for the azimuth and elevation estimates of the signal position is given by

$$\Psi = \left\langle \begin{bmatrix} \Delta\alpha \\ \Delta\varepsilon \end{bmatrix} \begin{bmatrix} \Delta\alpha & \Delta\varepsilon \end{bmatrix} \right\rangle = \left\langle \mathbf{U}^{-1}\mathbf{u}\mathbf{u}^{\mathrm{T}}\mathbf{U}^{-\mathrm{T}} \right\rangle = \mathbf{U}^{-1} \left\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \right\rangle \mathbf{U}^{-1} \tag{3.3.12}$$

where $\langle v \rangle$ means the expectation of a variable v, and we have used (3.3.7), and $(\mathbf{U}^{-1}\mathbf{u})^{\mathrm{T}} = \mathbf{u}^{\mathrm{T}}\mathbf{U}^{-\mathrm{T}}$, $\mathbf{U} = \mathbf{U}^{\mathrm{T}}$ (from (3.3.8)) and the fact that $\mathbf{U}$ is not considered to be a statistical variable here. Using (3.3.11) we have

$$\left\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \right\rangle = 4\pi^2 \left\langle \mathbf{E}^{\mathrm{T}}\mathbf{R}\mathbf{y}\mathbf{y}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{E} \right\rangle = 4\pi^2 \mathbf{E}^{\mathrm{T}}\mathbf{R} \left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle \mathbf{R}^{\mathrm{T}}\mathbf{E} \,. \tag{3.3.13}$$

From Appendix 3.2D we have

$$\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle = \left(\sigma_{\phi}{}^2 / n^2 \right) \mathbf{I} \tag{3.3.14}$$

where $\sigma_\phi^{\,2}$ is the phase variance of the errors, which are assumed to be equal for each element and independent between elements. **I** is the unit matrix of size *n*. Putting (3.3.14) into (3.3.13) we obtain

$$\left\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \right\rangle = \frac{4\pi^2 \sigma_\phi^{\,2}}{n^2} \mathbf{E}^{\mathrm{T}} \mathbf{R} \mathbf{R}^{\mathrm{T}} \mathbf{E} \,. \tag{3.3.15}$$

Using (3.2.31) and (3.3.10) we have

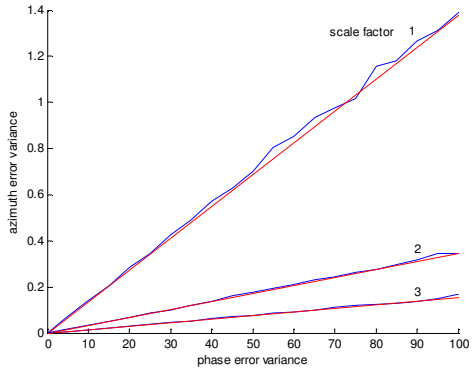
$$\left\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \right\rangle = \frac{4\pi^2 \sigma_\phi^{\,2}}{n} \mathbf{E}^{\mathrm{T}} \mathbf{M} \mathbf{E} = \frac{\sigma_\phi^{\,2}}{n} \mathbf{U} \,. \tag{3.3.16}$$

From (3.3.12) and (3.3.16) we have, finally

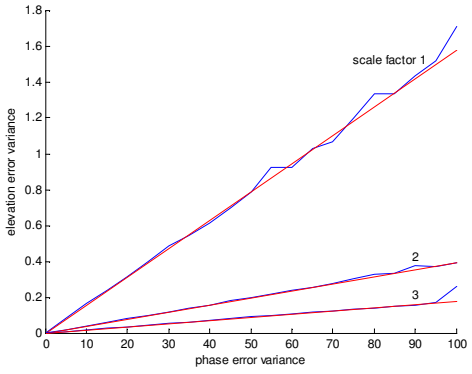
$$\Psi = \frac{\sigma_\phi^{\,2}}{n} \mathbf{U}^{-1} = \frac{\sigma_\phi^{\,2}}{4\pi^2 n} \left(\mathbf{E}^{\mathrm{T}} \mathbf{M} \mathbf{E} \right)^{-1} . \tag{3.3.17}$$

3.3.3 Simulation results (equal parallel antenna patterns)

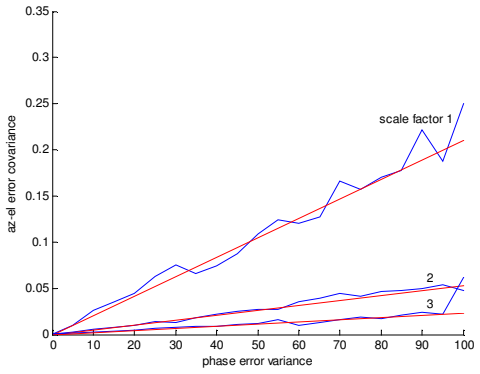
In this case we have a 2×2 theoretical error covariance matrix for comparison with the simulation results. In general the variance of the azimuth and elevation errors are of



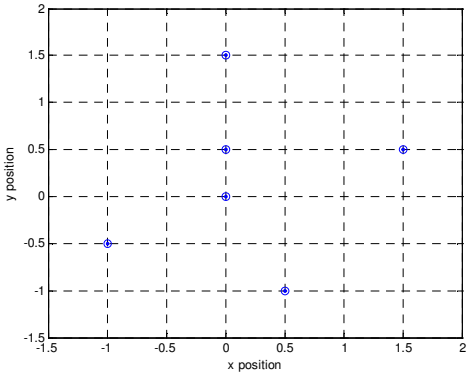
azimuth error variance



elevation error variance



azimuth-elevation error covariance



array R1

Figure 3.8 Error covariance results for array R1, target direction (30deg,40deg)

primary interest (given by the diagonal elements Ψ(1,1) and Ψ(2,2) in (3.3.17)) but the covariance of the azimuth and elevation errors (Ψ(1,2) or Ψ(2,1)) is available and we also found this in the simulation. Thus we have plotted the elements of Ψ, the variances and covariance of the errors, rather than the standard deviations (the square roots of the variances) as in §3.2.4 above. (The covariance may be positive or negative so the square root of this is not necessarily real.) The result, for the irregular 6 element planar array R1, is given in Figure 3.8 (where the element positions of R1 are also shown). For this case the target was at 30° in azimuth and 40º in elevation and 4000 trials were taken at each point. The variances and covariances are all in units of degrees squared, and we see that the input error variance range goes up to 100deg^2 or a standard deviation of 10deg. Again we use an array scaling factor as a parameter for a set of curves, confirming the relation that the error variance is proportional to the inverse square of the array size. We see good agreement between the theoretical relationship (given by (3.3.17)) and the simulation results, including for the covariance figure which gives a measure of the degree to which the two errors are linked.

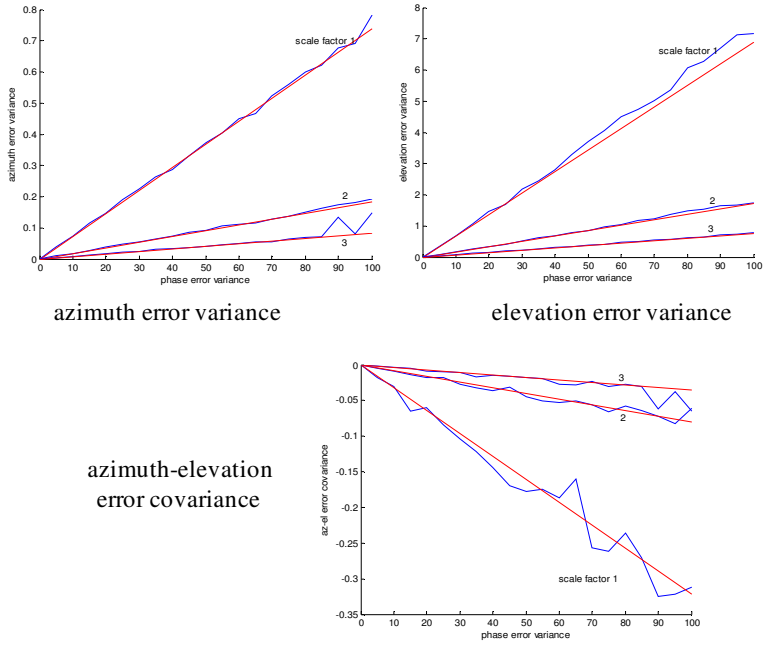


Figure 3.9 Error covariance results for array R1, target direction (120deg,20deg)

Figure 3.9 is for the same array but a different signal direction. Whereas in Figure 3.8 the azimuth and elevation errors are of comparable magnitude, in Figure 3.9 there is a factor of 10 between them. This may be partly due to the different azimuth direction, (the array being slightly narrower, from front to back, seen from this direction) and partly due to the lower target elevation, which reduces the aperture further, from the

point of view of the elevation measurement. In fact Figure 3.10 shows why the accuracy in elevation is so much lower in this case. This figure shows plots of the DF function (maximum value unity) in the region of the target position in the two cases, and for the

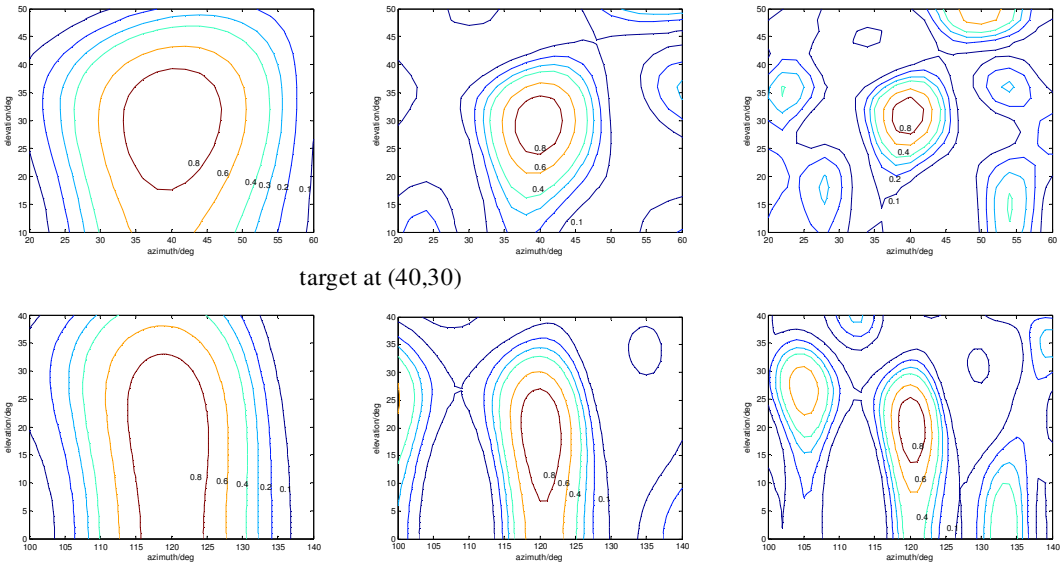
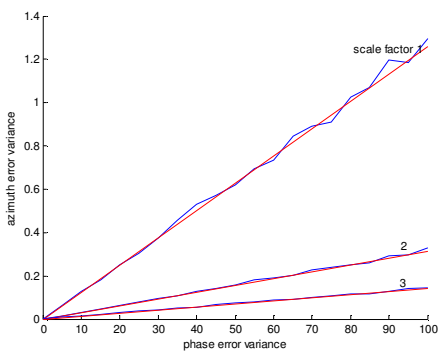


Figure 3.10 DF function plots for Array R1

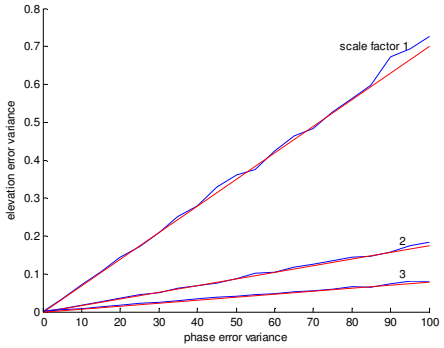
three scaling factors used. We see that for the target at (40º,30º) the function has a fairly circular peak, and the error sensitivity, shown in Figure 3.8, is similar for the two angle dimensions. However at (120º,20º) the peak is more of a ridge, with much greater extension in the elevation direction than in the azimuth direction. The sensitivity to errors would be expected to be much greater in elevation and this is shown in Figure 3.9. We also note that the pattern shrinks as the array is expanded, using the scaling factors, as expected. This brings more sidelobes into the pattern, as seen particularly with scale factor 3, and the increased probability of a gross error at high element error levels (if the search region is limited to that used for the plots).

Finally Figure 3.11 shows results for a 7-element array in the form of a regular heptagon on a circumcircle of one wavelength radius, but with variations in height (over about two wavelengths). The results for the azimuth error variance are seen to be quite close to those of the irregular 6-element array R1, with target at (40º,30º), shown in Figure 3.8. We note that the apertures of the two arrays in the horizontal plane are quite similar, both about 2 wavelengths, and this is a strong determinant of the error performance. The performance of the 3D heptagonal array in the measurement of elevation is much better than the planar array for the target at 120° (Figure 3.9), having a much greater vertical aperture.

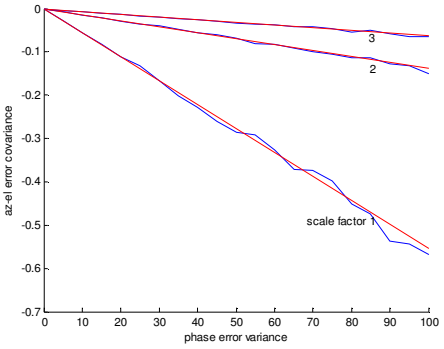
The plot of the function shown in Figure 3.11 shows a fairly circular peak, of about the same size as in Figure 3.10 for array R1 with target at (40º,30º) (and scale factor 3), so might be expected to give similar accuracy.



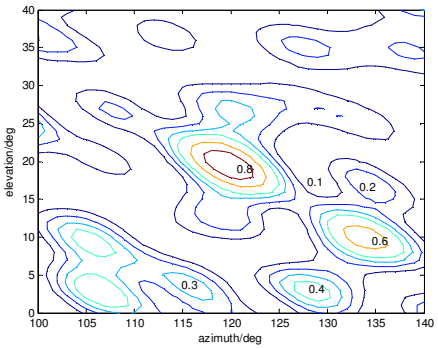
azimuth error variance



elevation error variance



azimuth-elevation error covariance



function plot (scale factor 3)

Figure 3.11 Error covariance results for heptagonal array, target at (120deg,20deg)

3.3.4 Non-parallel pattern case

Here we consider the case where the element gains are more general than for the case previously taken. In that case the shapes of the element gain patterns are supposed to be identical for all elements (‘parallel’) and also equally scaled (‘equal’). These conditions mean that, although the element patterns may vary with direction, for any given direction the gains of all the elements are equal. We now take the more general case where the element gains in any given direction may be different. This may be because the equal condition is removed (with the element patterns having the same shapes, so are still parallel) or because the parallel condition is removed. (The first of these seems to be physically unlikely, but an example might be where different levels of receiver gain are applied to the outputs of similar, equally aligned elements, so that different signal levels are received at the processor in different receiver channels.) In the second case the element patterns may be equally scaled but are directional and are not aligned; an example is a circular array of similar elements with directional patterns (perhaps

reflector-backed dipole elements) all directed outwards from the centre of the array. More generally still, all the element patterns may be different from each other; an example in this case might be elements mounted on poor sites (poor from the point of view of pattern integrity, such as elements mounted around a ship for HF DF operation). In all these cases we still have two remaining assumptions, if we are to be able to apply the results of the theory to estimate array DF accuracy. Firstly we suppose that the patterns are known, at least in a reasonably well-sampled form, and are given as complex gain factors (not as power gains). Secondly we suppose that the patterns are relatively slowly changing (in both amplitude and phase) so that they may be considered essentially constant over the range of the estimated accuracy. ('Relative' here means that we suppose that the change of gain with angle due to the element patterns is small compared with the effect of the phase shift due to the separation of the element phase centres.) This condition is generally satisfied if small, simple elements are used. (However, an exception is the case where pattern, rather than element displacement from the array centre, is the factor distinguishing the elements. An example of an array of this kind would be a three element array consisting of two orthogonally mounted loops (with figure-of-eight patterns) and a dipole (with an omnidirectional pattern) all cosited, with essentially coincident phase centres.) If strongly directional elements (requiring large apertures) are used, then these are at correspondingly larger separations and again the very rapid change of phase with angle due to the element separations will dominate the amplitude pattern. However an interesting case is that of an array which is partitioned so that a number of beams can be formed, using elements from all over the array for each beam, which may, from the choice of sets of elements, have phase centres near the array centre, so that we have effectively a small array of directional elements. In practice in this case the beams may be parallel or nearly so (a beam cluster) and the DF is applied in the area of the common, or central, look direction.

In this case the point response vector is given, as before, by

$$\mathbf{a}(\alpha,\varepsilon) = \left[a_1(\alpha,\varepsilon) \quad a_2(\alpha,\varepsilon) \quad . \quad . \quad . \quad a_n(\alpha,\varepsilon)\right] \tag{3.3.18}$$

but now the components of **a** are given by

$$a_k(\alpha,\varepsilon) = g_k \exp\left(2\pi i \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}(\alpha,\varepsilon)\right) \tag{3.3.19}$$

where g_k is the complex gain of the element, and the set of gains is normalized so that

$$\|\mathbf{a}\|^2 = \sum_k |g_k|^2 = 1. \tag{3.3.20}$$

(As stated above, g_k varies with angle but it is assumed only slowly compared with the phase factor, so this dependence is not indicated in (3.3.19)). From (3.3.19) and (3.3A.2) we have

$$\mathbf{a}^{\mathrm{H}}\mathbf{a}_\alpha = \sum_k a_k {}^{*}a_{\alpha k} = \sum_k 2\pi i \mathbf{r}_k{}^{\mathrm{T}}\mathbf{e}_\alpha \left|g_k\right|^2 . \tag{3.3.21}$$

In the previous analysis we took the array reference point to be the mean position of the elements, so that $\sum_k \mathbf{r}_k = \mathbf{0}$ and hence $\mathbf{a}^{\mathrm{H}}\mathbf{a}_\alpha = 0$, and similarly $\mathbf{a}^{\mathrm{H}}\mathbf{a}_\varepsilon = 0$. In this case, if we take the *weighted* mean of the element positions as the reference point, where the weighting of element k is by the factor $|g_k|^2$, the relative power gain of the element in the signal direction, then we have $\sum_k \left|g_k\right|^2 \mathbf{r}_k = \mathbf{0}$ (Appendix 3.3B) and so $\mathbf{a}^{\mathrm{H}}\mathbf{a}_\alpha = 0$ and $\mathbf{a}^{\mathrm{H}}\mathbf{a}_\varepsilon = 0$.

Similarly, in deriving **U** in §3.3.1, we have for D in (3.3.8), for example,

$$D = \mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{a}_{\alpha\alpha}) = \sum_k a_k {}^{*}a_{\alpha\alpha k} = -\sum_k \left(2\pi \mathbf{r}_k{}^{\mathrm{T}}\mathbf{e}_\alpha\right)^2 \left|g_k\right|^2 . \tag{3.3.22}$$

In this case we put

$$\sum_k \left(\mathbf{r}_k{}^{\mathrm{T}}\mathbf{e}_\alpha\right)^2 \left|g_k\right|^2 = \sum_k \mathbf{e}_\alpha{}^{\mathrm{T}}\mathbf{r}_k\mathbf{r}_k{}^{\mathrm{T}}\mathbf{e}_\alpha \left|g_k\right|^2 = \mathbf{e}_\alpha{}^{\mathrm{T}}\left(\sum_k \mathbf{r}_k\mathbf{r}_k{}^{\mathrm{T}}\left|g_k\right|^2\right)\mathbf{e}_\alpha = \mathbf{e}_\alpha{}^{\mathrm{T}}\mathbf{M}_1\mathbf{e}_\alpha$$

where $\mathbf{M}_1$ is now defined with the weights $|g_k|^2$ instead of $1/n$. Similar results follow for E and F so that **U** is given by (3.3.10) with $\mathbf{M}_1$, the weighted form of **M**, used in place of **M** as defined in (3.2.23) or (3.2.31). With $\mathbf{E} = \left[\mathbf{e}_\alpha \quad \mathbf{e}_\varepsilon\right]$ as before, **U** for the unequal gain case becomes

$$\mathbf{U}_1 = 4\pi^2 \begin{bmatrix} \mathbf{e}_\alpha{}^{\mathrm{T}} \\ \mathbf{e}_\varepsilon{}^{\mathrm{T}} \end{bmatrix} \mathbf{M}_1 \left[\mathbf{e}_\alpha \quad \mathbf{e}_\varepsilon\right] = 4\pi^2 \mathbf{E}^{\mathrm{T}}\mathbf{M}_1\mathbf{E} \tag{3.3.23}$$

We can put, for the weighted second moment matrix,

$$\mathbf{M}_1 = \sum_k \mathbf{r}_k\mathbf{r}_k{}^{\mathrm{T}}\left|g_k\right|^2 = \left[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \ldots \quad \mathbf{r}_n\right]\mathbf{G}\begin{bmatrix} \mathbf{r}_1{}^{\mathrm{T}} \\ \mathbf{r}_2{}^{\mathrm{T}} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{r}_n{}^{\mathrm{T}} \end{bmatrix} = \mathbf{R}\mathbf{G}\mathbf{R}^{\mathrm{T}} \tag{3.3.24}$$

where $\mathbf{G} = \mathrm{diag}(|g_1|^2 \;\; |g_2|^2 \;\; \ldots \;\; |g_n|^2)$, where the power gains have been normalized so

that $\sum_k |g_k|^2 = 1$ (see (3.3.20)).

In deriving $\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \rangle$ we require $\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \rangle$ which, if the errors are independent between channels, is given by $\sigma_\phi{}^2\mathbf{G}^2$, from (3.2D.3). Thus (3.3.13) becomes, for the unequal gain case,

$$\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \rangle = 4\pi^2\sigma_\phi{}^2\mathbf{E}^{\mathrm{T}}\mathbf{R}\mathbf{G}^2\mathbf{R}^{\mathrm{T}}\mathbf{E}.$$

We now put $\mathbf{M_2} = \mathbf{R}\mathbf{G}^2\mathbf{R}^{\mathrm{T}}$ where $\mathbf{M}_2$ is now another weighted second moment matrix, now with weights given by the *squares* of the element power gains. With similar results for the other matrix elements, we have

$$\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \rangle = 4\pi^2\sigma_\phi{}^2 \begin{bmatrix} \mathbf{e}_\alpha{}^{\mathrm{T}}\mathbf{M}_2\mathbf{e}_\alpha & \mathbf{e}_\alpha{}^{\mathrm{T}}\mathbf{M}_2\mathbf{e}_\varepsilon \\ \mathbf{e}_\varepsilon{}^{\mathrm{T}}\mathbf{M}_2\mathbf{e}_\alpha & \mathbf{e}_\varepsilon{}^{\mathrm{T}}\mathbf{M}_2\mathbf{e}_\varepsilon \end{bmatrix} = 4\pi^2\sigma_\phi{}^2\mathbf{E}^{\mathrm{T}}\mathbf{M}_2\mathbf{E} = \sigma_\phi{}^2\mathbf{U_2} \qquad (3.3.25)$$

where $\mathbf{E} = [\mathbf{e}_\alpha \quad \mathbf{e}_\varepsilon]$ as before, the matrix of the derivatives of the target direction vector.

From (3.3.12), (3.3.23) and (3.3.25) we have, for the error covariance matrix in the case of unequal element gains,

$$\Psi = \mathbf{U}^{-1}\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \rangle \mathbf{U}^{-1} = \sigma_\phi{}^2\mathbf{U_1}^{-1}\mathbf{U}_2\mathbf{U_1}^{-1}. \qquad (3.3.26)$$

We have put $\mathbf{U_1} = 4\pi^2\mathbf{E}^{\mathrm{T}}\mathbf{R}\mathbf{G}\mathbf{R}^{\mathrm{T}}\mathbf{E}$ and $\mathbf{U}_2 = 4\pi^2\mathbf{E}^{\mathrm{T}}\mathbf{R}\mathbf{G}^2\mathbf{R}^{\mathrm{T}}\mathbf{E}$ where **G** is given above (following (3.3.24)). If we put, for the equal gain case, **G** = **I**/*n* then, from (3.2.31) we see that $\mathbf{M}_1 = \mathbf{M}$ in this case, so $\mathbf{U}_1 = \mathbf{U}$ and similarly $\mathbf{U}_2 = \mathbf{U}/n$, and putting these into (3.3.26) gives the result (3.3.17) for the equal gain case.

3.3.5 Simulation results (unequal antenna gains)

For a simulation for the case of unequal element gains any set of unequal gains might be chosen, but here we have modelled a possible practical array. This consists of a set of directional elements placed uniformly round a circle and directed outwards. In this case we take the case of ideal reflector-backed dipoles, which, because of the reflector, have zero gain in the back half circle. The gain (relative to the dipole gain) is $2\sin((\pi/2)\cos\theta)$ in direction θ between –90º and +90º from the normal to the reflector, which is placed a quarter wavelength behind the dipole. (See [7], page 178, with the response shown in Figure 7.7). Assuming the dipoles are vertical, this gives directional patterns in azimuth but identical parallel patterns in elevation. (In this study where we assume strong

signals, with performance limited only by errors, the elevation patterns, which will be sinusoidal, are not actually significant.) For this simulation we used 14 elements, as shown in Figure 3.12, where the element gains in direction 10º¸ as an example, are indicated. Because half the elements are inactive, with gain zero, to any particular target, at most only seven elements are useful and one or two of these, for which the target is at the edge of the pattern, may have low gain, as shown in this figure, reducing the effective number of elements further. The basic array is of one wavelength radius, so the distance between elements is only 0.45 wavelengths.

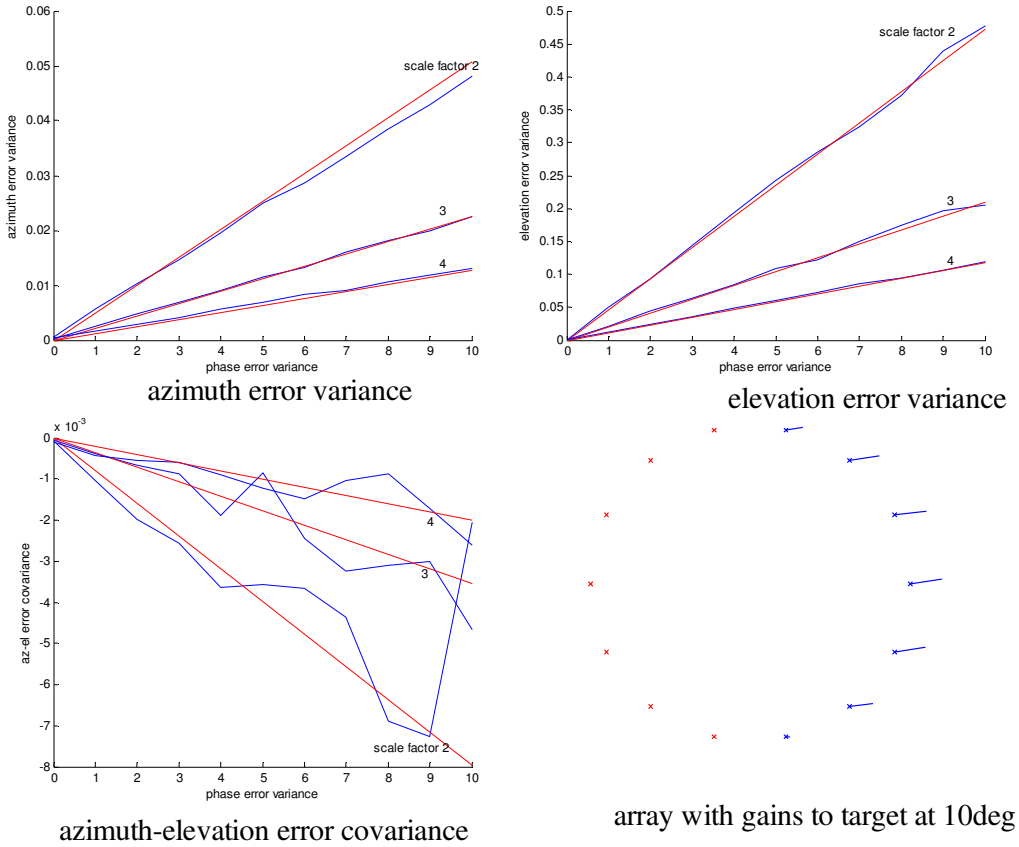


Figure 3.12 Error covariance results for array of directional elements (target at (30,40))

The results show a good agreement between simulation and theory, particularly for the azimuth and elevation error variances, though the azimuth-elevation covariance results are rather variable, but they are at quite a low level, much lower than the other variances, and could be made steadier by taking more trials (4000 were used here). The variance of the elevation errors is about ten times that of the azimuth errors, which is likely to be related to the considerably wider array aperture in azimuth than in elevation.

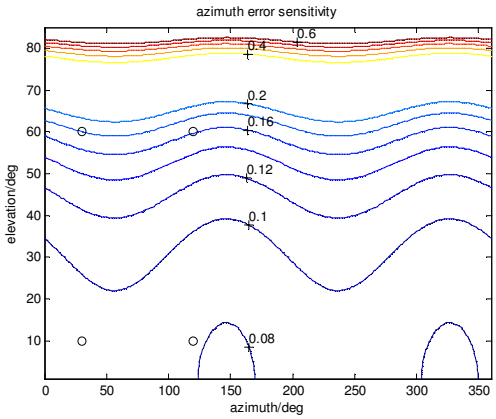
3.3.6 Error sensitivity

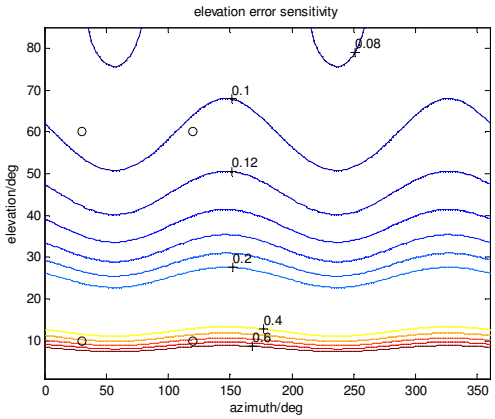
The theoretical expressions for the variances in the estimates of the angular positions of a target show that these are linearly related to the variances of the channel errors, for the model taken. (This model includes the supposition that we take the high signal to noise ratio case, so that only the effect of channel errors is significant.) If we take the square root of these variances then the standard deviations (s.d) of the DF errors are linearly related to the s.d. of the channel errors, specified in phase angle. If we set the input error values to unity then the expressions for the output errors can be taken to be the *error sensitivity* of the array; the expressions give the s.d. errors in the output estimates (azimuth and elevation), in degrees, for unit error (one degree s.d. in phase) at the input. If an estimate of the actual input error s.d. is known (e.g. 3º) then the DF s.d. values are given by multiplying the sensitivity by the input s.d. (by 3 in this example).

In Section 3.2.4 (Figs. 3.2, 3.3 and 3.5), Section 3.3.3 (Figs. 3.8, 3.9 and 3.11) and Section 3.3.5 (Figure 3.12) we see that the simulations confirm accurately the theoretical expressions, giving confidence in the theoretical expressions. However, these results are for specific target positions and it is desirable to obtain the performance for all target directions. Now, given an array, we can plot the potential performance of the array, using the theoretical sensitivity expressions, over any angular region of interest, instead of, perhaps, simulating the performance over a relatively limited set of points. Contour plots are given in Figure 3.13 for the azimuth and elevation error sensitivities for the horizontal planar array R1, shown in Figure 3.8. The plots show that the azimuth sensitivity rises rapidly for targets at high elevations where the differences in the ranges to the various elements with change of azimuth are very small making the effective aperture small. On the other hand at low elevations the array has maximum aperture in the azimuth plane, and the azimuth error sensitivity is lowest. (Note the change of scale of contour levels for these different regions.) While, in the previous linear theoretical plots it is possible to plot corresponding simulation results, this would not be so feasible for contour plots, so simulations were carried out at only four points, shown as small circles on the plot. The simulation results at these points (the s.d. determined over 2500 trials in each case) are given as the blue values in the small tables below the plots and these are seen to match very well the values read from the contour plots at these four points. The elevation plot shows the opposite form of sensitivity – the sensitivity is very high at low elevations where the effective elevation aperture is small, and low at high elevations where the variation of target-element ranges with elevation is equivalent to a

relatively large aperture. Again the simulation results show good agreement with the values read from the contour plot.

Theoretical plots:





Simulation results

azimuth sensitivity

		target azimuth 30	target azimuth 120
target elevation	60	0.18	0.16
	10	0.09	0.08

elevation sensitivity

		target azimuth 30	target azimuth 120
target elevation	60	0.09	0.11
	10	0.48	0.50

Figure 3.13 Error sensitivity for array R1

We note from (3.3.17) that the DF error variances (given in Ψ) are proportional to $\mathbf{U}^{-1}$ and from (3.3.10) that $\mathbf{U}$ is proportional to $\mathbf{M}$, which, in turn, is given by $\mathbf{M} = \sum_k \mathbf{r}_k \mathbf{r}_k^{\mathrm{T}} / n$, which can also be written (see (3.2.31)) as $\mathbf{M} = \mathbf{R}\mathbf{R}^{\mathrm{T}} / n$ (where the element position vectors are taken relative to the mean array position). It follows that if we double the array size (replacing $\mathbf{R}$ by $2\mathbf{R}$) $\mathbf{M}$ becomes 4 times as large and so also does $\mathbf{U}$, and then the error variances, in Ψ, are 1/4 the size. The standard deviations of the errors will then be half the values, with array $2\mathbf{R}$, of those they would be with array $\mathbf{R}$. In general we see that the error standard deviations are inversely proportional to the array size. (Running the sensitivity plot program with array 2×R1, i.e. with all coordinates twice those of R1, and contour intervals at 0.05 instead of 0.1, produces plots identical, except for the halved contour values, with those of Figure 3.13, and running the simulation program produces values close to half those given in Figure 3.13. The same result is seen for array 2×R2.)

From Figure 3.13 we see that for a planar array (of similar, parallel pattern elements) of aperture 2½ wavelengths (in both the X and Y coordinates, see R1, given below, and Figure 3.5) the error sensitivity is, over most of the angular space, less than 0.2, so it

would appear that if the aperture were one wavelength it should be in the region 0 to 0.5 (degrees in DF/degrees in phase error). This will not actually be the case, in general, because the elements will become too close in the sense that mutual coupling will be significant. However if the array aperture is made k wavelengths (k large enough to avoid significant mutual coupling, with elements at least ½ wavelength apart) then we expect sensitivities in the region 0 to 0.5/k (deg/deg). This seems to be a reasonable rule of thumb for the array accuracy, over most of the angular region, (not close to the array plane, for elevation accuracy, or close to the array zenith, for azimuth accuracy) for the case of strong signals (so that the performance is not limited by noise) with array manifold phase errors.

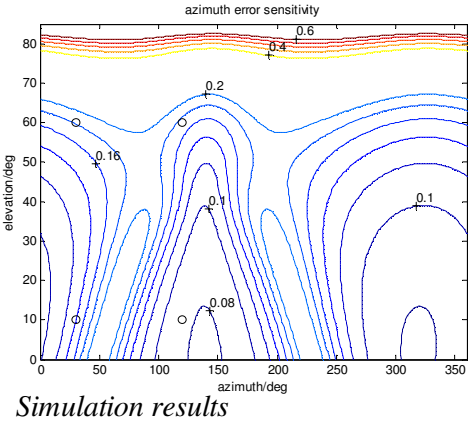
One more point of interest in Figure 3.13, the sensitivity plot for the planar array R1, is that there is 180º symmetry in azimuth, in the pattern – i.e. the pattern from 180º to 360º is identical to that from 0º to 180º, and we only need to plot one half of the azimuth range (e.g. 0º to 180º or –90º to +90º). This is because if we move 180º round the array, at a given elevation, the array looks exactly the same, except for a reflection normal to the azimuth look direction. This means that the array position variance, seen from this direction is exactly the same and so therefore is the sensitivity. This symmetry will not hold for non-planar arrays, in general (except for some particular, symmetric cases) as, with varying element position heights, the array will not look identical (other than a reflection) on moving 180º round it.

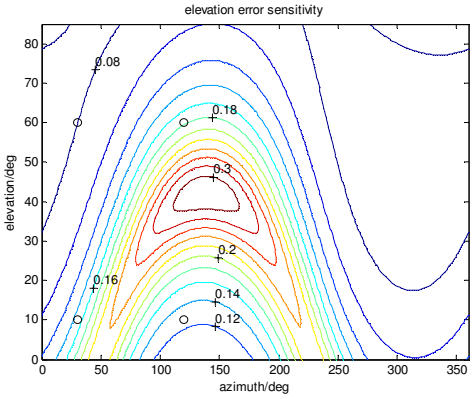
For a non-planar example we form array R2 by taking the elements of array R1 and varying them over a range of ±1 wavelengths in height above or below the plane of R1. The two arrays are given by

$$\mathrm{R1} = \begin{bmatrix} 0 & 0 & 0 & 0.5 & 1.5 & -1 \\ 0 & 0.5 & 1.5 & -1 & 0.5 & -0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathrm{R2} = \begin{bmatrix} 0 & 0 & 0 & 0.5 & 1.5 & -1.0 \\ 0 & 0.5 & 1.5 & -1.0 & 0.5 & -0.5 \\ 0 & 0.5 & 1.0 & -0.5 & -1.0 & 0 \end{bmatrix}$$

where the columns are the XYZ coordinates of the six elements and the units are wavelengths. The azimuth sensitivity plot for this array (Figure 3.14) again shows high sensitivity at high elevations, with a modified pattern at lower elevations. The elevation sensitivity plot now does not have the high sensitivity at low target elevations as the aperture in elevation is now comparable with the azimuth aperture. Again the simulation results, shown below the plots, agree well with the values read from the theoretical contour plots at the four target positions. We also note that the sensitivity patterns do

Theoretical plots:





Simulation results

azimuth sensitivity

		target azimuth	
		30	120
target elevation	60	0.18	0.17
	10	0.15	0.08

elevation sensitivity

		target azimuth	
		30	120
target elevation	60	0.08	0.18
	10	0.16	0.13

Figure 3.14 Error sensitivity for array R2

not have the 180º symmetry in azimuth shown in the planar array case.

3.3.7 Rule of thumb for regular circular arrays

The error sensitivity may also depend on the size of the array measured in the number of elements, but for a general, irregular array the relation may be quite weak. If we reduce the array R1 by removing the first two elements, leaving only the outer elements (3 to 6), which define the aperture (see Figure 3.8), we find there is only a little increase in the sensitivity, as shown in Table 3.1 below. This shows the azimuth and elevation sensitivities (using the simulation program) at the four points marked on the plots in Figures 3.13 and 3.14, for the basic array R1 and for the reduced, four element array. The ratios of the sensitivities of the reduced array to those of the full array are also shown. We see that the sensitivity ratios are little greater than unity (and even, in one case, just below, which may be an artefact due to the limited number of trials used in simulation) and vary with the target direction. The fact that the ratios are close to unity shows that the inner two elements of R1 contribute little to the array accuracy.

	elements 1 – 6				elements 3 – 6				ratios			
Az.	0.185	0.155	0.090	0.083	0.186	0.160	0.095	0.084	1.005	1.032	1.056	1.012
El.	0.092	0.103	0.476	0.538	0.093	0.107	0.475	0.553	1.011	1.034	0.998	1.028

Table 3.1 Error sensitivities at four target positions of array R1 and a reduced form of R1

We have not included any dependence on the number of elements, n, in the rule of thumb given above for general arrays, as we have seen above that the performance depends essentially on the elements defining the aperture (which we take to be a small number, perhaps three to five) with inner elements not greatly affecting the performance. An exception to the case where there are inner elements, not contributing very much to the array accuracy, is the case of circular arrays, where all the elements are equally significant, all being at one 'edge' of the aperture. In this case we can determine the dependence of the sensitivity on n. We see from (3.3.17) the error variances are proportional to n^{-1} (and hence the s.d. errors are proportional to $1/\sqrt{n}$) before taking into account the effect of **U**. As before, we note that **U** is proportional to **M**, or $\mathbf{RR}^{\mathrm{T}}/n$, which seems to depend on n, but in fact as n increases so does the product

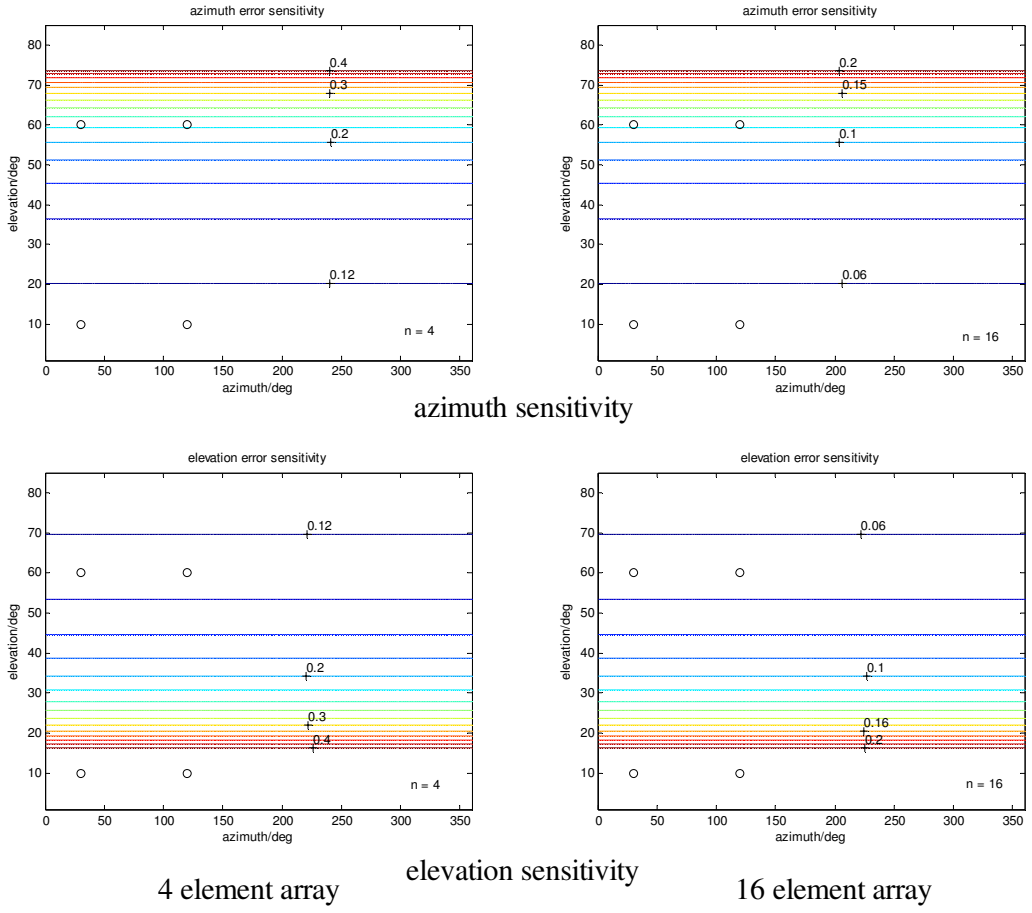


Figure 3.15 Error sensitivities for 4 and 16 element regular circular arrays

$\mathbf{RR}^{\mathrm{T}}$, proportionally (as **R** is of size $3\times n$), and so **M** and hence **U** do not depend on n, for the circular array where all the elements are equally significant. To confirm the $1/\sqrt{n}$ dependence of the sensitivity Figure 3.15 shows azimuth and elevation sensitivity plots for regular circular arrays of the same radius (one wavelength) for arrays of 4 and 16

elements. We see that the plots are identical except that the contour levels for the 16 element array are half those of the four element array, confirming the $1/\sqrt{n}$ relationship. We also note the uniformity of the patterns in the azimuth dimension – this is because the azimuth aperture is constant at any given elevation for these regular circular arrays of omnidirectional elements.

We note that both the azimuth and elevation sensitivities at the elevation of 45º are about 0.16 for the 4 element array and 0.08 for the 16 element array, or, more generally, using the $1/\sqrt{n}$ relationship, $0.32/\sqrt{n}$ for an n-element array. This is for an array of one wavelength radius, so, as the sensitivity is inversely proportional to radius we can say that the sensitivities at 45º elevation are given by $0.32/r\sqrt{n}$ for an n-element array of radius r. Now, in fact, we cannot pack as many elements as we like into a given array because mutual coupling will significantly change the element responses at separations under a half wavelength. Thus we could consider the case of regular circular arrays where the elements are at separation one half wavelength. In this case the radius of an n element array is given by $2r\sin(\pi/n) = ½$ or $r = 1/4\sin(\pi/n) \approx n/4\pi$ for large enough n. Putting this value of r into the expression above we have, for the sensitivities (in both azimuth and elevation) of a regular circular array of n elements at $\lambda/2$ separation, the values $1.28\pi/n\sqrt{n}$ or about $4/n\sqrt{n}$ and this could be a convenient rule of thumb for these arrays.

We note that the sensitivities vary considerably with elevation, being less than that at 45° for azimuth sensitivity at lower target elevations and for elevation sensitivity at higher elevations. The sensitivity is doubled at about 70° elevation for the azimuth sensitivity and at about 20° elevation for elevation sensitivity so within a factor of 2 the rule of thumb will give a useful estimate of the performance possible over most of the angle space.

Appendix 3.3A: Auxiliary results for 2D angle case

The elements of the manifold vector **a** are given for the linear array case in (3.2.20) but we now consider **a** as a function of both azimuth α and elevation ε. Thus we have (in the case of equal element gains)

$$a_k(\alpha,\varepsilon) = \exp\left(2\pi i \mathbf{r}_k^{\mathrm{T}} \mathbf{e}(\alpha,\varepsilon)\right)/\sqrt{n}\,. \tag{3.3A.1}$$

with the unit signal direction vector **e** dependent on both angles. Thus we have for the elements of $\mathbf{a}_\alpha$, replacing **d** in §2,

$$a_{\alpha k}(\alpha,\varepsilon) = \frac{\partial a_k(\alpha,\varepsilon)}{\partial \alpha} = 2\pi i \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\alpha(\alpha,\varepsilon) a_k(\alpha,\varepsilon) \tag{3.3A.2a}$$

and the elements of $\mathbf{a}_\varepsilon$ are given by

$$a_{\varepsilon k}(\alpha,\varepsilon) = \frac{\partial a_k(\alpha,\varepsilon)}{\partial \varepsilon} = 2\pi i \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\varepsilon(\alpha,\varepsilon) a_k(\alpha,\varepsilon) \tag{3.3A.2b}$$

Here we use $\mathbf{e}_\alpha$ and $\mathbf{e}_\varepsilon$ for the partial derivatives of $\mathbf{e}$ with respect to α and ε. Similarly the elements of $\mathbf{a}_{\alpha\alpha}$, $\mathbf{a}_{\alpha\varepsilon}$ and $\mathbf{a}_{\varepsilon\varepsilon}$ are given (omitting the indication of dependence on angle, for clarity) by

$$a_{\alpha\alpha k} = \frac{\partial^2 a_k}{\partial \alpha^2} = -\left(2\pi \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\alpha\right)^2 a_k + 2\pi i \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_{\alpha\alpha} a_k \tag{3.3A.3a}$$

$$a_{\alpha\varepsilon k} = \frac{\partial^2 a_k}{\partial \alpha \partial \varepsilon} = -(2\pi)^2 \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\alpha \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\varepsilon a_k + 2\pi i \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_{\alpha\varepsilon} a_k \tag{3.3A.3b}$$

$$a_{\alpha\alpha k} = \frac{\partial^2 a_k}{\partial \varepsilon^2} = -\left(2\pi \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_e\right)^2 a_k + 2\pi i \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_{\varepsilon\varepsilon} a_k . \tag{3.3A.3c}$$

($\mathbf{e}_{\alpha\alpha}$, $\mathbf{e}_{\alpha\varepsilon}$ and $\mathbf{e}_{\varepsilon\varepsilon}$ are the second partial derivatives of $\mathbf{e}$.) With these results and using $|a_k|^2 = 1/n$ and $\mathbf{r}_k{}^{\mathrm{T}} \dot{\mathbf{e}} = \dot{\mathbf{e}}^{\mathrm{T}} \mathbf{r}_k$ (for $\dot{\mathbf{e}}$ either $\mathbf{e}_\alpha$ or $\mathbf{e}_\varepsilon$) we have, first,

$$\mathbf{a}^{\mathrm{H}} \mathbf{a}_\alpha = \sum\nolimits_k a_k {}^* a_{\alpha k} = \sum\nolimits_k 2\pi i \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\alpha / n = 2\pi i \bar{\mathbf{r}}^{\mathrm{T}} \mathbf{e}_\alpha = 0 \tag{3.3A.4}$$

where $\bar{\mathbf{r}} = \sum_k \mathbf{r}_k / n = 0$, as the array reference is the mean position of the elements, and similarly $\mathbf{a}^{\mathrm{H}} \mathbf{a}_\varepsilon = 0$. Also

$$\mathrm{Re}(\mathbf{a}^{\mathrm{H}} \mathbf{a}_{\alpha\alpha}) = \mathrm{Re}\left(\sum\nolimits_k a_k {}^* a_{\alpha\alpha k}\right) = -\sum\nolimits_k \left(2\pi \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\alpha\right)^2 / n = -\sum\nolimits_k |a_{\alpha k}|^2 = -\|\mathbf{a}_\alpha\|^2 . \tag{3.3A.5}$$

As before, using (3.2.31),

$$\sum\nolimits_k \left(\mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\alpha\right)^2 / n = \sum\nolimits_k \mathbf{e}_\alpha{}^{\mathrm{T}} \mathbf{r}_k \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\alpha / n = \mathbf{e}_\alpha{}^{\mathrm{T}} \left(\sum\nolimits_k \mathbf{r}_k \mathbf{r}_k{}^{\mathrm{T}} / n\right) \mathbf{e}_\alpha = \mathbf{e}_\alpha{}^{\mathrm{T}} \mathbf{M} \mathbf{e}_\alpha \tag{3.3A.6}$$

so we have, from (3.3A.5) and (3.3A.6)

$$D = \mathrm{Re}(\mathbf{a}^{\mathrm{H}} \mathbf{a}_{\alpha\alpha}) = -\|\mathbf{a}_\alpha\|^2 = -4\pi^2 \mathbf{e}_\alpha{}^{\mathrm{T}} \mathbf{M} \mathbf{e}_\alpha \tag{3.3A.7}$$

where $\mathbf{M}$ is the array second moment matrix (or array position covariance matrix, also given in (3.2.31)). Similarly we obtain

$$E = \mathrm{Re}(\mathbf{a}^{\mathrm{H}} \mathbf{a}_{\alpha\varepsilon}) = -4\pi^2 \mathbf{e}_\alpha{}^{\mathrm{T}} \mathbf{M} \mathbf{e}_\varepsilon \text{ and } F = \mathrm{Re}(\mathbf{a}^{\mathrm{H}} \mathbf{a}_{\varepsilon\varepsilon}) = -4\pi^2 \mathbf{e}_\varepsilon{}^{\mathrm{T}} \mathbf{M} \mathbf{e}_\varepsilon .$$

We also see, from (3.3A.2), that

$$\mathrm{Re}(a_{\alpha k} * \Delta a_k) = \mathrm{Re}\left((2\pi i \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\alpha a_k) * \Delta a_k\right) = 2\pi \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\alpha \,\mathrm{Re}(-i a_k * \Delta a_k)$$

$$= 2\pi \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\alpha \,\mathrm{Im}(a_k * \Delta a_k)$$

so that

$$B = \mathrm{Re}(\mathbf{a}_\alpha{}^{\mathrm{H}} \Delta\mathbf{a}) = \sum_k \mathrm{Re}(a_{\alpha k} * \Delta a_k) = \sum_k 2\pi \mathbf{r}_k{}^{\mathrm{T}} \mathbf{e}_\alpha \,\mathrm{Im}(a_k * \Delta a_k)$$

$$= 2\pi \sum_k \left(\mathbf{R}^{\mathrm{T}} \mathbf{e}_\alpha\right)_k y_k = 2\pi \mathbf{e}_\alpha{}^{\mathrm{T}} \mathbf{R}\mathbf{y} \qquad (3.3A.8a)$$

where y_k = Im(a_k*Δa_k) as before, and we have used $\sum_k v_k y_k = \mathbf{v}^{\mathrm{T}}\mathbf{y}$ in general. Similarly

$$C = \mathrm{Re}(\mathbf{a}_\varepsilon{}^{\mathrm{H}} \Delta\mathbf{a}) = 2\pi \mathbf{e}_\varepsilon{}^{\mathrm{T}} \mathbf{R}\mathbf{y} \,. \qquad (3.3A.8b)$$

Appendix 3.3B: Weighted mean position

Let the element positions measured from some general reference point be $\mathbf{r}_1'$, $\mathbf{r}_2'$, . . ., $\mathbf{r}_n'$, and a weighted mean position be given by

$$\bar{\mathbf{r}} = \sum_k w_k \mathbf{r}_k' \qquad (3.3B.1)$$

where w_1, . . . w_n are a set of weights such that $\sum_k w_k = 1$. Now let the element positions measured relative to this mean position be given by $\mathbf{r}_1$, $\mathbf{r}_2$, . . ., $\mathbf{r}_n$ where $\mathbf{r}_k = \mathbf{r}_k' - \bar{\mathbf{r}}$, then the weighted mean of the positions is now given by

$$\sum_k w_k \mathbf{r}_k = \sum_k w_k \mathbf{r}_k' - \bar{\mathbf{r}} \sum_k w_k = \bar{\mathbf{r}} - \bar{\mathbf{r}} = \mathbf{0} \,. \qquad (3.3B.2)$$

In the case of equal parallel patterns the weights are all equal and of strength 1/n, but in the general case the weights are given by $|g_k|^2$ (k = 1 to n).

3.4. MULTIPLE TARGETS

3.4.1 MUSIC function in multiple target case

For the multiple target case we use the fact that the MUSIC 'spectral' function is a power-like measure (i.e. real and positive) of the degree to which the PSVs (point source vectors) over the parameter range of interest lie in the signal (sub)space – the vector space spanned by the PSVs of the signals received. This measure may be the squared lengths of the components of the PSVs *orthogonal to* this space, as in the original form of MUSIC, in which case we look for minima, or equivalently, and the measure taken here, the squared lengths of the components lying *in* this space, and look for maxima. In either case the PSVs should be standardized in length, most conveniently taken to be normalized. (This definition is consistent with that taken above (see (3.1.1)) for the single target case.)

One way of finding a basis for this space is by performing an eigenanalysis on the data covariance matrix, formed from the received waveforms, and taking the eigenvectors corresponding to non-trivial (principal) eigenvalues, those not corresponding to the receiver noise level. This begs questions such as how to decide which eigenvalues are essentially at the noise level, how much integration is required or how uniform or 'white' the noise is between channels or in the external environment. In this study, where we are considering the performance as limited by errors affecting the knowledge of the PSVs (channel errors or element position errors), we take the case of strong signals, so that noise effects are negligible, and are indeed neglected. In this case, assuming accurate arithmetic, the principal eigenvectors obtained will form an accurate basis for the received signal space – i.e. the m-dimensional vector space (with m significant signals present) defined by the m signal PSVs (albeit modified by the receiving system, with errors). Putting this alternatively, both the set of m significant eigenvectors and the set of m signal PSVs span the same space and each eigenvector is a linear combination of the signal PSVs and *vice versa*. (However, the eigenvector basis, unlike the signal PSV basis, is an orthogonal set. This is a desirable attribute but not a necessary one for forming the MUSIC spectrum.) It was noted in 1987 [6] that, if an estimate of the noise covariance matrix is available, (which is often assumed, generally implicitly, when using eigenanalysis for the processing) then a signal space basis can be found more easily than by eigenanalysis – for example by Gram-Schmidt processing (producing an orthogonal basis) or even more simply by Gaussian elimination (giving a non-orthogonal basis). Having obtained a signal space basis then the principle of

MUSIC is, given a stored set of PSVs covering the parameter region of interest with suitably fine sampling (the stored 'manifold'), to find the parameter '(power) spectrum' over this region in the form of the squared distance of the PSVs in (or orthogonal to) the signal space. The PSVs which lie precisely *in* this space are taken to be those with the parameter values corresponding to the actual targets. Because of inaccuracies (finite precision arithmetic, for example, or, the subject of this study, the mismatch between the actual signal PSVs seen by the system and the stored values) we look for the m stored PSVs which lie most nearly in the signal space. To do this we project the manifold PSVs into this space and look for the m maximum values. If $\mathbf{B}$ is an $n\times m$ matrix of m signal space basis vectors (for an n-element sensor system) then the projection of an n-vector $\mathbf{v}$ is given by $\mathbf{P_B v}$ where the projection matrix is $\mathbf{P_B} = \mathbf{B}(\mathbf{B}^H\mathbf{B})^{-1}\mathbf{B}^H$. Any other basis for the same space, $\mathbf{B'}$, can be written as $\mathbf{B'} = \mathbf{BK}$ where $\mathbf{K}$ is a non-singular $m\times m$ matrix, and this represents the fact that the vectors of one basis are all linear combinations of the vectors of the other basis. Inserting $\mathbf{BK}$ into the expression for $\mathbf{P_{B'}}$ we find that the factors of $\mathbf{K}$ and $\mathbf{K}^{-1}$ all cancel out and that $\mathbf{P_{B'}} = \mathbf{P_B}$. This is not surprising, as it simply states that the projection into a vector space is independent of the basis used for that space. From this we conclude that instead of using the basis of eigenvectors for the projection we can equally well use the basis of the signal PSVs. Of course in practice these are not known and the eigenvector basis (or other basis, such as Gram-Schmidt) derived from the observed data covariance matrix, must be used, but for the theoretical analysis the actual PSV basis is equivalent to these other, experimentally determined, bases and this is what is used below.

Thus the MUSIC function for a parameter set $\boldsymbol{\theta}$ is given by

$$f(\boldsymbol{\theta}) = \left\|\mathbf{P_A a}(\boldsymbol{\theta})\right\|^2 = \left(\mathbf{P_A a}(\boldsymbol{\theta})\right)^H \mathbf{P_A a}(\boldsymbol{\theta}) = \mathbf{a}(\boldsymbol{\theta})^H \mathbf{P_A}^H \mathbf{P_A a}(\boldsymbol{\theta}) = \mathbf{a}(\boldsymbol{\theta})^H \mathbf{P_A a}(\boldsymbol{\theta}) \qquad (3.4.1)$$

where we have used the Hermitian and idempotent properties of projection matrices ($\mathbf{P_A}^H = \mathbf{P_A}$, and $\mathbf{P_A}^2 = \mathbf{P_A}$) and $\mathbf{A}$ is the $n\times m$ matrix of the m signal PSVs. This form is consistent with that used for the single target case (as in (3.1.1), (3.2.6) and (3.3.2)). In (3.1.1) for example we could put this in the form

$$f(\theta) = \left|\mathbf{a}(\theta)^H \mathbf{a}(\phi)\right|^2 = \mathbf{a}(\theta)^H \mathbf{a}(\phi)\left(\mathbf{a}(\theta)^H \mathbf{a}(\phi)\right)^H = \mathbf{a}(\theta)^H \mathbf{a}(\phi)\mathbf{a}(\phi)^H \mathbf{a}(\theta)$$

$$= \mathbf{a}(\theta)^H \mathbf{P}_{\mathbf{a}(\phi)} \mathbf{a}(\theta) \qquad (3.4.2)$$

where we have put

$$\mathbf{P}_{\mathbf{a}(\phi)} = \mathbf{a}(\phi)\mathbf{a}(\phi)^{\mathrm{H}} = \frac{\mathbf{a}(\phi)\mathbf{a}(\phi)^{\mathrm{H}}}{\left\|\mathbf{a}(\phi)\right\|^2} \tag{3.4.3}$$

as in this case **a** is taken to be normalized, i.e. ‖**a**(ϕ)‖ = 1. The last form in (3.4.3) is the general form for projecting into a space of a single dimension, whose basis (not necessarily normalized) is given by **a**(ϕ). (It is easily checked that $\mathbf{P}_{\mathbf{a}(\phi)}\mathbf{a}(\phi) = \mathbf{a}(\phi)$, i.e. $\mathbf{P}_{\mathbf{a}(\phi)}$ projects $\mathbf{a}(\phi)$ onto itself.) When the basis vector was not normalized, as in (3.2.1) where the vector is **a** + Δ**a**, we effectively divided by its modulus, as in (3.4.3); in fact we multiplied by *h* where $h^2 = 1/\left\|\mathbf{a} + \Delta\mathbf{a}\right\|^2$, see (3.2.2). The form given in (3.4.1) will be used in the multiple target case, i.e. $\mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}(\boldsymbol{\theta})$, where $\mathbf{A} = \left[\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \quad \cdots \quad \mathbf{a}_m\right]$ for the case of *m* targets, and **θ** is in general a 2-component angle parameter (azimuth and elevation).

3.4.2 Measurement error due to PSV error; multiple targets, single parameter case

Following the approach of §3.2.1, extended to the case of multiple targets, we have, for a target with the true parameter value α, the true set of target PSVs **A**, the errors in **A** given by Δ**A**, and the resultant error in α given by Δα,

$$f(\alpha + \Delta\alpha) = \mathbf{a}(\alpha + \Delta\alpha)^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}(\alpha + \Delta\alpha) \tag{3.4.4}$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \Delta\mathbf{A}$. We assume Δα and Δ**A** are small and that the actual measurement error is given by the value of Δα that maximizes *f* round the parameter value α. We expand the function in terms of Δα to second order, in order to find the maximum, and take terms to only second order of smallness in terms of products involving Δα and Δ**A** (i.e. excluding terms in $\Delta\alpha^2$ and Δ**A**, and in Δα and $\Delta\mathbf{A}^2$ and higher order products, of course).

We now expand **a**(α+Δα) up to second order in Δα, as before (as in (3.2.3));

$$\mathbf{a}(\alpha + \Delta\alpha) = \mathbf{a}(\alpha) + \mathbf{d}(\alpha)\Delta\alpha + \tfrac{1}{2}\mathbf{s}(\alpha)\Delta\alpha^2 \tag{3.4.5}$$

where **d** and **s** are the first and second derivatives of **a** with respect to α. (NB: We do not include a normalizing factor *k*, as given in (3.2.5) as, because of the definition of **a**, it was seen in §3.2.1 that the expanded form is also normalized. Although the finite expansion in (3.4.5) is not exact, the norm of **a**(α+Δα) is unity to the second order accuracy taken here.) Then (3.4.4) becomes

$$f(\alpha+\Delta\alpha)=\left(\mathbf{a}(\alpha)+\mathbf{d}(\alpha)\Delta\alpha+\tfrac{1}{2}\mathbf{s}(\alpha)\Delta\alpha^2\right)^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\left(\mathbf{a}(\alpha)+\mathbf{d}(\alpha)\Delta\alpha+\tfrac{1}{2}\mathbf{s}(\alpha)\Delta\alpha^2\right) \tag{3.4.6}$$

$$=A+2B\Delta\alpha+C\Delta\alpha^2+\cdots \tag{3.4.7}$$

taking terms up to second order only. The error is given by the value of $\Delta\alpha$ which maximizes (3.4.7). We have $df/d\Delta\alpha=0$ when $2B+2C\Delta\alpha=0$, i.e.

$$\Delta\alpha=-B/C\,. \tag{3.4.8}$$

Taking the terms in $\Delta\alpha$ and $\Delta\alpha^2$ in the expansion of (3.4.6) we have

$$2B=\mathbf{a}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{d}+\mathbf{d}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}=2\,\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{d}) \tag{3.4.9}$$

and

$$C=\tfrac{1}{2}\mathbf{a}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{s}+\tfrac{1}{2}\mathbf{s}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}+\mathbf{d}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{d}=\mathrm{Re}(\mathbf{a}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{s})+\mathbf{d}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{d}\,. \tag{3.4.10}$$

(We have used $\mathbf{y}^{\mathrm{H}}\mathbf{P}\mathbf{x}=(\mathbf{x}^{\mathrm{H}}\mathbf{P}\mathbf{y})^{\mathrm{H}}=(\mathbf{x}^{\mathrm{H}}\mathbf{P}\mathbf{y})^*$, as $\mathbf{x}^{\mathrm{H}}\mathbf{P}\mathbf{y}$ is a scalar so its value is the same as its transpose, and we have also used $\mathbf{P}=\mathbf{P}^{\mathrm{H}}$.) Now if we drop all terms of third order (and higher) in smallness (where $\Delta\alpha$ and $\Delta\mathbf{A}$ are both of first order) we see that for B we only need terms up to first order in $\Delta\mathbf{A}$ in the expansion of $\mathbf{P}_{\tilde{\mathbf{A}}}$ and for C which multiplies $\Delta\alpha^2$ we take terms of zeroth order only. From Appendix 3.4A we have, to first order in $\Delta\mathbf{A}$,

$$\mathbf{P}_{\tilde{\mathbf{A}}}=\mathbf{P}_{\mathbf{A}}+\mathbf{Q}_{\mathbf{A}}\Delta\mathbf{A}\mathbf{A}^{+}+\mathbf{A}^{+\mathrm{H}}\Delta\mathbf{A}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}} \tag{3.4.11}$$

where $\mathbf{A}^{+}=(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}$ is a pseudoinverse of $\mathbf{A}$ (such that $\mathbf{A}^{+}\mathbf{A}=(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{A}=\mathbf{I}_m$) and $\mathbf{Q_A}=\mathbf{I}-\mathbf{P_A}$. For target r we have from (3.4B.4) $\Delta\mathbf{A}\mathbf{A}^{+}\mathbf{a}_r=\Delta\mathbf{a}_r$, where $\Delta\mathbf{a}_r$ is column r of $\Delta\mathbf{A}$. Also using $\mathbf{P_A}\mathbf{a}_r=\mathbf{a}_r$ and $\mathbf{Q_A}\mathbf{a}_r=\mathbf{0}$, we have, putting these results into (3.4.11),

$$\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}_r=\mathbf{a}_r+\mathbf{Q}_{\mathbf{A}}\Delta\mathbf{a}_r \tag{3.4.12}$$

Then, from (3.4.9) we have

$$B_r=\mathrm{Re}(\mathbf{a}_r{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{d}_r)=\mathrm{Re}(\mathbf{d}_r{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}_r)=\mathrm{Re}\left(\mathbf{d}_r{}^{\mathrm{H}}(\mathbf{a}_r+\mathbf{Q}_{\mathbf{A}}\Delta\mathbf{a}_r)\right)$$

$$=\mathrm{Re}(\mathbf{d}_r{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\Delta\mathbf{a}_r) \tag{3.4.13}$$

as $\mathrm{Re}(\mathbf{a}_r{}^{\mathrm{H}}\mathbf{d}_r)=\mathbf{0}$ from (3.2A.2).

From (3.4.11) the zeroth order approximation to $\mathbf{P}_{\tilde{\mathbf{A}}}$ is $\mathbf{P}_{\mathbf{A}}$, of course, so putting this into (3.4.10) and using $\mathbf{P}_{\mathbf{A}}\mathbf{a}_r = \mathbf{a}_r$ (and $\mathbf{a}_r{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}} = \mathbf{a}_r{}^{\mathrm{H}}$) we have

$$C_r = \mathrm{Re}(\mathbf{a}_r{}^{\mathrm{H}}\mathbf{s}_r) + \mathbf{d}_r{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{d}_r .$$

Using (3.2A.6) (or (3.3A.5)) we have, using the definition of $\mathbf{Q}_{\mathbf{A}}$,

$$C_r = -\|\mathbf{d}_r\|^2 + \mathbf{d}_r{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{d}_r = -\mathbf{d}_r{}^{\mathrm{H}}\mathbf{d}_r + \mathbf{d}_r{}^{\mathrm{H}}(\mathbf{I} - \mathbf{Q}_{\mathbf{A}})\mathbf{d}_r = -\mathbf{d}_r{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_r . \qquad (3.4.14)$$

The error on the estimate of the parameter α_r for the rth target due to the error $\Delta\mathbf{A}$ in $\mathbf{A}$ is, from (3.4.8),

$$\Delta\alpha_r = -B_r / C_r \qquad (3.4.15)$$

where B_r and C_r are given in (3.4.13) and (3.4.14), so we have, finally

$$\Delta\alpha_r = \frac{\mathrm{Re}\left(\mathbf{d}_r{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\Delta\mathbf{a}_r\right)}{\mathbf{d}_r{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{d}_r} . \qquad (3.4.16)$$

[This is consistent with the result (3.2.9) for a single target. In this case (m =1, $\mathbf{A} = \mathbf{a}$) we have $\mathbf{Q}_{\mathbf{A}} = \mathbf{Q}_{\mathbf{a}} = \mathbf{I} - \mathbf{a}\mathbf{a}^{\mathrm{H}} / \|\mathbf{a}\|^2$ so $\mathbf{d}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}} = \mathbf{d}^{\mathrm{H}} - \mathbf{d}^{\mathrm{H}}\mathbf{a}\mathbf{a}^{\mathrm{H}}$ (as $\mathbf{a}$ is normalized) and the numerator of (3.4.16) becomes $\mathrm{Re}\left(\mathbf{d}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\Delta\mathbf{a}\right) = \mathrm{Re}\left(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}\right) - \mathrm{Re}\left(\mathbf{d}^{\mathrm{H}}\mathbf{a}\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a}\right)$. However we have $\mathrm{Re}\left(\mathbf{d}^{\mathrm{H}}\mathbf{a}\right) = 0$, so $\mathbf{d}^{\mathrm{H}}\mathbf{a}$ is imaginary, and $\mathbf{d}^{\mathrm{H}}\mathbf{a} = -\mathbf{a}^{\mathrm{H}}\mathbf{d}$. Then the numerator becomes

$$B = \mathrm{Re}\left(\mathbf{d}^{\mathrm{H}}\Delta\mathbf{a}\right) + \mathrm{Re}\left(\mathbf{a}^{\mathrm{H}}\mathbf{d}\mathbf{a}^{\mathrm{H}}\Delta\mathbf{a}\right)$$

in agreement with (3.2.9). The denominator of (3.4.16) is seen immediately to be the same as that of (3.2.9) on substituting for $\mathbf{Q}_{\mathbf{a}}$.]

3.4.3 Measurement error due to PSV error; multiple targets, two parameter case

The analysis for m targets follows closely that for the single target given in §3.3.1 with the exception that the projection matrix $\mathbf{P}_{\tilde{\mathbf{A}}}$ is now included, where $\tilde{\mathbf{A}}$ is of size $n \times m$. In the two parameter case the expression for the MUSIC function in (3.4.4) is replaced by

$$f(\alpha + \Delta\alpha, \varepsilon + \Delta\varepsilon) = \mathbf{a}(\alpha + \Delta\alpha, \varepsilon + \Delta\varepsilon)^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}(\alpha + \Delta\alpha, \varepsilon + \Delta\varepsilon) . \qquad (3.4.17)$$

Using the expansion of $\mathbf{a}$ as before (3.3.1) (but setting k = 1) we can put f in the form of (3.3.5) but now we have, instead of (3.4.9) and (3.4.10),

$$B = \mathrm{Re}\left(\mathbf{a}_\alpha{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}\right),\ C = \mathrm{Re}\left(\mathbf{a}_\varepsilon{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}\right) \tag{3.4.18}$$

and

$$\begin{aligned} &D = \mathrm{Re}\left(\mathbf{a}_{\alpha\alpha}{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}\right) + \mathbf{a}_\alpha{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}_\alpha,\ E = \mathrm{Re}\left(\mathbf{a}_{\alpha\varepsilon}{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}\right) + \mathrm{Re}\left(\mathbf{a}_\alpha{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}_\varepsilon\right), \\ &F = \mathrm{Re}\left(\mathbf{a}_{\varepsilon\varepsilon}{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}\right) + \mathbf{a}_\varepsilon{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}_\varepsilon. \end{aligned} \tag{3.4.19}$$

Replacing $\mathbf{d}$ in (3.4.9) by $\mathbf{a}_\alpha$ or $\mathbf{a}_\varepsilon$ we can use the result obtained in §4.2 for B_r (see (3.4.13)) for B_r and C_r in (3.4.18), considering target r, to obtain

$$B_r = \mathrm{Re}\left(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\Delta\mathbf{a}_r\right) \text{ and } C_r = \mathrm{Re}\left(\mathbf{a}_{r\varepsilon}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\Delta\mathbf{a}_r\right). \tag{3.4.20}$$

(The notation $\mathbf{a}_{r\alpha}$, for example, means the partial derivative of the vector $\mathbf{a}_r$ with respect to α.)

As in deriving C_r from C in §3.4.2, we note that we replace $\mathbf{P}_{\tilde{\mathbf{A}}}$ by $\mathbf{P}_{\mathbf{A}}$ in D, E and F, so that D, for example, becomes, considering target r,

$$D_r = \mathrm{Re}\left(\mathbf{a}_{r\alpha\alpha}{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}_r\right) + \mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}_{r\alpha} = \mathrm{Re}\left(\mathbf{a}_{r\alpha\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}\right) + \mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}_{r\alpha} = \mathrm{Re}\left(\mathbf{a}_{r\alpha\alpha}{}^{\mathrm{H}}\mathbf{a}\right) + \mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}_{r\alpha}$$

$$= -\left\|\mathbf{a}_{r\alpha}\right\|^2 + \mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}_{r\alpha} = -\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{a}_{r\alpha} + \mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}_{r\alpha} = -\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\alpha} \tag{3.4.21}$$

using $\mathbf{P}_{\mathbf{A}}\mathbf{a}_r = \mathbf{a}_r$ (from (3.4B.1)), (3.3A.5) and $\mathbf{Q}_{\mathbf{A}} = \mathbf{I} - \mathbf{P}_{\mathbf{A}}$. This differs from the single signal case in that there we note that $\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}_{r\alpha} = \mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{a}_r\mathbf{a}_r{}^{\mathrm{H}}\mathbf{a}_{r\alpha} = 0$, from (3.3A.4). For F_r we have, similarly,

$$F_r = -\mathbf{a}_{r\varepsilon}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\varepsilon}$$

but for E_r we note that

$$E_r = \mathrm{Re}\left(\mathbf{a}_{r\alpha\varepsilon}{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}_r\right) + \mathrm{Re}\left(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\tilde{\mathbf{A}}}\mathbf{a}_{r\varepsilon}\right) = \mathrm{Re}\left(\mathbf{a}_{r\alpha\varepsilon}{}^{\mathrm{H}}\mathbf{a}_r\right) + \mathrm{Re}\left(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}_{r\varepsilon}\right). \tag{3.4.22}$$

We see, in a derivation similar to that of (3.3A.5), we have

$$\mathrm{Re}\left(\mathbf{a}^{\mathrm{H}}\mathbf{a}_{\alpha\varepsilon}\right) = \mathrm{Re}\left(\sum_k a_k{}^* a_{\alpha\varepsilon k}\right) = -\sum_k (2\pi)^2\, \mathbf{r}_k{}^{\mathrm{T}}\mathbf{e}_\alpha \mathbf{r}_k{}^{\mathrm{T}}\mathbf{e}_\varepsilon \Big/ n = -\sum_k a_{\alpha k}{}^* a_{\varepsilon k}$$

$$= -\mathbf{a}_\alpha{}^{\mathrm{H}}\mathbf{a}_\varepsilon \tag{3.4.23}$$

using equations (3.3A.3b) and then (3.3A.2). From (3.4.22) and (3.4.23) we have

$$E_r = \mathrm{Re}\left(\mathbf{a}_{r\alpha\varepsilon}{}^{\mathrm{H}}\mathbf{a}_r\right) + \mathrm{Re}\left(\mathbf{a}_{\alpha r}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}_{\varepsilon r}\right) = -\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{a}_{r\varepsilon} + \mathrm{Re}\left(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}_{r\varepsilon}\right).$$

$$= -\mathrm{Re}\left(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\varepsilon}\right) \tag{3.4.24}$$

Putting these results into (3.3.7), giving the parameter errors in the case of two parameters, but now for the multiple target case, we have, for the errors on the estimates of target *r*,

$$\begin{bmatrix}\Delta\alpha_r\\ \Delta\varepsilon_r\end{bmatrix} = -\begin{bmatrix}D_r & E_r\\ E_r & F_r\end{bmatrix}^{-1}\begin{bmatrix}B_r\\ C_r\end{bmatrix} = \mathbf{U}_r{}^{-1}\mathbf{u}_r \tag{3.4.25}$$

where

$$\mathbf{U}_r = \begin{bmatrix}\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\alpha} & \mathrm{Re}\left(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\varepsilon}\right)\\ \mathrm{Re}\left(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\varepsilon}\right) & \mathbf{a}_{r\varepsilon}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\varepsilon}\end{bmatrix} \text{ and } \mathbf{u}_r = \begin{bmatrix}\mathrm{Re}\left(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\Delta\mathbf{a}_r\right)\\ \mathrm{Re}\left(\mathbf{a}_{r\varepsilon}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\Delta\mathbf{a}_r\right)\end{bmatrix}. \tag{3.4.26}$$

These correspond to **U** and **u** given in (3.3.8) and (3.3.9) as applicable to the single signal case. (For example, $U(1,1) = \mathbf{a}_\alpha{}^{\mathrm{H}}\left(\mathbf{I}-\mathbf{a}\mathbf{a}^{\mathrm{H}}\right)\mathbf{a}_\alpha = \left\|\mathbf{a}_\alpha\right\|^2 = -\mathrm{Re}\left(\mathbf{a}_{\alpha\alpha}{}^{\mathrm{H}}\mathbf{a}\right)$ using (3.2A.6) and the fact that $\mathbf{a}^{\mathrm{H}}\mathbf{a}_\alpha = 0$.)

3.4.4 Measurement error due to PSV error; two targets, two parameter case

Equations (3.4.25) and (3.4.26) cannot be easily simplified into a form corresponding to (3.3.10) in the general multiple signal case, but the general case is not of great interest theoretically – little can be determined in the way of simple statements about the accuracy. For well separated signals (particularly if their number *m* is well below the number of array elements *n*) we expect the accuracy of measurement of each of the signals to approach that given by the single signal case, given in §3.3, and for close signals we expect the accuracy of measurement of one signal to be affected by the proximity of the others. In this section we take the case of only two close signals; this is both more tractable and also of more general practical value.

Taking the (1,1) element of $\mathbf{U}_r$ we have

$$U_r(1,1) = \mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\alpha} = \mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{a}_{r\alpha} - \mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}_{r\alpha} \tag{3.4.27}$$

From (3.3A.2) and (3.2.31)

$$\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{a}_{r\alpha} = 4\pi^2\mathbf{e}_{r\alpha}{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_{r\alpha} = 4\pi^2\mathbf{e}_{r\alpha}{}^{\mathrm{T}}\left(\mathbf{R}\mathbf{R}^{\mathrm{T}}/n\right)\mathbf{e}_{r\alpha}. \tag{3.4.28}$$

(This is the same as –D in §3.3.1 (see (3.3.10) and (3.3.8)), obtained equivalently via (3.3A.6) and (3.3A.7).) The second term on the right side of (3.4.27) is given, for target 1, using (3.4C.5), by

$$\mathbf{a}_{1\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}_{1\alpha} = \mathbf{a}_{1\alpha}{}^{\mathrm{H}}(\mathbf{a}_1\mathbf{a}_1{}^{\mathrm{H}} - \gamma\mathbf{a}_1\mathbf{a}_2{}^{\mathrm{H}} - \gamma^*\mathbf{a}_2\mathbf{a}_1{}^{\mathrm{H}} + \mathbf{a}_2\mathbf{a}_2{}^{\mathrm{H}})\mathbf{a}_{1\alpha}/\mu = \mathbf{a}_{1\alpha}{}^{\mathrm{H}}\mathbf{a}_2\mathbf{a}_2{}^{\mathrm{H}}\mathbf{a}_{1\alpha}/\mu \quad (3.4.29)$$

where we have used $\mathbf{a}_{1\alpha}{}^{\mathrm{H}}\mathbf{a}_1 = 0$ from (3.3A.4), $\gamma = \mathbf{a}_1{}^{\mathrm{H}}\mathbf{a}_2$ and $\mu = 1 - |\gamma|^2$.

Now (using (3.3.2a))

$$\mathbf{a}_{1\alpha}{}^{\mathrm{H}}\mathbf{a}_2 = \sum\nolimits_k -2\pi i\mathbf{r}_k{}^{\mathrm{T}}\mathbf{e}_{1\alpha}a_{1k}{}^*a_{2k} = -2\pi i\sum\nolimits_k \left(\mathbf{R}^{\mathrm{T}}\mathbf{e}_{1\alpha}\right)_k b_k = -2\pi i\mathbf{e}_{1\alpha}{}^{\mathrm{T}}\mathbf{R}\mathbf{b} \quad (3.4.30)$$

where $b_k = a_{1k}{}^*a_{2k}$ as in (3.4C.3). Using (3.4.30) in (3.4.29) gives

$$\mathbf{a}_{1\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\mathbf{a}_{1\alpha} = \mathbf{a}_{1\alpha}{}^{\mathrm{H}}\mathbf{a}_2\left(\mathbf{a}_{1\alpha}{}^{\mathrm{H}}\mathbf{a}_2\right)^*/\mu = 4\pi^2\mathbf{e}_{1\alpha}{}^{\mathrm{T}}\mathbf{R}\mathbf{b}\mathbf{b}^{\mathrm{H}}\mathbf{R}^{\mathrm{T}}\mathbf{e}_{1\alpha}/\mu\,. \quad (3.4.31)$$

Thus we have, from (3.4.27), (3.4.28) and (3.4.31),

$$U_1(1,1) = 4\pi^2\mathbf{e}_{1\alpha}{}^{\mathrm{T}}\mathbf{R}\left(\frac{1}{n}\mathbf{I} - \frac{\mathbf{b}\mathbf{b}^{\mathrm{H}}}{\mu}\right)\mathbf{R}^{\mathrm{T}}\mathbf{e}_{1\alpha}\,. \quad (3.4.32)$$

We note that the matrix $\mathbf{R}\left(1/n\,\mathbf{I} - \mathbf{b}\mathbf{b}^{\mathrm{H}}/\mu\right)\mathbf{R}^{\mathrm{T}}$ is Hermitian so it follows that $U_1(1,1)$ is real, and so we need only the real part of this matrix and we can put

$$U_1(1,1) = 4\pi^2\mathbf{e}_{1\alpha}{}^{\mathrm{T}}\tilde{\mathbf{M}}\mathbf{e}_{1\alpha} \quad (3.4.33)$$

where

$$\tilde{\mathbf{M}} = \mathrm{Re}\left\{\mathbf{R}\left(\frac{1}{n}\mathbf{I} - \frac{\mathbf{b}\mathbf{b}^{\mathrm{H}}}{\mu}\right)\mathbf{R}^{\mathrm{T}}\right\} = \mathbf{M} - \frac{\mathbf{R}\,\mathrm{Re}(\mathbf{b}\mathbf{b}^{\mathrm{H}})\mathbf{R}^{\mathrm{T}}}{\mu}. \quad (3.4.34)$$

We clearly have a similar result for target 2, replacing 1 by 2, so we can replace 1 in (3.4.33) by r (r = 1 or 2). We also see that we have a similar result, with ε replacing α, for element (2,2) of $\mathbf{U}_r$. For the off-diagonal elements of $\mathbf{U}_r$, following the argument above, we obtain

$$U_r(1,2) = \mathrm{Re}\left(4\pi^2\mathbf{e}_{r\alpha}{}^{\mathrm{T}}\mathbf{R}\left(1/n\,\mathbf{I} - \mathbf{b}\mathbf{b}^{\mathrm{H}}/\mu\right)\mathbf{R}^{\mathrm{T}}\mathbf{e}_{r\varepsilon}\right) = 4\pi^2\mathbf{e}_{r\alpha}{}^{\mathrm{T}}\tilde{\mathbf{M}}\mathbf{e}_{r\varepsilon} = U_r(2,1)\,. \quad (3.4.35)$$

With these results, we see that, for the two signal case, we have a modified form of (3.3.10):

$$\mathbf{U}_r = 4\pi^2\begin{bmatrix}\mathbf{e}_{r\alpha}{}^{\mathrm{T}}\tilde{\mathbf{M}}\mathbf{e}_{r\alpha} & \mathbf{e}_{r\alpha}{}^{\mathrm{T}}\tilde{\mathbf{M}}\mathbf{e}_{r\varepsilon} \\ \mathbf{e}_{r\varepsilon}{}^{\mathrm{T}}\tilde{\mathbf{M}}\mathbf{e}_{r\alpha} & \mathbf{e}_{r\varepsilon}{}^{\mathrm{T}}\tilde{\mathbf{M}}\mathbf{e}_{r\varepsilon}\end{bmatrix} = 4\pi^2\mathbf{E}_r{}^{\mathrm{T}}\tilde{\mathbf{M}}\mathbf{E}_r \quad (3.4.36)$$

with $\tilde{\mathbf{M}}$ given in (3.4.34) and $\mathbf{E}_r$ as in (3.3.10).

For the first component of $\mathbf{u}_r$ we have

$$u_r(1) = \mathrm{Re}\left(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\Delta\mathbf{a}_r\right) = \mathrm{Re}\left(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\Delta\mathbf{a}_r\right) - \mathrm{Re}\left(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\Delta\mathbf{a}_r\right). \quad (3.4.37)$$

Similarly to the derivation of (3.4.30) we have

$$\mathbf{a}_{1\alpha}{}^{\mathrm{H}}\Delta\mathbf{a}_1 = \sum\nolimits_k -2\pi i\mathbf{r}_k{}^{\mathrm{T}}\mathbf{e}_{1\alpha}a_{1k} * \Delta a_{1k} = -2\pi i\sum\nolimits_k \left(\mathbf{R}^{\mathrm{T}}\mathbf{e}_{1\alpha}\right)_k z_{1k} = -2\pi i\mathbf{e}_{1\alpha}{}^{\mathrm{T}}\mathbf{R}\mathbf{z}_1 \quad (3.4.38)$$

where $z_{1k} = a_{1k} * \Delta a_{1k}$.

From (3.4.37), (3.4.38) and (3.4C.9) we have

$$u_1(1) = \mathrm{Re}\left(-2\pi i\mathbf{e}_{1\alpha}{}^{\mathrm{T}}\mathbf{R}\mathbf{z}_1 + 2\pi i\mathbf{e}_{1\alpha}{}^{\mathrm{T}}\mathbf{R}\mathbf{b}\mathbf{b}^{\mathrm{H}}(n\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}})\mathbf{z}_1/\mu\right)$$

$$= 2\pi\mathbf{e}_{1\alpha}{}^{\mathrm{T}}\mathbf{R}\,\mathrm{Im}\left\{\left(\mathbf{I} - \mathbf{b}\mathbf{b}^{\mathrm{H}}(n\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}})/\mu\right)\mathbf{z}_1\right\} \quad (3.4.39)$$

$$= 2\pi\mathbf{e}_{1\alpha}{}^{\mathrm{T}}\mathbf{R}\,\mathrm{Im}\left\{\left(\mathbf{I} - \mathbf{b}\mathbf{b}^{\mathrm{H}}\mathbf{J}/\mu\right)\mathbf{z}_1\right\} \quad (3.4.40)$$

where we put

$$\mathbf{J} = n\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}}. \quad (3.4.41)$$

Clearly $u_1(2)$ is as $u_1(1)$ with ε replacing α, and for $u_2(1)$ and $u_2(2)$ we replace 1 by 2 in the suffices, so in general we can put

$$\mathbf{u}_r = 2\pi\mathbf{E}_r{}^{\mathrm{T}}\mathbf{R}\,\mathrm{Im}\left\{\left(\mathbf{I} - \mathbf{b}\mathbf{b}^{\mathrm{H}}\mathbf{J}/\mu\right)\mathbf{z}_r\right\}. \quad (3.4.42)$$

3.4.5 Error covariance matrix, two targets

As before the error covariance matrix is given by (3.3.12) except that we use (3.4.36) and (3.4.42) for $\mathbf{U}_r$ and $\mathbf{u}_r$, for target r (r = 1 or 2), giving

$$\mathbf{\Psi}_r = \mathbf{U}_r{}^{-1}\left\langle\mathbf{u}_r\mathbf{u}_r{}^{\mathrm{T}}\right\rangle\mathbf{U}_r{}^{-1}. \quad (3.4.43)$$

In Appendix 3.4D (equations (3.4D.1) and (3.4D.5)) we find an expression for $\left\langle\mathbf{u}_r\mathbf{u}_r{}^{\mathrm{T}}\right\rangle$ from (3.4.42), and we see that $\left\langle\mathbf{u}_r\mathbf{u}_r{}^{\mathrm{T}}\right\rangle$ depends on both $\sigma_\phi{}^2$, the phase error variance and $\sigma_a{}^2$, the fractional amplitude variance, and in different ways (i.e. proportional to $\sigma_\phi{}^2\mathbf{S}\mathbf{S}^{\mathrm{T}} + \sigma_a{}^2\mathbf{T}\mathbf{T}^{\mathrm{T}}$ where $\mathbf{S}\mathbf{S}^{\mathrm{T}}$ and $\mathbf{T}\mathbf{T}^{\mathrm{T}}$ are different functions of $\mathbf{b}$). It follows that, only if we could find that both $\mathbf{S}\mathbf{S}^{\mathrm{T}}$ and $\mathbf{T}\mathbf{T}^{\mathrm{T}}$ were proportional to $\tilde{\mathbf{M}}$, as in (3.4.36), could we find $\left\langle\mathbf{u}_r\mathbf{u}_r{}^{\mathrm{T}}\right\rangle$ proportional to $\mathbf{U}$. Thus, as this is not the case, we have the rather cumbersome form for the parameter error covariance (for target r)

$$\Psi_r = \frac{4\pi^2}{n^2}\mathbf{U}_r^{-1}\mathbf{E}_r^{\mathrm{T}}\mathbf{R}\left(\sigma_\phi^2\mathbf{S}\mathbf{S}^{\mathrm{T}} + \sigma_a^2\mathbf{T}\mathbf{T}^{\mathrm{T}}\right)\mathbf{R}^{\mathrm{T}}\mathbf{E}_r\mathbf{U}_r^{-1}. \qquad (3.4.44)$$

where **S** and **T** are respectively the real and imaginary parts of **K** (see (3.4D.2)) and $\mathbf{U}_r$ is given in (3.4.36). However, in the case where the amplitude and phase errors have the same variance, we find that $\left\langle \mathbf{u}_r\mathbf{u}_r^{\mathrm{T}} \right\rangle$ is given by $(\sigma_\phi^2/n)\mathbf{U}$ (see (3.4D.9)) and in this case we have the simpler form

$$\mathbf{\Psi}_r = \frac{\sigma_\phi^2}{n^2}\mathbf{U}_r^{-1} \qquad (3.4.45)$$

with $\mathbf{U}_r$ given in (3.4.36).

3.4.6 Simulation results

3.4.6.1 Comparison of two target and single target cases

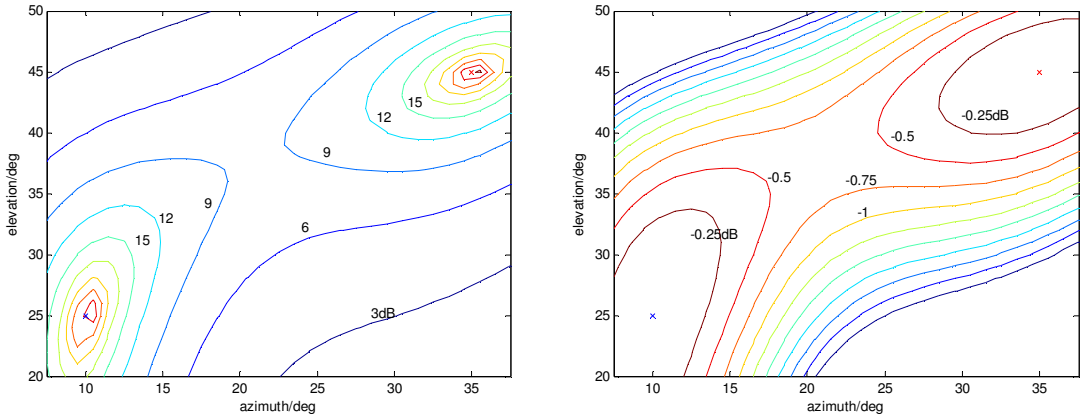
In this section we present results for some pairs of targets, giving both simulation results and the values given by the theory in equations (3.4.36), (3.4.44) and (3.4.45). In the case of two targets it is not possible to produce general plots, whether linear or contour plots, covering all target possibilities, as there are four independent target variables (azimuth and elevation for each of two targets). One simplification would be to fix one target position and give contour plots of the s.d. azimuth and elevation errors over all positions for the second target. However, there would be a degree of arbitrariness in the choice of position of the first target, and in principle a set of results over a considerable number of first target positions would be desirable. Instead of attempting any such general forms of results, we limit the examples here to a number of target positions, showing that the theoretical results agree well with the simulation results, and exploring a few interesting cases.

1: (10,25)	2 target simulation		2 target theory		single targets sim'n.	
2: (35,45)	1	2	1	2	1	2
azimuth s.d.	0.54	0.87	0.54	0.87	0.47	0.64
elevation s.d.	1.47	0.71	1.46	0.71	1.03	0.57

Table 3.2 Example results for two target angle estimation with channel errors

The results for the first example are given in Table 3.2. This is for the irregular 2D array R1, illustrated in Figure 3.8. In this case the targets are at (10º,25º) and (35°,45°) in azimuth-elevation coordinates and the s.d. of the phase errors is 5° (with equal level amplitude errors, i.e. equal to the phase error s.d in radians). The results in the first four columns show excellent agreement between simulation and theory. The last two

columns show the s.d. errors that would be given in the cases of single targets in the positions of targets 1 and 2 – i.e. the results if target 2 and target 1 respectively, were not present. These results are better (i.e. with lower s.d. errors) showing that the interaction between the targets in the two target case makes the angle estimation more



(a) reciprocal form (b) linear form

Figure 3.16 MUSIC function with two targets

sensitive to system errors, as might be expected. Figure 3.16 shows the MUSIC function plotted over an angular region including the two targets (with small crosses marking their positions). In Figure 3.16(a) the function is plotted as in Figure 2.4(b) and the contours are at 3dB intervals. In Figure 3.16(b) it is in linear form, as used in the peak finding process and the contours are ¼dB intervals. The function peaks are not quite at the target positions, due to the errors included in the simulation. In particular we note the tendency of the dip between the two peaks to fill in, with a ridge being seen between them. This rapidly fills as the targets are moved closer, until the two peaks merge and only one peak is seen. In this case the targets are not resolved.

In Table 3.3 we investigate the effect of moving the targets closer together. Again the channel error s.d. was set at 5° but the array was 2×R1 – i.e. R1 doubled in size. This should reduce the variance of the estimation errors (proportional to $\mathbf{RR}^{\mathrm{T}}$) by a factor of 4 and the s.d. by a factor of 2. (This is what might be expected – doubling the aperture of the array should increase the precision of measurement by a factor of 2.) This is seen (in part (a) of the table) to be approximately the case, though the target positions and relative separations (and hence the degree of interaction) are different from Table 3.2. Again we see that in the single target cases, with targets at the positions 1 and 2, the results are better, in the absence of interaction between the targets. In Table 3.3(a) the targets are separated by 8 degrees in both azimuth and elevation, in (b) by 7 degrees, in

1: (10,15)	2 target simulation		2 target theory		single targets sim.	
2: (18,23)	1	2	1	2	1	2
azimuth s.d.	0.29	0.32	0.29	0.32	0.22	0.24
elevation s.d.	0.95	0.59	0.95	0.59	0.84	0.54

(a) Separation 8 degrees in azimuth and elevation

1: (10,15)	2 target simulation		2 target theory		single targets sim.	
2: (17,22)	1	2	1	2	1	2
azimuth s.d.	0.34	0.36	0.32	0.35	0.22	0.24
elevation s.d.	0.99	0.63	0.96	0.62	0.84	0.57

(b) Separation 7 degrees in azimuth and elevation

1: (10,15)	2 target simulation		2 target theory		single targets sim.	
2: (16,21)	1	2	1	2	1	2
azimuth s.d.	0.44	0.42	0.37	0.40	0.22	0.23
elevation s.d.	1.02	0.67	0.98	0.66	0.84	0.59

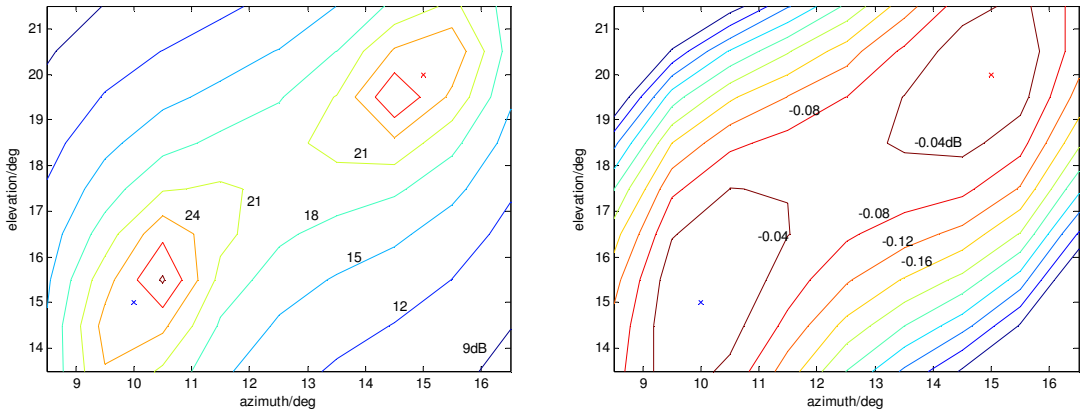
(c) Separation 6 degrees in azimuth and elevation

1: (10,15)	2 target simulation		2 target theory		single targets sim.	
2: (15,20)	1	2	1	2	1	2
azimuth s.d.	0.51	0.80	0.44	0.46	0.22	0.23
elevation s.d.	1.06	0.94	1.02	0.71	0.85	0.62

(d) Separation 5 degrees in azimuth and elevation

Table 3.3 Effect of target separation

(c) by 6 degrees and in (d) by 5 degrees. (The single target results are the same in all four cases for target 1, which is not moved, and change a little for target 2, as its position varies slightly as seen by the array.) At 8 degrees (and also at higher values of separation, such as 10° and 15°, not shown) the agreement between simulation (over



(a) reciprocal form (b) linear form

Figure 3.17 MUSIC function with two close targets

10000 trials) and theory is excellent. At 7°, as the targets are closer, the theory shows poorer performance (higher error variance), and so do the simulation results, but now the simulation results diverge slightly from the theory, being a little worse. This trend increases as the separation is reduced to 6° and 5°, and is attributable to the gradual failure of the small error approximations that are clearly valid at higher separations. In fact Figure 3.17 shows that, at 5º separation in both angles, the two peaks have almost merged, with the dip between them very much shallower than in Figure 3.16. (The contours are at 3dB intervals in Figure 3.16(a) and 0.04dB in part (b) of the figure, for comparison with Figure 3.16). NB: We can see from the angular nature of the contour lines that the function is only evaluated at 1° intervals. However the peak positions have been estimated using 2D quadratic interpolation, which gives very good accuracy with modest computational requirements.) The results are also found to be much more variable, even over runs of 10000 trials, suggesting that some errors may have high values. This is observed mainly for target 2. The estimated half-beamwidth is about 13 degrees along the line of the targets (computed using (3.4E.6) from Appendix 3.4E) so the target separation (about 7°) is only about 1/4 beamwidth. We conclude that the theoretical expressions give the s.d. of the errors very accurately for separations to well under a half beamwidth, successfully taking into account the interaction of the target responses, with some falling off in accuracy as the targets are brought closer still, but

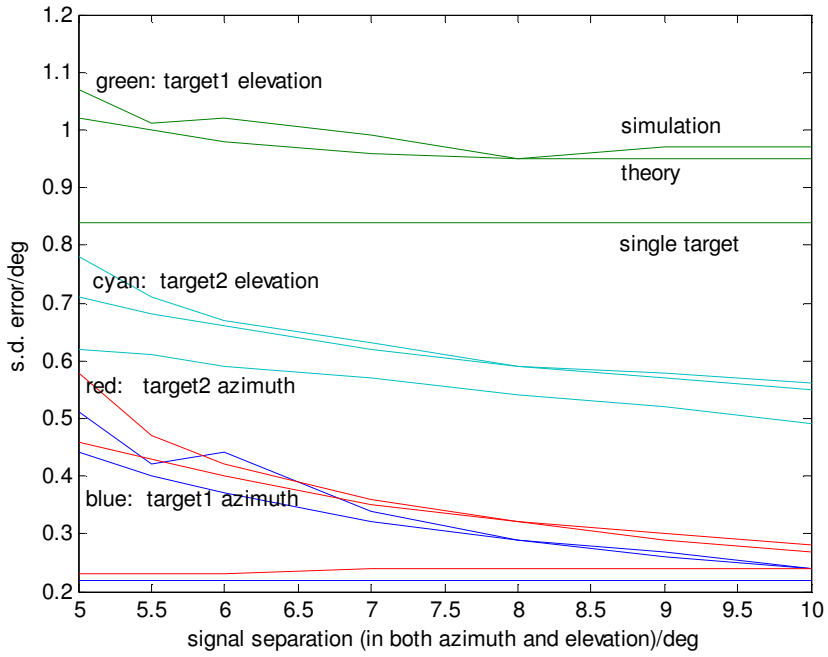


Figure 3.18 Effect of target separation on errors

then we are starting to approach the limit of resolution.

The values in Table 3.3 (with some extra values) are plotted in Figure 3.18. Each of the three sets of curves shows the simulation results (the highest values at 5 degrees separation), the theoretical values and, the lowest curve, the value in the case when only a single signal is present. These confirm the deductions already made – the good agreement between the simulation and the two signal theory above, say, 7° separation, with some divergence when the targets are closer, and the lower error values when the targets are present on their own, as might be expected. The two signal errors fall a little as the separation increases but not as low as the single target values, even at high separations, in the case considered here of amplitude errors as well as phase errors.

3.4.6.2 Effects of phase and amplitude errors

1: (10,20)	2 target simulation		2 target theory		single target theory	
2: (18,28)	1	2	1	2	1	2
azimuth s.d.	0.31	0.35	0.31	0.35	0.23	0.25
elevation s.d.	0.80	0.55	0.80	0.54	0.63	0.45

(a) phase and amplitude errors, same in the two channels

1: (10,20)	2 target simulation		2 target theory		targets theory single	
2: (18,28)	1	2	1	2	1	2
azimuth s.d.	0.31	0.36	0.31	0.35	0.22	0.25
elevation s.d.	0.80	0.54	0.80	0.54	0.63	0.45

(b) phase and amplitude errors, different in the two channels

1: (10,20)	2 target simulation		2 target theory		single targets theory	
2: (18,28)	1	2	1	2	1	2
azimuth s.d.	0.22	0.24	0.23	0.25	0.23	0.25
elevation s.d.	0.64	0.46	0.64	0.45	0.62	0.44

(c) phase errors only, same in the two channels

1: (10,20)	2 target simulation		2 target theory		single targets sim.	
2: (18,28)	1	2	1	2	1	2
azimuth s.d.	0.21	0.24	0.21	0.24	0.002	0.002
elevation s.d.	0.47	0.28	0.47	0.28	0.008	0.005

(d) amplitude errors only, same in the two channels

Table 3.4 Effect of different error models

In Table 3.4 we show the effect of different forms of error, taking the case of the targets separated by 8° in both azimuth and elevation, with targets at (10°,20°) and (18°,28°). In part (a) of the table we again see excellent agreement between simulation and theory

and also the rather better performance achieved in the single signal cases. In this case the errors for the simulation were set to be the same in the two channels. In part (b) the errors were taken to be independent between the two channels, though again with zero mean normal distributions of equal variance in phase and relative amplitude (5° s.d.). We see that there is little difference in the simulation results, any differences being likely to be due to statistical fluctuation. There is no difference in the theory results, in cases (a) and (b), which does not distinguish between the cases, and again there is almost complete agreement (to 2 decimal places) between simulation and theory. In part (c) we applied phase errors only and again there is good match between simulation and theory. We note the estimation errors in the two signal case are lower than in the case of both amplitude and phase errors, as seems reasonable. However in the single target cases the errors are essentially the same as for the case with amplitude errors also present, confirming that amplitude errors have no effect on the angle estimation in the single target cases (at least for small errors, allowing first order approximations to be used in the theory), as previously noted. What is also of interest is that the two target errors with only phase channel errors are essentially the same as the single target errors, suggesting that interactions between the target responses occur only as a result of amplitude errors.

Finally, in part (d) we see that with amplitude errors only there are significant estimation errors in the two target case (and again the simulation results match the theory) but in the single target cases the errors are very small. In fact if we run the simulation with zero s.d. channel errors we obtain exactly the same results, showing that the non-zero results given in the single target case are just the result of the limit on the accuracy of the peak position determination used. (Higher accuracy would be possible at higher computational cost, but this degree of accuracy is clearly satisfactory for the rest of the results.)

A further study of amplitude and phase errors only is given in Table 3.5. Here we see that with phase errors only, even with the target separation as low as 4 degrees (in both azimuth and elevation) the simulation results match the theoretical results very well (in part (a) of the table). Also the single target theory matches the two target theory almost exactly. In part (b) we give the result for targets at only 4 degrees separation with only amplitude errors present and we see that there is now a divergence of the simulation results from the theory, unlike the case of 8 degrees separation, seen in Table 3.4(d). (The single target results in theory should be zero; here the simulation results confirm

this, within the accuracy used.) These results indicate that the difference between the simulation results and theory, when the targets are close, shown in Figure 3.17, are due to the presence of amplitude errors.

1: (10,20)	2 target simulation		2 target theory		single target theory	
2: (14,24)	1	2	1	2	1	2
azimuth s.d.	0.21	0.22	0.23	0.24	0.23	0.24
elevation s.d.	0.63	0.53	0.63	0.52	0.63	0.52

(a) phase errors only

1: (10,20)	2 target simulation		2 target theory		single target theory	
2: (14,24)	1	2	1	2	1	2
azimuth s.d.	0.62	0.69	0.54	0.57	(0)	(0)
elevation s.d.	0.79	0.63	0.72	0.51	(0)	(0)

(b) amplitude errors only

Table 3.5 Target at 4 degrees separation with phase or amplitude errors only

3.4.6.3 Phase errors only

Here we look more closely at the result shown in Table 3.4(c) above and in Table 3.5(a). These appeared to show that in the case of phase errors only the two target results might be exactly the same (allowing for small differences due to taking a limited statistical sample) as if the targets were present independently. A study of the theoretical expressions did not show that those for the two target theory in the absence of amplitude errors could be made to equal those for the single target case. A possibility appeared to be that the two signal theory was not quite correct and that it should be the same as the

1: (10,20)	2 target simulation		2 target theory		single target theory	
2: (150,65)	1	2	1	2	1	2
azimuth s.d.	0.283	0.622	0.285	0.626	0.226	0.459
elevation s.d.	0.675	0.257	0.676	0.259	0.628	0.256

(a) widely separated targets

1: (10,20)	2 target simulation		2 target theory		single target theory	
2: (18,28)	1	2	1	2	1	2
azimuth s.d.	0.227	0.251	0.230	0.252	0.226	0.247
elevation s.d.	0.642	0.457	0.642	0.455	0.628	0.446

(b) close targets

Table 3.6 Higher accuracy results, phase errors only

single target theory when only phase errors are present. However, taking results to three decimal places, both in simulation and theory, confirmed (1) that the two signal theory

matches the simulation results very well and (2) also that both the two signal theory and simulation differ slightly from the single signal theory. This is shown in Table 3.6 for two cases, a pair of widely separated targets and two fairly close targets (8 degrees separation in both dimensions).

These results show that the two targets interact, even when widely separated in angle, but with only phase errors, the error performance is very close to the case of single targets.

3.4.6.4 Conclusions of two signal error study on MUSIC accuracy

In conclusion, we see that the theory for the two target angle estimation accuracy is very accurate, for target separations down to considerably less than one beamwidth (3dB) separation, and moderately accurate below that, approaching the separation at which the targets cease to be resolved. We also see that, while only channel phase errors affect the estimation performance when only a single target is present (at least for moderate amplitude error levels), amplitude errors do affect the estimation performance when two signals are present. However if only phase errors are present the two target results are very close to the single target results, indicating only a very small effect on the accuracy of estimation of each target due to the presence of the other in this case.

3.4.7 Comparison with the Cramer-Rao bound, two parameter case

In §3.2.5 we found that the expressions for the variance in the estimate of the value of a single parameter for a single target when limited only by noise (given by the CRB) and the variance when limited only by errors in the array data have precisely the same functional form. The two expressions only differ by scaling factors proportional to the signal to noise ratio and the variance on the array errors, respectively. The comparison showed that the array phase error variance σ_ϕ^2 is related to the integrated signal to noise ratio (iSNR) in each channel by

$$\text{iSNR} = \frac{pa^2}{\psi} = \frac{1}{2\sigma_\phi^2}, \tag{3.4.46}$$

(see (3.2.37)). Here we show that these results apply also to the single target, two parameter case, and for one form of the two target case, though not quite for the general two target case. In this section we consider only the EPP form of array.

In the case of a single target with two parameters, the CRB is given (from (2.D.2) by

$$\mathbf{B}(\alpha,\varepsilon)=\frac{\psi}{2npa^2}\left(\begin{bmatrix}\mathbf{a}_\alpha{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{a}_\alpha & \mathrm{Re}(\mathbf{a}_\alpha{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{a}_\varepsilon)\\ \mathrm{Re}(\mathbf{a}_\varepsilon{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{a}_\alpha) & \mathbf{a}_\varepsilon{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{a}_\varepsilon\end{bmatrix}\right)^{-1} \tag{3.4.47}$$

(We have replaced $\tilde{s}$ by the mean square signal amplitude a^2, and $\mathbf{d}_1$ and $\mathbf{d}_2$ by $\mathbf{a}_\alpha$ and $\mathbf{a}_\varepsilon$, to correspond to the notation of this section. We have also included the factor n required because of the different scaling of vectors in the CRB expression and the array error analysis, as discussed in §2.5.) Putting $\mathbf{Q}_a = \mathbf{I} - \mathbf{a}\mathbf{a}^{\mathrm{H}}$ and using $\mathbf{a}^{\mathrm{H}}\mathbf{a}_\alpha = 0$ and $\mathbf{a}^{\mathrm{H}}\mathbf{a}_\varepsilon = 0$, this becomes

$$\mathbf{B}(\alpha,\varepsilon)=\frac{\psi}{2npa^2}\left(\begin{bmatrix}\|\mathbf{a}_\alpha\|^2 & \mathrm{Re}(\mathbf{a}_\alpha{}^{\mathrm{H}}\mathbf{a}_\varepsilon)\\ \mathrm{Re}(\mathbf{a}_\varepsilon{}^{\mathrm{H}}\mathbf{a}_\alpha) & \|\mathbf{a}_\varepsilon\|^2\end{bmatrix}\right)^{-1}. \tag{3.4.48}$$

The parameter estimate covariance matrix in the case of array errors is given in (3.3.17) by

$$\mathbf{\Psi}(\alpha,\varepsilon)=\frac{\sigma_\phi{}^2}{n}\mathbf{U}^{-1} \tag{3.4.49}$$

where $\mathbf{U}$ is given in (3.3.8). Using (3.3A.5) and equivalents for the other components of $\mathbf{U}$ we have

$$\mathbf{U}=\begin{bmatrix}\|\mathbf{a}_\alpha\|^2 & \mathrm{Re}(\mathbf{a}_\alpha{}^{\mathrm{H}}\mathbf{a}_\varepsilon)\\ \mathrm{Re}(\mathbf{a}_\varepsilon{}^{\mathrm{H}}\mathbf{a}_\alpha) & \|\mathbf{a}_\varepsilon\|^2\end{bmatrix} \tag{3.4.50}$$

so that we see, from (3.4.48) and (3.4.50) that

$$\mathbf{B}(\alpha,\varepsilon)=\frac{\psi}{2npa^2}\mathbf{U}^{-1}. \tag{3.4.51}$$

As in the single parameter case, the forms of the expressions in (3.4.49) and (3.4.51) are the same, and the relationship between the scaling factors is the same as in that case, given in (3.4.46) above.

We now consider the case of two targets, and two parameters. For target r the CRB error covariance matrix is given. from (2.D.18), by

$$\mathbf{B}(\alpha_r,\varepsilon_r)=\frac{\psi}{2npa^2}\left(\begin{bmatrix}\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\alpha} & \mathrm{Re}(\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\varepsilon})\\ \mathrm{Re}(\mathbf{a}_{r\varepsilon}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\alpha}) & \mathbf{a}_{r\varepsilon}{}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}\mathbf{a}_{r\varepsilon}\end{bmatrix}\right)^{-1}=\frac{\psi}{2npa^2}\mathbf{U}_r{}^{-1} \tag{3.4.52}$$

where $\mathbf{U}_r$ is given in (3.4.26). From (3.4.43) we have, for the covariance matrix of parameter errors due to array errors only,

$$\boldsymbol{\Psi}(\alpha_r, \varepsilon_r) = \mathbf{U}^{-1} \left\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \right\rangle \mathbf{U}^{-1} \tag{3.4.53}$$

so if we could show that

$$\left\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \right\rangle = \frac{\sigma_\phi^{\,2}}{n} \mathbf{U} \tag{3.4.54}$$

then, again, **B** and **Ψ** would have the same functional form, and the relationship between the array error variance and the integrated signal to noise ratio would be the same as for the single target case. In fact we show in Appendix 3.4D that (3.4.54) does not quite hold (though the relationship is close) in the most general case, where the amplitude and phase errors have different variances, so that the CRB (with appropriate choice of SNR) is not quite a correct model for the effect of array errors in the general two (or higher) signal case. However in the case where these variances are equal we find (see the discussion leading to (3.4D.8)) that (3.4.54) does hold and it follows that

$$\boldsymbol{\Psi}(\alpha_r, \varepsilon_r) = \frac{\sigma_\phi^{\,2}}{n} \mathbf{U}^{-1} \tag{3.4.55}$$

in this case. This has the same form as (3.4.52) for the CRB with the same relationship between the variances (in both phase and amplitude) of the components of the PSVs and the effective integrated signal to noise ratio.

3.4.8 Accuracy of IMP in the two target case

Following the fruitful approach to the analysis of the effect of PSV errors on the performance of MUSIC taken in earlier sections, an attempt was made to apply the same method to the IMP function with two targets. However, this function, in the form of a ratio with a denominator that is error sensitive as well as its numerator, is more complex than that of the MUSIC function with two targets. Furthermore the errors on the PSVs near both targets seem to be combined in the expression, as the errors on signal 1 affect the projection matrix $\mathbf{Q}_1$ which enters the expression (in both numerator and denominator) when attempting to find the function peak in the region of signal 2. The IMP expression also includes both signal power levels (which can be combined into a single parameter, the power ratio, when we consider no noise, or negligible noise, to be present), which is not the case for MUSIC. It seemed that the expressions relating the parameter error estimates to the PSV errors were becoming rather complex, with little likelihood of obtaining compact and convenient expressions, so this study was left but simulations were nevertheless carried out to determine whether IMP appeared to be

more or less accurate than MUSIC.

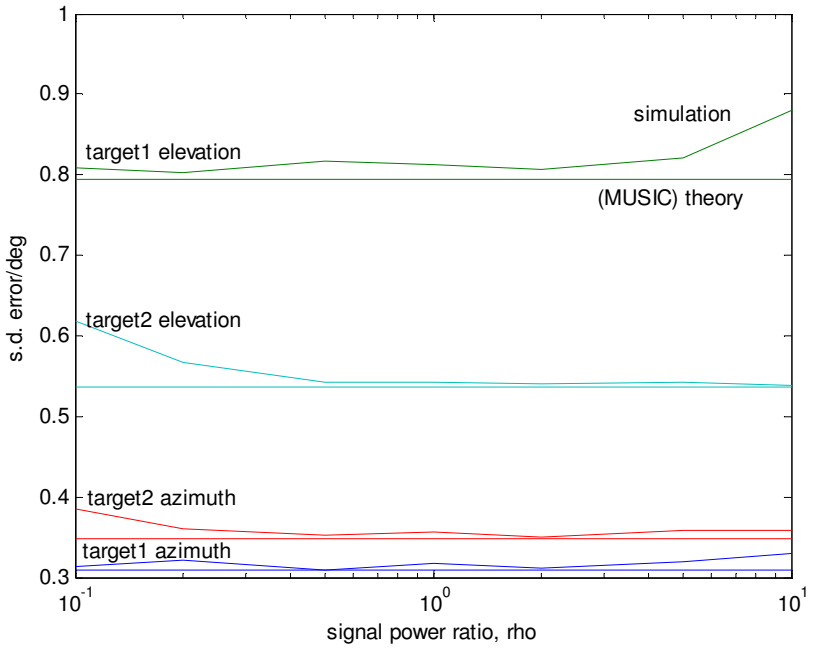
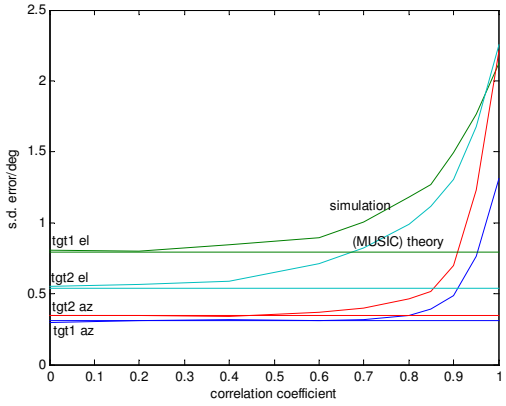


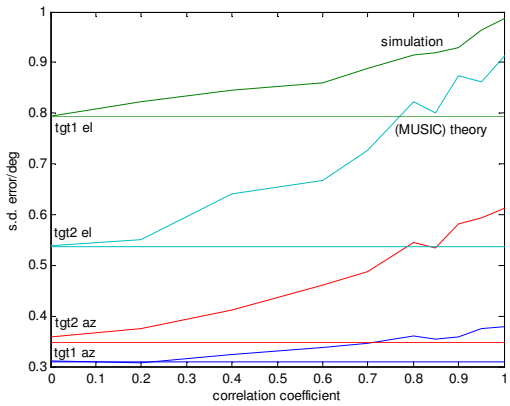
Figure 3.19 Effect of differential target signal strength on errors using IMP

Before comparing IMP and MUSIC the first investigations were to find out whether the relative signal strengths affected the performance of IMP, and also how a degree of correlation might affect the performance. IMP differs from MUSIC in being able, in principle, to determine the parameters of partially correlated, or even fully correlated, signals. (Strictly, MUSIC can also be used with partially correlated signals, but with less effectiveness as the degree of correlation increases. With two fully coherent signals MUSIC will search for only one signal with a PSV which is a power-weighted combination of the PSVs of the two signals, and generally this will not correspond closely to any manifold PSV. IMP, in principle, at least, should find the two PSVs which minimize the received power.) Figure 3.19 shows the effect of varying the ratio $\rho = p_2/p_1$ where p_1 and p_2 are the powers of signals 1 and 2, considered to be high compared with receiver noise. We see that over a large range (from about 0.2 to about 5) there is very little sensitivity to this parameter (or possibly none, considering that any small variations may due to the finite set of cases (4000) taken for the statistics). We note that when p_2 is relatively small enough ($\rho = 0.1$) the accuracy to target 2 starts to fall (or the error s.d. values rise) and correspondingly when p_1 is relatively small ($\rho = 10$) the accuracy to target 1 starts to fall. (In this case we used the array 2×R1 again and the targets were at (10°,20°) and (18°,28°) with 5deg rms phase errors and equivalent

level amplitude errors.) Below each simulation curve is plotted the theoretical s.d. value for MUSIC, and it is seen that IMP essentially gives the same performance.



(a) real coefficient



(b) imaginary coefficient

Figure 3.20 Effect of signal correlation on errors using IMP

Defining a correlation coefficient by $\gamma = \mathbf{u}_1^{\mathrm{H}}\mathbf{u}_2 \big/ n\sqrt{p_1 p_2}$ where $p_k = \|\mathbf{u}_k\|^2 / n \;\; (k = 1,2)$ and $\mathbf{u}_k$ is the set of n samples of the waveform of signal k, we see that γ may be complex, in general, and rises to value 1 when the signals are perfectly correlated. In Figure 3.20 we show the effect of correlation on IMP accuracy. In part (a) the coefficient is real and plotted over the range 0 to 1. We see that the errors rise very little up to $\gamma = 0.4$, rising more sharply as the correlation approaches unity. It is more difficult for IMP to separate fully correlated signals (and impossible for MUSIC) which is reflected in the sharply increased errors for this condition. (In compiling the error statistics a small fraction of the trials were rejected, and repeated with another set of errors, when IMP failed to resolve the targets.) In the case of the correlation coefficient being imaginary, shown in part (b), there is still an increase of errors as the signals become more closely correlated, but less dramatically as γ approaches unity. The sensitivity to correlation is likely to differ with signal separation, but this has not been explored here. In particular

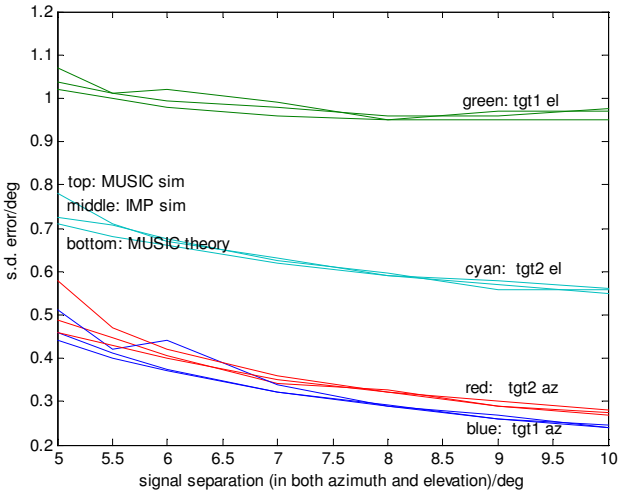


Figure 3.21 Comparison of accuracy of MUSIC and IMP with signal separation

we note that for a modest degree of correlation the performance is not greatly affected, and is close to that given by the MUSIC theory for two (uncorrelated) targets.

In Figure 3.21 we show the performance of both MUSIC and IMP with varying signal separation. The MUSIC results are those already presented in Figure 3.18 and the IMP results are for the same conditions (with the signal ratio ρ set at 0.5 and the correlation coefficient to zero). As already seen before, the MUSIC curves are close to the theoretical curves for all four target parameters, and here we see that IMP is equally close, in general – in fact slightly closer at the lowest target separations. This might be because of the better resolution achievable by IMP, with MUSIC approaching the limit of its capability. In any case, we see again that the IMP performance matches the MUSIC theory very well.

1: (10,20)	2 target simulation		2 target theory		IMP simulation	
2: (18,28)	1	2	1	2	1	2
azimuth s.d.	0.31	0.35	0.31	0.35	0.31	0.36
elevation s.d.	0.80	0.55	0.80	0.54	0.81	0.54

(a) phase and amplitude errors, same in the two channels

1: (10,20)	2 target simulation		2 target theory		IMP simulation	
2: (18,28)	1	2	1	2	1	2
azimuth s.d.	0.31	0.36	0.31	0.35	0.31	0.35
elevation s.d.	0.80	0.54	0.80	0.54	0.80	0.54

(b) phase and amplitude errors, different in the two channels

1: (10,20)	2 target simulation		2 target theory		IMP simulation	
2: (18,28)	1	2	1	2	1	2
azimuth s.d.	0.22	0.24	0.23	0.25	0.23	0.25
elevation s.d.	0.64	0.46	0.64	0.45	0.64	0.45

(c) phase errors only, same in the two channels

1: (10,20)	2 target simulation		2 target theory		IMP simulation	
2: (18,28)	1	2	1	2	1	2
azimuth s.d.	0.21	0.24	0.21	0.24	0.21	0.25
elevation s.d.	0.47	0.28	0.47	0.28	0.49	0.30

(d) amplitude errors only, same in the two channels

Table 3.7 Effect of different error models comparing IMP and MUSIC

Finally, in Table 3.7, we compare IMP and MUSIC with different error conditions. Again the MUSIC results have been given in Table 3.4, but here we add IMP results for the same target positions. We see that the IMP results, as well as the MUSIC results, are

very close to the MUSIC theory for all the error conditions. (There is an indication that IMP may be slightly more sensitive in the case of amplitude errors only, but the IMP results were taken over only 4000 cases, rather than 10000 for MUSIC (as the IMP program takes longer) so the statistics may be more variable.) As with the other results, these generally indicate that the performance of IMP is essentially the same as that of MUSIC, and that the MUSIC theory derived here is an accurate description of this performance.

Appendix 3.4A: First order approximation for perturbed projection matrix

Let $\tilde{\mathbf{A}} = \mathbf{A} + \delta\mathbf{A}$ where $\delta\mathbf{A}$ is a small perturbation to $\mathbf{A}$. (In general $\mathbf{A}$ is $n \times m$ with $m \leq n$.) We require to find an approximation to the projection matrix

$$\mathbf{P}_{\tilde{\mathbf{A}}} = \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^{\mathrm{H}}\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^{\mathrm{H}} \tag{3.4A.1}$$

to first order in $\delta\mathbf{A}$. First we have

$$\tilde{\mathbf{A}}^{\mathrm{H}}\tilde{\mathbf{A}} = (\mathbf{A}+\delta\mathbf{A})^{\mathrm{H}}(\mathbf{A}+\delta\mathbf{A}) = \mathbf{A}^{\mathrm{H}}\mathbf{A} + \mathbf{A}^{\mathrm{H}}\delta\mathbf{A} + \delta\mathbf{A}^{\mathrm{H}}\mathbf{A} + \delta\mathbf{A}^{\mathrm{H}}\delta\mathbf{A}$$

$$= \mathbf{A}^{\mathrm{H}}\mathbf{A}(\mathbf{I}+\mathbf{Z}) \tag{3.4A.2}$$

where

$$\mathbf{Z} = (\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}(\mathbf{A}^{\mathrm{H}}\delta\mathbf{A} + \delta\mathbf{A}^{\mathrm{H}}\mathbf{A}) \tag{3.4A.3}$$

to first order in $\delta\mathbf{A}$. Then the inverse, from (3.4A.2), is

$$(\tilde{\mathbf{A}}^{\mathrm{H}}\tilde{\mathbf{A}})^{-1} = (\mathbf{I}+\mathbf{Z})^{-1}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1} = (\mathbf{I}-\mathbf{Z}+\mathbf{Z}^{2}-\ \ldots\ )(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}$$

$$= (\mathbf{I}-\mathbf{Z})(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1} \tag{3.4A.4}$$

to first order in $\mathbf{Z}$ and hence also in $\delta\mathbf{A}$. From (3.4A.1), using (3.4A.4) to first order in $\delta\mathbf{A}$,

$$\mathbf{P}_{\tilde{\mathbf{A}}} = (\mathbf{A}+\delta\mathbf{A})(\tilde{\mathbf{A}}^{\mathrm{H}}\tilde{\mathbf{A}})^{-1}(\mathbf{A}+\delta\mathbf{A})^{\mathrm{H}} = (\mathbf{A}+\delta\mathbf{A})(\mathbf{I}-\mathbf{Z})(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}(\mathbf{A}+\delta\mathbf{A})^{\mathrm{H}}$$

$$= \mathbf{A}(\mathbf{I}-\mathbf{Z})(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}} + \delta\mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}} + \mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\delta\mathbf{A}^{\mathrm{H}},$$

where we have neglected terms including both $\mathbf{Z}$ and $\delta\mathbf{A}$ as these are of second and third order. Substituting $\mathbf{Z}$ from (3.4A.3) and putting

$$\mathbf{A}^{+} = (\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}} \tag{3.4A.6}$$

where $\mathbf{A}^{+}$ is a pseudoinverse of $\mathbf{A}$, such that $\mathbf{A}^{+}\mathbf{A} = \mathbf{I}_m$ (where $\mathbf{A}$ is $n \times m$ and $m < n$), we have

$$\begin{aligned}\mathbf{P}_{\tilde{\mathbf{A}}} &= \mathbf{P}_{\mathbf{A}} - \mathbf{A}\mathbf{Z}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}} + \delta\mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}} + \mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\delta\mathbf{A}^{\mathrm{H}} \\ &= \mathbf{P}_{\mathbf{A}} - \mathbf{P}_{\mathbf{A}}\delta\mathbf{A}\mathbf{A}^{+} - \mathbf{A}^{+\mathrm{H}}\delta\mathbf{A}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}} + \delta\mathbf{A}\mathbf{A}^{+} + \mathbf{A}^{+\mathrm{H}}\delta\mathbf{A}^{\mathrm{H}} \\ &= \mathbf{P}_{\mathbf{A}} + \mathbf{Q}_{\mathbf{A}}\delta\mathbf{A}\mathbf{A}^{+} + \mathbf{A}^{+\mathrm{H}}\delta\mathbf{A}^{\mathrm{H}}\mathbf{Q}_{\mathbf{A}}, \end{aligned} \tag{3.4A.7}$$

where $\mathbf{Q}_{\mathbf{A}} = \mathbf{I} - \mathbf{P}_{\mathbf{A}}$.

Appendix 3.4B: Results on products including a vector in A

(1) We note that $\mathbf{P}_{\mathbf{A}}\mathbf{A} = \mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{A} = \mathbf{A}$, where $\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \ldots \quad \mathbf{a}_m]$, from which we can see that

$$\mathbf{P}_{\mathbf{A}}\mathbf{a}_r = \mathbf{a}_r \tag{3.4B.1}$$

where $\mathbf{a}_r$ is the rth column of $\mathbf{A}$.

(2) Similarly $\mathbf{Q}_{\mathbf{A}}\mathbf{A} = (\mathbf{I}_n - \mathbf{P})\mathbf{A} = \mathbf{A} - \mathbf{A} = \mathbf{0}_{n\times m}$ from which, considering column r of $\mathbf{A}$, we have

$$\mathbf{Q}_{\mathbf{A}}\mathbf{a}_r = \mathbf{0}_n \ . \tag{3.4B.2}$$

(3) We have $\mathbf{A}^{+}\mathbf{A} = (\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{A} = \mathbf{I}_m = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \ldots \quad \mathbf{u}_m]$ where $\mathbf{u}_r$ is column r of the $m \times m$ identity matrix. Taking column r of $\mathbf{A}$ and $\mathbf{I}$ we see that

$$\mathbf{A}^{+}\mathbf{a}_r = \mathbf{u}_r \tag{3.4B.3}$$

and so

$$\mathbf{B}\mathbf{A}^{+}\mathbf{a}_r = \mathbf{b}_r \tag{3.4B.4}$$

where $\mathbf{b}_r$ is column r of $\mathbf{B}$ (which is a matrix of m columns).

Appendix 3.4C: Auxiliary results for case of two signals

(a) Projection matrix

We first obtain an expression for $\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}$. In this case the matrix of signal PSVs contains only two vectors, thus $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2]$ so

$$\mathbf{A}^{\mathrm{H}}\mathbf{A} = \begin{bmatrix} \mathbf{a}_1{}^{\mathrm{H}} \\ \mathbf{a}_2{}^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1{}^{\mathrm{H}}\mathbf{a}_1 & \mathbf{a}_1{}^{\mathrm{H}}\mathbf{a}_2 \\ \mathbf{a}_2{}^{\mathrm{H}}\mathbf{a}_1 & \mathbf{a}_2{}^{\mathrm{H}}\mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} 1 & \gamma \\ \gamma^* & 1 \end{bmatrix}$$

and so

$$\left(\mathbf{A}^{\mathrm{H}}\mathbf{A}\right)^{-1} = \begin{bmatrix} 1 & \gamma \\ \gamma^* & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -\gamma \\ -\gamma^* & 1 \end{bmatrix} \Big/ (1 - |\gamma|^2) \tag{3.4C.1}$$

where $\gamma = \mathbf{a}_1{}^{\mathrm{H}}\mathbf{a}_2$. Using (3.3A.1) for the components of $\mathbf{a}_r$ we have

$$\gamma = \sum_k a_{1k}{}^* a_{2k} = \sum_k \exp\left(2\pi i \mathbf{r}_k{}^{\mathrm{T}}(\mathbf{e}_2 - \mathbf{e}_1)\right)\Big/ n = \sum_k \exp\left(2\pi i \mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e}\right)\Big/ n \tag{3.4C.2}$$

where $\Delta\mathbf{e} = \mathbf{e}_2 - \mathbf{e}_1$. If we put

$$b_k = a_{1k}{}^* a_{2k} \tag{3.4C.3}$$

then we can put

$$\gamma = \sum_k a_{1k}{}^* a_{2k} = \sum_k b_k = \mathbf{1}^{\mathrm{T}}\mathbf{b}\,. \tag{3.4C.4}$$

The projection matrix is thus, from (3.4C.1),

$$\mathbf{P} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \begin{bmatrix} 1 & -\gamma \\ -\gamma^* & 1 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1{}^{\mathrm{H}} \\ \mathbf{a}_2{}^{\mathrm{H}} \end{bmatrix} \Big/ \mu = (\mathbf{a}_1\mathbf{a}_1{}^{\mathrm{H}} - \gamma\mathbf{a}_1\mathbf{a}_2{}^{\mathrm{H}} - \gamma^*\mathbf{a}_2\mathbf{a}_1{}^{\mathrm{H}} + \mathbf{a}_2\mathbf{a}_2{}^{\mathrm{H}})/\mu\,, \tag{3.4C.5}$$

where μ is given, using (3.4C.4), by

$$\mu = 1 - |\gamma|^2 = 1 - \mathbf{1}^{\mathrm{T}}\mathbf{b}\mathbf{b}^{\mathrm{H}}\mathbf{1}\,, \tag{3.4C.6}$$

(b) $\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\Delta\mathbf{a}_r$

To evaluate $\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\Delta\mathbf{a}_r$, we have, using (3.3A.4) in (3.4C.5),

$$\begin{aligned} \mathbf{a}_{1\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}} &= \mathbf{a}_{1\alpha}{}^{\mathrm{H}}(\mathbf{a}_1\mathbf{a}_1{}^{\mathrm{H}} - \gamma\mathbf{a}_1\mathbf{a}_2{}^{\mathrm{H}} - \gamma^*\mathbf{a}_2\mathbf{a}_1{}^{\mathrm{H}} + \mathbf{a}_2\mathbf{a}_2{}^{\mathrm{H}})/\mu\,. \\ &= \mathbf{a}_{1\alpha}{}^{\mathrm{H}}\mathbf{a}_2(\mathbf{a}_2{}^{\mathrm{H}} - \gamma^*\mathbf{a}_1{}^{\mathrm{H}})/\mu \end{aligned} \tag{3.4C.7}$$

Now we note, from (3.4C.3) and (3.3A.1), that $a_{1k}b_k = |a_{1k}|^2 a_{2k} = a_{2k}/n$, so that

$$\begin{aligned} (\mathbf{a}_2{}^{\mathrm{H}} - \gamma^*\mathbf{a}_1{}^{\mathrm{H}})\Delta\mathbf{a}_1 &= \sum_k (a_{2k}{}^* - \gamma^* a_{1k}{}^*)\Delta a_{1k} = \sum_k (nb_k{}^* - \gamma^*)a_{1k}{}^*\Delta a_{1k} \\ &= \sum_k (nb_k{}^* - \gamma^*)z_{1k} = n\mathbf{b}^{\mathrm{H}}\mathbf{z}_1 - \mathbf{b}^{\mathrm{H}}\mathbf{1}\mathbf{1}^{\mathrm{T}}\mathbf{z}_1 = \mathbf{b}^{\mathrm{H}}(n\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}})\mathbf{z}_1 \end{aligned} \tag{3.4C.8}$$

where we have substituted for γ from (3.4C.4) and also put $\sum_k z_{1k} = \mathbf{1}^{\mathrm{T}}\mathbf{z}_1$. Combining

(3.4.30), (3.4C.7) and (3.4C.8) gives, for target *r*,

$$\mathbf{a}_{r\alpha}{}^{\mathrm{H}}\mathbf{P}_{\mathbf{A}}\Delta\mathbf{a}_r = -2\pi i\mathbf{e}_\alpha{}^{\mathrm{T}}\mathbf{R}\mathbf{b}\mathbf{b}^{\mathrm{H}}(n\mathbf{I}-\mathbf{1}\mathbf{1}^{\mathrm{T}})\mathbf{z}_r/\mu\,. \tag{3.4C.9}$$

(c) Alternative form for μ

We note that $|b_k| = |a_{1k}||a_{2k}| = \left(1/\sqrt{n}\right)^2 = 1/n$ and so

$$\|\mathbf{b}\|^2 = \sum_k |b_k|^2 = \sum_k 1/n^2 = 1/n = \mathbf{b}^{\mathrm{H}}\mathbf{b}\,,$$

so we can put μ, from (3.4C.6) in the form

$$\mu = 1-\mathbf{1}^{\mathrm{T}}\mathbf{b}\mathbf{b}^{\mathrm{H}}\mathbf{1} = n\mathbf{b}^{\mathrm{H}}\mathbf{b}-\mathbf{b}^{\mathrm{H}}\mathbf{1}\mathbf{1}^{\mathrm{T}}\mathbf{b} = \mathbf{b}^{\mathrm{H}}(n\mathbf{I}-\mathbf{1}\mathbf{1}^{\mathrm{T}})\mathbf{b} = \mathbf{b}^{\mathrm{H}}\mathbf{J}\mathbf{b} \tag{3.4C.10}$$

using (3.4.41).

Appendix 3.4D: Expectation of $\mathbf{uu}^{\mathrm{T}}$

From (3.4.42) we can put (dropping the suffix *r* in this section, for less cluttered notation) the expectation of $\mathbf{uu}^{\mathrm{T}}$ can be written in the form

$$\left\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \right\rangle = 4\pi^2\mathbf{E}^{\mathrm{T}}\mathbf{R}\left\langle \mathrm{Im}(\mathbf{K}\mathbf{z})\,\mathrm{Im}(\mathbf{K}\mathbf{z})^{\mathrm{T}} \right\rangle \mathbf{R}^{\mathrm{T}}\mathbf{E} = 4\pi^2\mathbf{E}^{\mathrm{T}}\mathbf{R}\mathbf{L}\mathbf{R}^{\mathrm{T}}\mathbf{E} \tag{3.4D.1}$$

where

$$\mathbf{K} = \mathbf{I}-\mathbf{b}\mathbf{b}^{\mathrm{H}}\mathbf{J}/\mu \tag{3.4D.2}$$

(with **J** given in (3.4.41)) and

$$\mathbf{L} = \left\langle \mathrm{Im}(\mathbf{K}\mathbf{z})\,\mathrm{Im}(\mathbf{K}\mathbf{z})^{\mathrm{T}} \right\rangle. \tag{3.4D.3}$$

Now let $\mathbf{K} = \mathbf{S} + i\mathbf{T}$ and $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ where **S**, **T**, **x** and **y** are real, then we have

$$\mathbf{L} = \left\langle (\mathbf{S}\mathbf{y}+\mathbf{T}\mathbf{x})(\mathbf{y}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}+\mathbf{x}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}) \right\rangle$$

$$= \mathbf{S}\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle \mathbf{S}^{\mathrm{T}} + \mathbf{S}\left\langle \mathbf{y}\mathbf{x}^{\mathrm{T}} \right\rangle \mathbf{T}^{\mathrm{T}} + \mathbf{T}\left\langle \mathbf{x}\mathbf{y}^{\mathrm{T}} \right\rangle \mathbf{S}^{\mathrm{T}} + \mathbf{T}\left\langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \right\rangle \mathbf{T}^{\mathrm{T}}. \tag{3.4D.4}$$

Now in Appendix 2D we found that $\left\langle \mathbf{y}\mathbf{y}^{\mathrm{T}} \right\rangle = \left(\sigma_\phi{}^2/n^2\right)\mathbf{I}$ where $\sigma_\phi{}^2$ is the phase error variance. Following the same argument we find that $\left\langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \right\rangle = \left(\sigma_a{}^2/n^2\right)\mathbf{I}$, where $\sigma_a{}^2$ is the amplitude error variance, and we also assume the independence of these errors, so that $\left\langle \mathbf{x}\mathbf{y}^{\mathrm{T}} \right\rangle = \mathbf{0}$. With these results we have, from (3.4D.4),

$$\mathbf{L} = (\sigma_\phi{}^2/n^2)\mathbf{S}\mathbf{S}^{\mathrm{T}} + (\sigma_a{}^2/n^2)\mathbf{T}\mathbf{T}^{\mathrm{T}}. \tag{3.4D.5}$$

In the case of equal variance errors we have

$$\mathbf{L} = (\sigma_\phi{}^2/n^2)(\mathbf{SS}^{\mathrm{T}} + \mathbf{TT}^{\mathrm{T}}) = (\sigma_\phi{}^2/n^2)\,\mathrm{Re}\left(\mathbf{KK}^{\mathrm{H}}\right). \tag{3.4D.6}$$

We note

$$\mathbf{KK}^{\mathrm{H}} = \mathbf{I} - \mathbf{bb}^{\mathrm{H}}\mathbf{J}/\mu - \mathbf{Jbb}^{\mathrm{H}}/\mu + \mathbf{bb}^{\mathrm{H}}\mathbf{JJbb}^{\mathrm{H}}/\mu^2$$

but using $\mathbf{J}^2 = n\mathbf{J}$ (from (3.4.41)) and $\mathbf{b}^{\mathrm{H}}\mathbf{Jb} = \mu$ from (3.4C.10) the last term becomes $n\mathbf{bb}^{\mathrm{H}}/\mu$ and substituting for $\mathbf{J}$ then gives

$$\mathbf{KK}^{\mathrm{H}} = \mathbf{I} - n\mathbf{bb}^{\mathrm{H}}/\mu + (\gamma^*\mathbf{b1}^{\mathrm{T}} + \gamma\mathbf{1b}^{\mathrm{H}})/\mu\,. \tag{3.4D.7}$$

Now $\mathbf{R1} = n\bar{\mathbf{r}}$ where $\bar{\mathbf{r}}$ is the centroid of the array – the mean position of the array elements, and if we take the array positions in $\mathbf{R}$ as relative to the centroid then we have $\bar{\mathbf{r}} = \mathbf{0}$ so that

$$\mathbf{R}\mathrm{Re}(\mathbf{KK}^{\mathrm{H}})\mathbf{R}^{\mathrm{T}} = \mathrm{Re}(\mathbf{RKK}^{\mathrm{H}}\mathbf{R}^{\mathrm{T}}) = \mathbf{R}(\mathbf{I} - n\,\mathrm{Re}(\mathbf{bb}^{\mathrm{H}})/\mu)\mathbf{R}^{\mathrm{T}} = n\tilde{\mathbf{M}}\,, \tag{3.4D.8}$$

using (3.4.34). From (3.4D.8), (3.4D.6) and (3.4D.1) we have, for the equal variance case,

$$\left\langle \mathbf{uu}^{\mathrm{T}} \right\rangle = (4\pi^2\sigma_\phi{}^2/n)\mathbf{E}^{\mathrm{T}}\tilde{\mathbf{M}}\mathbf{E} = (\sigma_\phi{}^2/n)\mathbf{U}\,. \tag{3.4D.9}$$

Appendix 3.4E: Approximate beamwidth of general array

The normalized gain in direction (α,ε) when steered in direction (α_0,ε_0) is given by

$$g(\alpha,\varepsilon;\alpha_0,\varepsilon_0) = \frac{1}{n}\sum\nolimits_k \exp\left(2\pi i\mathbf{r}_k{}^{\mathrm{T}}(\mathbf{e}(\alpha,\varepsilon) - \mathbf{e}(\alpha_0,\varepsilon_0))\right). \tag{3.4E.1}$$

Expanding the exponentials, putting $\Delta\mathbf{e} = \mathbf{e}(\alpha,\varepsilon) - \mathbf{e}(\alpha_0,\varepsilon_0)$, we have

$$g(\alpha,\varepsilon;\alpha_0,\varepsilon_0) = 1 + \frac{2\pi i}{n}\sum\nolimits_k \mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e} - \frac{(2\pi)^2}{2!n}\sum\nolimits_k\left(\mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e}\right)^2 - \frac{(2\pi)^3 i}{3!n}\sum\nolimits_k\left(\mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e}\right)^3 +$$

$$+\frac{(2\pi)^4}{4!n}\sum\nolimits_k\left(\mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e}\right)^4 + \ \ldots$$

and if we take the array centroid as the origin for the element coordinates then the second term disappears, as $\sum_k \mathbf{r}_k = \mathbf{0}$. Also, although this does not make the fourth term zero, except for arrays with a degree of symmetry, it will be the case that this term will generally be low compared with the even power terms. Thus we have

$$g(\alpha,\varepsilon;\alpha_0,\varepsilon_0) \approx 1-\frac{(2\pi)^2}{2!n}\sum_k\left(\mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e}\right)^2+\frac{(2\pi)^4}{4!n}\sum_k\left(\mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e}\right)^4+\ldots \qquad (3.4E.2)$$

The 3dB gain is given by the value of $\Delta\mathbf{e}$ such that $g = 1/\sqrt{2}$. If we ignore the last term we have

$$\frac{(2\pi)^2}{2!n}\sum_k\left(\mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e}\right)^2 \approx 1-\frac{1}{\sqrt{2}} = 0.293\,. \qquad (3.4E.3)$$

If we put

$$g \approx 1-T_1+T_2 = 1-\frac{(2\pi)^2}{2!}S_2+\frac{(2\pi)^4}{4!}S_4 \qquad (3.4E.4)$$

where $S_2 = \sum_k\left(\mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e}\right)^2\Big/n$ and $S_4 = \sum_k\left(\mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e}\right)^4\Big/n$ then we can compare the terms T_1 and T_2 taking two extreme cases. In the first case we suppose $\mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e} = \rho \;\; \forall k$, the same value for all the elements. In this case we see that $S_2 = \rho^2$ and $S_4 = \rho^4$, so that $T_1 = \dfrac{(2\pi\rho)^2}{2!}$ and $T_2 = \dfrac{(2\pi\rho)^4}{4!}$ and so

$T_2/T_1 = (2\pi\rho)^2\,2!/4!$. Also, from (3.4E.3) see that $T_1 = (2\pi\rho)^2/2! \approx 0.293$ so that

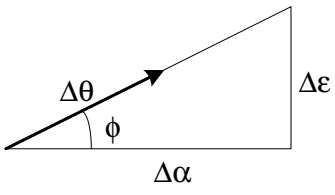
$$T_2/T_1 \approx 0.293(2!)^2/4! = 0.293/6 < 0.05\,.$$

In the other case we suppose that only one component is significant so that $\mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e} = \rho$ for only one value of k. In this case $S_2 = \rho^2/n$ and $S_4 = \rho^4/n$, and following through the argument we find

$$T_2/T_1 \approx 0.293n/6\,.$$

If $n = 6$ (as in the array R1) then $T_2/T_1 \approx 0.293n/6 < 0.3$. Thus we see that the fourth power term is between 3/10 (when $n = 6$) and 1/20 of the magnitude of the second power term, with, in practice, being rather closer to the lower value, in general. Thus it is reasonable to neglect the fourth power term to obtain a fairly good approximation to the beamwidth.

Now we want to find the half-beamwidth $\Delta\theta$ of a beam at the position of target 1 along the line between target 1 and target 2. If target 2 is at position $(\Delta\alpha,\Delta\varepsilon)$ relative to target 1 then this line is at an angle ϕ, given by $\tan\phi = \Delta\varepsilon/\Delta\alpha$, as illustrated. We can put



$$\Delta\mathbf{e} \approx \mathbf{e}_\alpha\Delta\alpha + \mathbf{e}_\varepsilon\Delta\varepsilon = (\mathbf{e}_\alpha\cos\phi + \mathbf{e}_\varepsilon\sin\phi)\Delta\theta = \mathbf{e}_\phi\Delta\theta \qquad (3.4E.5)$$

where $\mathbf{e}_\alpha$ and $\mathbf{e}_\varepsilon$ are the partial derivatives of $\mathbf{e}(\alpha,\varepsilon)$ with respect to α and ε respectively. The estimate of $\Delta\theta$ is then given, from (3.4E.3), by

$$\frac{(2\pi)^2}{2!n}\sum\nolimits_k \left(\mathbf{r}_k{}^{\mathrm{T}}\mathbf{e}_\phi\Delta\theta\right)^2 = 0.293$$

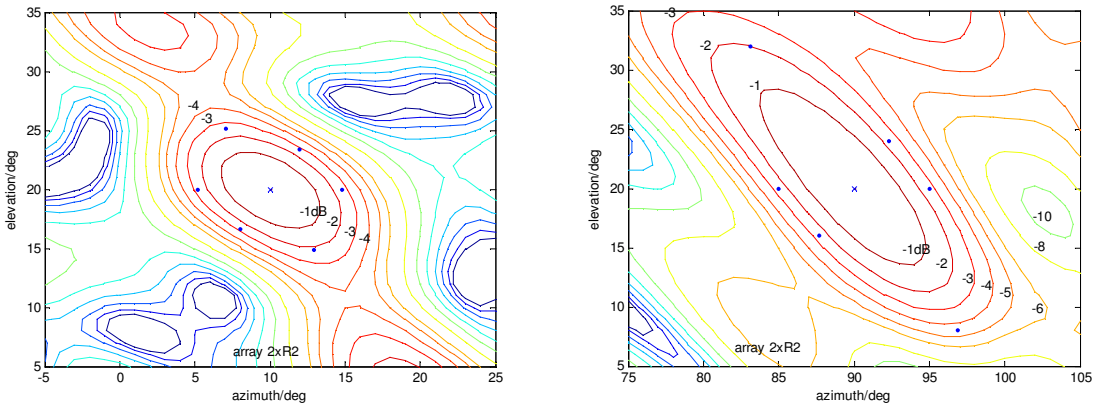
Using the result obtained earlier (see (3.3A.6) for example),

$$\sum\nolimits_k \left(\mathbf{r}_k{}^{\mathrm{T}}\mathbf{e}_\phi\right)^2 = \mathbf{e}_\phi{}^{\mathrm{T}}\mathbf{R}\mathbf{R}^{\mathrm{T}}\mathbf{e}_\phi = n\mathbf{e}_\phi{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_\phi\,,$$

we have

$$\Delta\theta^2 = \frac{0.293}{2\pi^2\mathbf{e}_\phi{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_\phi} = \frac{0.015}{\mathbf{e}_\phi{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_\phi} \qquad (3.4E.6)$$

with $\mathbf{M} = \mathbf{R}\mathbf{R}^{\mathrm{T}}/n$ and $\mathbf{e}_\phi$ given in (3.4E.5).



(a) beam centred at (10,20) (b) beam centred at (90,20)

Figure 3.22 Beam patterns for an irregular 3D array

As $\mathbf{e}(\alpha,\varepsilon)^{\mathrm{T}} = [\cos\alpha\cos\varepsilon \quad \sin\alpha\cos\varepsilon \quad \sin\varepsilon]$

we see that $\mathbf{e}_\alpha(\alpha,\varepsilon)^{\mathrm{T}} = [-\sin\alpha\cos\varepsilon \quad \cos\alpha\cos\varepsilon \quad 0]$

and $\mathbf{e}_\varepsilon(\alpha,\varepsilon)^{\mathrm{T}} = [-\cos\alpha\sin\varepsilon \quad -\sin\alpha\sin\varepsilon \quad \cos\varepsilon]$.

Two beam patterns are shown in Figure 3.22 for the array R2 (defined in §3.3.6) doubled in size to reduce the beam pattern. The array R2 was formed by taking the irregular planar array R1 and varying the element heights to form a 3D array, for which there is no simple expression for the beamwidth. In Figure 3.22(a) the beam was

pointed at 10° in azimuth and 20° in elevation. The contour levels, relative to the peak of the beam, are at –1, -2, . . . , -6, -8, -10, . . . ,-20. The blue dots show the half-beamwidth positions – i.e. the estimated 3dB points – calculated using (3.4E.6) at ϕ values 0°, 60°, . . . , 300°, and we see that these are reasonably accurate, though generally lying slightly within the 3dB contour, and in one case quite near the 2dB value. Moving the beam to azimuth 90° (in (b)) we see that there is one point at –2dB, four between –2 and –3dB, and one between –4 and –5dB. However, as an approximation, it is generally quite good and relatively simple.

The half-beamwidth in the azimuth or elevation plane is simply given by setting ϕ to 0º or 90º and $\mathbf{e}_\phi$ to $\mathbf{e}_\alpha$ or $\mathbf{e}_\varepsilon$.

3.5. SUMMARY OF CHAPTER 3

3.5.1 General approach

In this chapter we have limited the error study to the case of large signals only, ignoring the effect of noise in further limiting the parameter estimation accuracy. The results, therefore, show the best that can be achieved with given error levels, and hence to what levels the errors (of the stored manifold PSVs relative to the actual vectors seen by the signals) should be reduced in order to achieve (at best) some required accuracy of parameter estimation. The effect of noise on accuracy, at least in the single target case is given by the Cramer-Rao bound.

We have developed the theory for the accuracy of MUSIC in stages from the simplest case – a uniform linear array, so estimating a single angle parameter only, and for a single target – up to the case of multiple targets using general (planar or volume) arrays, with two angle parameters. In principle only the most complicated case might have been presented, from which the results for simpler cases could have been derived, but the advantages of this progressive presentation are that it should be easier for the reader to follow the theory as the complexity increases in stages, and that any user interested in a simpler case can find the result without having to understand and simplify a higher level result.

The single target accuracy is of most interest as it is also close to the accuracy of measurement in multiple target cases when the targets are well separated. The next case of interest is that of two close targets, and these are the cases studied. The single target case is the same for both MUSIC and IMP, which have essentially the same function in

this case. For the two target case theoretical results were found only for MUSIC; the approach used very effectively for MUSIC becomes too difficult when applied to the more complex IMP function, so IMP results have been obtained in simulation only.

Also the study has been applied almost solely to the case of equal, parallel pattern (EPP) arrays. These form a very substantial proportion of practical arrays. The case of non-EPP arrays is of less easy practical application because the performance will depend on the actual element patterns and a considerable amount of information on the element patterns and orientations will be needed in order to use the theoretical results. However, this case was tackled, in the single target case, and theoretical expressions obtained, confirmed, as for the EPP theory, by simulations.

3.5.2 Results achieved

(a) As expected, the angular accuracy of an array with errors depends directly on the array aperture – that is to say that the s.d. error is inversely proportional to the linear extent of the array normal to the signal direction. This is not particularly original or interesting, but more so is the fact that the theory shows that **the error variance** (or squared s.d. error) **is inversely proportional to the second moment of the array positions** about their mean.

(b) The result **for the single target case** is equivalent to the Cramer-Rao Bound (CRB) with an error variance replacing the signal to noise ratio figure – i.e. **the array errors are equivalent to a specific level of S/N** as given in Figure 3.7.

(c) There is **excellent agreement between theory and simulation** for the arrays chosen. (See Figures 3.2, 3.3, 3.5, 3.8, 3.9, 3.11, 3.12. . . , Tables 3.2 – 3.7)

(d) The results are, in general, for arrays of equal parallel pattern elements (an EPP array) – in practice the most common form – but **the case of non-parallel pattern elements, with a single target**, was taken and results obtained, again confirmed in simulation. (In this case the centroid is of the element positions weighted by the element power gains in the signal direction. Thus the centroid varies with signal direction, in general.)

(e) An expression for **the error sensitivity** on a single target for an EPP array (the s.d. error with *unit* variance phase error) was defined (§3.3.6) and could be used for any specified array, to indicate the performance to be expected using the array (for azimuth and elevation estimation). (Figures 3.13 and 3.14 show the advantage of increasing the elevation aperture for a non-planar array relative to the planar form.

(f) From the error sensitivity study two **rules of thumb**, giving the approximate general performance in terms of sensitivity, were obtained for an array with a single target were devised: (1) for regular circular planar EPP arrays (again a quite likely form of array) with half wavelength spacing in terms of the number of elements and (2) for more general arrays in terms of the aperture (in wavelengths). These are intended to give an rough, but easily obtained, estimate of the performance to be expected from an array.

(g) The expression for the s.d. errors in the **two target case** was obtained. This shows there is interaction between the signals, when relatively close (e.g. within a beamwidth) reducing the performance. The **results were confirmed** to well below a beamwidth (in fact down **to about ¼ beamwidth** separation) when some of the approximations made start to fail.

(h) In all **the single target cases** the theory showed that **only the phase errors** (or imaginary parts) of the errors in the components of the PSVs **contributed to parameter estimation errors** (within the reasonable approximations used). This was confirmed in simulation, where, with only amplitude errors included, the parameter errors were negligible . . .

(k) . . . but **with two targets amplitude errors also affect the performance**. This was also confirmed in simulation, with excellent agreement.

(l) Not directly related to the error study, but using some of the theory developed, **a useful approximate expression for the beamwidth of an irregular array** in any given direction was obtained.

(m) Simulation results indicated that the **accuracy of IMP was very similar to that of MUSIC**, for which the theory has been confirmed. This is a reasonable result on the basis that in the absence of errors (limited by noise alone) the accuracy of both methods approaches the CRB, while with errors we have, for MUSIC, an expression *of the form* of the CRB, so we see that errors have a noise-like effect on MUSIC (statistically). On this basis it is not surprising that the accuracy of IMP with errors is similar.

(n) If phase errors are dominant (relative to amplitude errors) then the performance with two signals present is found to be very close to the performance to the two single signal cases. This is useful as the single signal case is more easily obtained and the rules of thumb are applicable.

Chapter 4: Resolution of IMP

4.1. RESOLUTION BASED ON THE METHOD OF SPEIRS *ET AL*

4.1.1 Negligible errors

An approach to determining the potential resolution performance of IMP, and similar methods, is proposed in SMCR99 [41]. This is based on the detectability of a second target when the first target has been removed by setting a null on it. The principle is that if the second target is above the threshold level for detection with a null directed onto the first target (assumed to be the larger) then the second target will be detected and an estimate of its position will be found, and in this sense the two targets will have been resolved. This has the weakness, for practical systems, that it assumes the first signal's position has been found accurately, and the null has been placed correctly at this position, so one implicit assumption is that there are no system errors, that the signals are large enough for the effect of noise on the position estimates to be negligible and that the signals are uncorrelated (and the observation interval is long enough to use this fact). Given these assumptions, this is a promising basis for an estimate of the resolution in good (or ideal) conditions, and so will represent a bound, which might be approached closely in practice in some cases. Speirs *et al* also consider that the null, rather than being a perfect one (given by orthogonal projection, as in IMP, using vectors $\mathbf{Q}_{a0}\mathbf{a}(\boldsymbol{\theta})$ looking in direction $\boldsymbol{\theta}$ with a null in direction $\boldsymbol{\theta}_0$ with the PSV $\mathbf{a}_0$) could be chosen to be signal strength dependent, as given by the Wiener adaptive array solution (using vectors $\mathbf{R}^{-1}\mathbf{a}(\boldsymbol{\theta})$, where $\mathbf{R}$ is the system covariance matrix). In this latter case, as the null is not quite so deep it follows that the attenuation of the second signal will not be so great and the resolution will be better. However this effect is likely to be small in general, and as it does not correspond to the IMP-type approach, we will not consider this variation and consider only the projection null. Also, although the paper considers an example of the three signal case, this case is difficult to cover in general terms as there are too many variables (the three target strengths and the three separations of the targets from each other) so we restrict attention to the main problem, the resolution of two close targets. In the Speirs paper plots are given of the power loss due to a nearby null but this loss and the limiting separation for resolution are not given in terms of the array parameters. The aim of this section is to obtain an expression for this loss, leading to expressions for the resolution limit and the resolution improvement factor.

With the assumption of a null accurately placed at the position of target 1, the signal to

noise power gain at target 2 is proportional to

$$g_{21} = \mathbf{a}_2^{\mathrm{H}}\mathbf{Q}_1\mathbf{a}_2 = \mathbf{a}_2^{\mathrm{H}}\left(\mathbf{I} - \mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}/\|\mathbf{a}_1\|^2\right)\mathbf{a}_2 = \|\mathbf{a}_2\|^2 - \left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_2\right|^2/\|\mathbf{a}_1\|^2 \quad . \qquad (4.1.1)$$

[If a weight vector $\mathbf{w}$ is applied to an array then the signal to noise power ratio for a signal in some direction with PSV $\mathbf{a}_2$ is proportional to $|\mathbf{w}^{\mathrm{T}}\mathbf{a}_2|^2/\|\mathbf{w}\|^2$. (This is the 'white noise' case where the noise is uniform across the receiver channels. We make this common assumption here.) If nulls are set in directions with PSVs contained in $\mathbf{A}$ the weight vector maximizing the signal to noise ratio is given, (see Appendix 2B, eq. (2.B.12)) within a constant, by $\mathbf{w}^* = \mathbf{Q}_\mathbf{A}\mathbf{a}_2$, or equivalently $\mathbf{w}^{\mathrm{T}} = \mathbf{a}_2^{\mathrm{H}}\mathbf{Q}_\mathbf{A}$ (as $\mathbf{Q}_\mathbf{A} = \mathbf{Q}_\mathbf{A}^{\mathrm{H}}$), and then the ratio in (2.B.2) becomes $r_0|\mathbf{a}_2^{\mathrm{H}}\mathbf{Q}_\mathbf{A}\mathbf{a}_2|^2/\mathbf{a}_2^{\mathrm{H}}\mathbf{Q}_\mathbf{A}\mathbf{a}_2 = r_0\mathbf{a}_2^{\mathrm{H}}\mathbf{Q}_\mathbf{A}\mathbf{a}_2$. (We have used $\|\mathbf{Q}_\mathbf{A}\mathbf{a}_2\|^2 = \mathbf{a}_2^{\mathrm{H}}\mathbf{Q}_\mathbf{A}^{\mathrm{H}}\mathbf{Q}_\mathbf{A}\mathbf{a}_2$ and $\mathbf{Q}_\mathbf{A}^{\mathrm{H}}\mathbf{Q}_\mathbf{A} = \mathbf{Q}_\mathbf{A}^2 = \mathbf{Q}_\mathbf{A}$ for a projection matrix, and also the fact that $\mathbf{a}_0^{\mathrm{H}}\mathbf{Q}_\mathbf{A}\mathbf{a}_0$ is real as $\mathbf{Q}_\mathbf{A}$ is Hermitian.) In the case considered here we have a single null, so we can put $\mathbf{A} = \mathbf{a}_1$, and so $\mathbf{Q}_1 = \mathbf{Q}_{\mathbf{a}_1} = \mathbf{I} - \mathbf{a}_1\mathbf{a}_1^{\mathrm{H}}/\|\mathbf{a}_1\|^2$.]

Without the null the power gain ratio is $g_{20} = \|\mathbf{a}_2\|^2$ (on putting $\mathbf{I}$ for $\mathbf{Q}_1$ in (4.1.1)) so the power loss is

$$l_{21} = g_{21}/g_{20} = 1 - \left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_2\right|^2/\|\mathbf{a}_1\|^2\|\mathbf{a}_2\|^2 \; .$$

We see that this loss expression would be the same if we multiplied $\mathbf{a}_1$ or $\mathbf{a}_2$ by an arbitrary (non-zero) constant so we could use the normalized vectors here. Taking the PSVs to be normalized, we can write

$$l_{21} = 1 - \left|\mathbf{a}_1^{\mathrm{H}}\mathbf{a}_2\right|^2 = 1 - |\gamma|^2 = \mu \qquad (4.1.2)$$

where γ and μ are as in Chapter 3 (eq. (3.4.29)). The gain loss in the direction of signal 1 with a null at the position of signal 2 is given by

$$l_{12} = \frac{g_{12}}{g_{10}} = \frac{\mathbf{a}_1^{\mathrm{H}}\mathbf{Q}_2\mathbf{a}_1}{\|\mathbf{a}_1\|^2} = \frac{\mathbf{a}_1^{\mathrm{H}}\left(\mathbf{I} - \mathbf{a}_2\mathbf{a}_2^{\mathrm{H}}/\|\mathbf{a}_2\|^2\right)\mathbf{a}_1}{\|\mathbf{a}_1\|^2} = 1 - \frac{\left|\mathbf{a}_2^{\mathrm{H}}\mathbf{a}_1\right|^2}{\|\mathbf{a}_1\|^2\|\mathbf{a}_2\|^2}$$

and we see that $l_{12} = l_{21}$, i.e. the power gains to the two signals suffer the same loss due to a null placed on the other one.

To express μ in terms of the array position coordinates we use the approach of Appendix 3.4E of Chapter 3 (taking the common case of an EPP array). The gain γ is given (following (3.4E.1)) by

$$\gamma = \mathbf{a}_1{}^{\mathrm{H}}\mathbf{a}_2 = \frac{1}{n}\sum\nolimits_k \exp\left(2\pi i \mathbf{r}_\mathbf{k}{}^{\mathrm{T}}(\mathbf{e}(\alpha_2,\varepsilon_2) - \mathbf{e}(\alpha_1,\varepsilon_1))\right) \qquad (4.1.3)$$

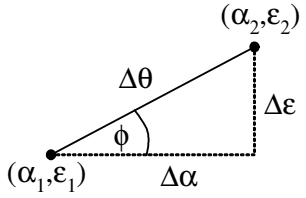
Expanding as before with $\Delta\mathbf{e} = \mathbf{e}_2 - \mathbf{e}_1$ (the unit direction vectors in the directions of signals 1 and 2, i.e. $\mathbf{e}_k = \mathbf{e}(\alpha_k,\varepsilon_k)$) then putting $\sum_k \mathbf{r}_k = \mathbf{0}$ (using centroid coordinates) and taking the second order approximation, as in Chapter 3, we have

$$\gamma \approx 1 + \frac{2\pi i}{n}\sum\nolimits_k \mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e} - \frac{(2\pi)^2}{2!n}\sum\nolimits_k \left(\mathbf{r}_k{}^{\mathrm{T}}\Delta\mathbf{e}\right)^2 = 1 - 2\pi^2\mathbf{e}_\phi{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_\phi\Delta\theta^2 \quad (4.1.4)$$

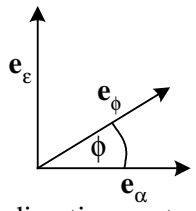
using $\Delta\mathbf{e} \approx \mathbf{e}_\phi\Delta\theta$ and $\sum_k \left(\mathbf{r}_k{}^{\mathrm{T}}\mathbf{e}_\phi\right)^2 = \mathbf{e}_\phi{}^{\mathrm{T}}\mathbf{R}\mathbf{R}^{\mathrm{T}}\mathbf{e}_\phi = n\mathbf{e}_\phi{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_\phi$ (from (3.3A.6)). Finally, again taking the lowest order approximation, we have

$$\mu \approx 4\pi^2\mathbf{e}_\phi{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_\phi\Delta\theta^2 \qquad (4.1.5)$$

as the loss in power gain. Here $\Delta\theta^2 = \Delta\alpha^2 + \Delta\varepsilon^2 = (\alpha_2 - \alpha_1)^2 + (\varepsilon_2 - \varepsilon_1)^2$ and $\mathbf{e}_\phi = \mathbf{e}_\alpha\cos\phi + \mathbf{e}_\varepsilon\sin\phi$, (as illustrated below) where $\mathbf{e}_\alpha$ and $\mathbf{e}_\varepsilon$ are the derivatives with respect to α and ε of the direction vector, evaluated in the immediate region of the signals – for example at the position of signal 1, or of signal 2, or at the midpoint between them, and $\tan\phi = \Delta\varepsilon/\Delta\alpha$.



target positions



direction vector derivatives

Now let the power level of signal 2 be a factor ρ above the minimum level for detection, then the targets are at the limit of detection (and equally, in the case of IMP, at the limit of resolution) when the separation is such that signal 2 is reduced to the minimum detection level, i.e. the loss is equal to $1/\rho$. The limiting separation for resolution $\Delta\theta_{\mathrm{res}}$ (in radians) is thus given by $\mu = 1/\rho$ or

$$\Delta\theta_{\mathrm{res}}{}^2 = 1\Big/(4\pi^2\mathbf{e}_\phi{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_\phi\,\rho)\,. \qquad (4.1.6)$$

The resolution limits in azimuth and elevation, rather than the general direction ϕ (in the α-ε parameter space) are given by

$$\Delta\alpha_{\mathrm{res}} = 1\Big/2\pi\sqrt{\mathbf{e}_\alpha{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_\alpha\rho} \text{ and } \Delta\varepsilon_{\mathrm{res}} = 1\Big/2\pi\sqrt{\mathbf{e}_\varepsilon{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_\varepsilon\rho}\,. \qquad (4.1.7)$$

This gives the resolution in terms of the array element positions (in the moment matrix **M**) and the derivatives of the unit direction vector in the region of the targets being

resolved, $\mathbf{e}_\alpha$ and $\mathbf{e}_\varepsilon$, as well as the weaker signal power relative to the minimum detectable power ρ.

We can also define a figure of merit for resolution, which is the ratio of the array beamwidth (approximately the resolution limit in the simple, beamforming case) to the minimum separation for resolution in the superresolution case. This ratio is essentially the resolution improvement factor given by the superresolution processing, and the higher the figure the better the resolution. For irregular arrays the beamwidth is not simply defined, but we can use the expression obtained in Appendix 3.4E (equation (3.4E.6)). This expression gives the half-beamwidth (along direction ϕ in parameter space), so if θ_B is the beamwidth, this equation becomes

$$\left(\frac{\theta_B}{2}\right)^2 = \frac{0.293}{2\pi^2 \mathbf{e}_\phi{}^T \mathbf{M} \mathbf{e}_\phi} \tag{4.1.8}$$

and combining this with (4.1.5) we have

$$\mu = 8 \times 0.293 (\Delta\theta / \theta_B)^2 = 2.34 (\Delta\theta / \theta_B)^2 = 2.34 / r^2 \tag{4.1.9}$$

Combining these expressions (with $\mu = 1/\rho$ at the resolution limit) we have

$$r = \frac{\theta_B}{\Delta\theta_{res}} = 1.53\sqrt{\rho} \tag{4.1.10}$$

where r is the resolution improvement factor. This is a remarkably simple result, not requiring any details of the array (as both the detection limit and the beamwidth depend on the array moment $\mathbf{M}$ in the same way). The improvement seems to increase indefinitely (as $\sqrt{\rho}$) with the power level of the weaker signal, but in practice, of course, errors will limit this factor. If the weaker signal is 10dB above the minimum detectable level this gives a resolution improvement factor of 4.8, i.e. the signals should be at the limit of resolution at about 0.2 beamwidths separation, and at 20dB the value of r will be about 15.3 indicating resolution at 0.065 beamwidths, which seems optimistic.

4.1.2 Resolution with errors

Errors will limit the cancellation of signal 1 (taken to be the stronger of the two) and so, if it is strong enough there will be a residue above noise level. If the detection level is left at a value defined by noise level (i.e. to give a suitably low false alarm rate on noise), which is appropriate for weaker signals, this residue of the strong signal will generate false detections at essentially 100% probability when no second signal is

present. (If there is a second signal present then false alarms will be generated after the second scan, and so on.) Thus it is essential to raise the threshold in this case, to a certain factor below the peak of the function on the first scan. The factor will need to be based on the estimated levels of the errors (e.g. 4º rms in phase or the equivalent 0.6dB rms in amplitude), or perhaps could be found experimentally from observed residues with given input signals.

In order to estimate the resolution performance with errors we need first to determine the residue level. The gain to signal 1 with an error corrupted null on this signal is

$$g_{11e} = \mathbf{a}_1^{\mathrm{H}} \mathbf{Q}_{\tilde{\mathbf{a}}_1} \mathbf{a}_1 = 1 - \frac{\left|\mathbf{a}_1^{\mathrm{H}} \tilde{\mathbf{a}}_1\right|^2}{\left\|\tilde{\mathbf{a}}_1\right\|^2} = \frac{\left\|\tilde{\mathbf{a}}_1\right\|^2 - \left|\mathbf{a}_1^{\mathrm{H}} \tilde{\mathbf{a}}_1\right|^2}{\left\|\tilde{\mathbf{a}}_1\right\|^2} \qquad (4.1.11)$$

where $\tilde{\mathbf{a}}_1 = \mathbf{a}_1 + \delta\mathbf{a}_1$ is the error vector (not normalized). Noting that

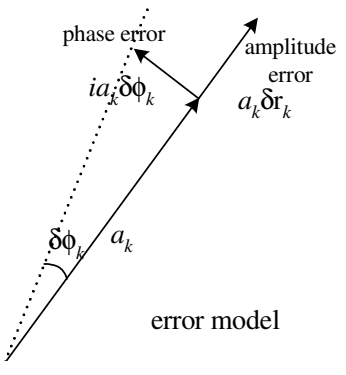
$$\left\|\tilde{\mathbf{a}}_1\right\|^2 = 1 + 2\,\mathrm{Re}\left(\mathbf{a}_1^{\mathrm{H}} \delta\mathbf{a}_1\right) + \left\|\delta\mathbf{a}_1\right\|^2 \text{ and } \left|\mathbf{a}_1^{\mathrm{H}} \tilde{\mathbf{a}}_1\right|^2 = \left|1 + \mathbf{a}_1^{\mathrm{H}} \delta\mathbf{a}_1\right|^2 = 1 + 2\,\mathrm{Re}\left(\mathbf{a}_1^{\mathrm{H}} \delta\mathbf{a}_1\right) + \left|\mathbf{a}_1^{\mathrm{H}} \delta\mathbf{a}_1\right|^2$$

(where we have taken $\left\|\mathbf{a}_1\right\|^2 = 1$) equation (4.1.11) becomes

$$g_{11e} = \frac{1 + 2\,\mathrm{Re}\left(\mathbf{a}_1^{\mathrm{H}} \delta\mathbf{a}_1\right) + \left\|\delta\mathbf{a}_1\right\|^2 - \left(1 + 2\,\mathrm{Re}\left(\mathbf{a}_1^{\mathrm{H}} \delta\mathbf{a}_1\right) + \left|\mathbf{a}_1^{\mathrm{H}} \delta\mathbf{a}_1\right|^2\right)}{\left\|\tilde{\mathbf{a}}_1\right\|^2} = \frac{\left\|\delta\mathbf{a}_1\right\|^2 - \left|\mathbf{a}_1^{\mathrm{H}} \delta\mathbf{a}_1\right|^2}{\left\|\tilde{\mathbf{a}}_1\right\|^2}$$

or $$g_{11e} = \left\|\delta\mathbf{a}_1\right\|^2 - \left|\mathbf{a}_1^{\mathrm{H}} \delta\mathbf{a}_1\right|^2$$

to the lowest order approximation in $\delta\mathbf{a}_1$. We note that $g_{11e} \geq 0$ (by the Schwarz inequality, with equality only if $\delta\mathbf{a}_1 = k\mathbf{a}_1$ for some constant k) and has maximum value $\|\delta\mathbf{a}_1\|^2$. With the error in channel k given by $a_k(\delta r_k + i\delta\phi_k)$ where δr_k is the relative amplitude error and $\delta\phi_k$ the phase error, as shown, then we have



$$\mathbf{a}_1^{\mathrm{H}} \delta\mathbf{a}_1 = \sum_k a_{1k}^* a_{1k} (\delta r_k + i\delta\phi_k) = \sum_k (\delta r_k + i\delta\phi_k)/n .$$

We see that we can put $\mathbf{a}_1^{\mathrm{H}} \delta\mathbf{a}_1 = \overline{\delta r} + i\overline{\delta\phi}$ where $\overline{\delta r}$ and $\overline{\delta\phi}$ are the mean relative amplitude and phase errors over the n array channels. As we take the errors to be from zero mean distributions we expect these

mean values to be low, tending to zero as n becomes large, and so we expect $\left|\mathbf{a}_1^{\mathrm{H}}\delta\mathbf{a}_1\right|^2$ to be small compared with $\left\|\delta\mathbf{a}_1\right\|^2$, and we take $g_{11e} \approx \left\|\delta\mathbf{a}_1\right\|^2$. Thus we have

$$g_{11e} \approx \left\|\delta\mathbf{a}_1\right\|^2 = \sum\nolimits_k \left|a_k(\delta r_k + i\delta\phi_k)\right|^2 = \sum\nolimits_k \left|\delta r_k + i\delta\phi_k\right|^2 \Big/ n \approx \overline{\delta r^2} + \overline{\delta\phi^2}$$

$$= \psi_a + \psi_\phi = \psi_e \tag{4.1.12}$$

where $\psi_a = \overline{\delta r^2}$ and $\psi_\phi = \overline{\delta\phi^2}$ are the variances of the amplitude and phase errors and ψ_e is the total error variance. We have neglected the cross terms as these will be small, assuming the amplitude and phase errors are uncorrelated.

On the first scan (with no nulls inserted) the peak level will be essentially $p_1 + p_2 + p_n$ where p_k is the power received when the beam is pointed at signal k and p_n is the noise power level. (We assume that the targets to be resolved are close together in terms of the beamwidth, in which case the gains will be almost the same to both targets). We do not know, at this stage, whether there is only one target or more close targets giving this peak; in any case, if a null is directed, during the second scan, towards the position of the peak of the first scan, then, because of errors, a gain of g_{11e} (equal to the total error variance ψ_e, as shown in (4.1.12) above) will be applied, rather than a perfect null. Thus we must set a threshold related to the residue $\psi_e(p_1 + p_2)$ plus the noise contribution p_n, rather than related to the noise level alone. A suitable threshold, in the case of looking for a target in the presence of noise alone, is given in Appendix 4A (eq. (4.A.6)). With a signal residue present we need to raise this threshold. If we treat the residue as a noise-like signal then we replace ψ_n in this expression for the threshold by this power level. We will detect a second signal if the residue of signal 1, with the error limited null, plus signal 2 with a loss due to the proximity of the null directed on signal 1 (given by μ), together exceed the new threshold. Thus we require

$$\psi_e p_1 + \mu p_2 + p_n \geq (\psi_e(p_1 + p_2) + p_n)(1 + u_\beta)$$

or

$$\mu p_2 \geq \psi_e p_2(1 + u_\beta) + (\psi_e p_1 + p_n)u_\beta\,. \tag{4.1.13}$$

(u_β defines the threshold on noise, see Appendix 4A), taking into account the integration factor p and the required false alarm rate, β.) The limit of resolution is when this is an equality, and using (4.1.9) for μ and rearranging we find

$$r = 1.53 \Big/ \sqrt{\psi_e (u_\beta \, p_1 / p_2 + 1 + u_\beta) + u_\beta p_n / p_2} \; . \tag{4.1.14}$$

If $\psi_e p_1$, the residue of signal 1 after imperfect nulling, dominates p_n then we have, more simply,

$$r = 1.53 \Big/ \sqrt{\psi_e (u_\beta \, p_1 / p_2 + 1 + u_\beta)} \, . \tag{4.1.15}$$

We see that the smaller the error variance the larger r and the larger the signal power ratio p_1/p_2 the smaller r, both expected results. We also note that the performance is now limited by the error levels and the power ratio of the signals, not on the actual signal to noise ratios, as long as they are high enough to require a redefinition of the threshold (based on the expected residue.)

We note that without errors ($\psi_e = 0$) the inequality (4.1.13) becomes $\mu p_2 \geq u_\beta p_n$, leading to

$$r = 1.53 \Big/ \sqrt{u_\beta p_n / p_2} = 1.53 \sqrt{\rho} \tag{4.1.16}$$

where $u_\beta p_n$ is the minimum detectable power level and $\rho = p_2 \big/ u_\beta p_n$ is the ratio of signal 2 power to this minimum, as in §4.1.1 above (eq. (4.1.10)). We can now put (4.1.14) in the form

$$r = 1.53 \Big/ \sqrt{\psi_e (u_\beta \, p_1 / p_2 + 1 + u_\beta) + 1/\rho} \tag{4.1.17}$$

which reduces to (4.1.10) if we put $\psi_e = 0$. To put in some example values, let us set β at 3.2×10^{-5} (giving a false alarm rate of 3.2% over a scan of 1000 points at beamwidth intervals), as in Appendix 4A, then the normal distribution cumulative probability threshold t_β is 4 standard deviations. Let the number of samples per channel p be 100 then we have $u_\beta = t_\beta \sqrt{(2/p)} = 0.57$. Also let the phase errors be 5º rms so that ψ_ϕ = 0.0076 and the amplitude errors be the same level (corresponding to about 0.73dB rms) then $\psi_e = 0.0152$ (giving a null depth of 18.2dB). Let there be 10dB difference between the signals, then we find, from (4.1.17) (and ignoring $1/\rho$ for strong signals), that

$$r = 4.6.$$

This suggests that the signals can be resolved, given these parameter values, down to about 0.22 beamwidths. We note that the signals are in fact quite large, as we assume the first peak is over 18dB above noise level, so that the residue is high enough to use a residue-based threshold, and furthermore there is an integration factor of 100 which

effectively increases the signal to noise ratio by this factor. If we took the smaller signal (before integration) to be 20dB above the detection level, then the threshold detection condition in the negligible error case is given, from (4.1.10) or (4.1.16), by $r = 15.3$, so we see that errors can limit the resolution considerably.

4.1.3 Modified method based on null depth (error free case)

We can modify the method of Speirs *et al* [41] by following the IMP method more closely. (We take the error free case here and uncorrelated signals, as before.) We carry out the first scan to find the peak position, then place a null at this point (rather than at the position of the larger signal) for the second scan. In the region of the signals we scan a unit gain beam except for the null, so we find the residues of the two signals resulting from this null and compare this with the threshold. In Appendix 4B we find that the peak position on the first scan is given by

$$\boldsymbol{\theta}_{\text{pk}} = \frac{p_1}{p_1 + p_2}\boldsymbol{\theta}_1 + \frac{p_2}{p_1 + p_2}\boldsymbol{\theta}_2 \tag{4.1.18}$$

where $\boldsymbol{\theta}_k = \begin{bmatrix} \alpha_k \\ \varepsilon_k \end{bmatrix}$, in general, and we see that $\boldsymbol{\theta}_{\text{pk}}$ is between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ and weighted as the signal strengths, as might be expected. We have, from (4.1.2) and (4.1.5), an expression for the gain loss due to a null at a point at relative position $(\Delta\alpha,\Delta\varepsilon)$, so we need these angle differences for the two signal positions relative to the position of the null.

We have, using the result (4.1.18),

$$\Delta\alpha_1 = \alpha_1 - \alpha_{\text{pk}} = \alpha_1 - \left(\frac{p_1\alpha_1}{p_1 + p_2} + \frac{p_2\alpha_2}{p_1 + p_2} \right) = \frac{p_2(\alpha_1 - \alpha_2)}{p_1 + p_2}$$

$$= -\frac{p_2\Delta\alpha}{p_1 + p_2} \tag{4.1.19a}$$

where $\Delta\alpha = \alpha_2 - \alpha_1$. Similarly

$$\Delta\alpha_2 = \alpha_2 - \alpha_{\text{pk}} = \frac{p_1\Delta\alpha}{p_1 + p_2} \tag{4.1.19b}$$

and we have corresponding results for $\Delta\varepsilon_1$ and $\Delta\varepsilon_2$:

$$\Delta\varepsilon_1 = \varepsilon_1 - \varepsilon_{pk} = -\frac{p_2\Delta\varepsilon}{p_1 + p_2} \quad \text{and} \quad \Delta\varepsilon_2 = \varepsilon_2 - \varepsilon_{pk} = \frac{p_1\Delta\varepsilon}{p_1 + p_2}. \qquad (4.1.20)$$

From (4.1.5) the gain loss at the position of signal k due to a null is given by μ_k where

$\mu_k \approx 4\pi^2 \mathbf{e}_\phi{}^{\mathrm{T}} \mathbf{M} \mathbf{e}_\phi \Delta\theta_k{}^2$ and $\Delta\theta_k{}^2 = \Delta\alpha_k{}^2 + \Delta\varepsilon_k{}^2$ so, from (4.1.19) and (4.1.20),

$$\Delta\theta_1{}^2 = \left(\frac{p_2}{p_1 + p_2}\right)^2 (\Delta\alpha^2 + \Delta\varepsilon^2) = \left(\frac{p_2}{p_1 + p_2}\right)^2 \Delta\theta^2 \text{ and } \Delta\theta_2{}^2 = \left(\frac{p_1}{p_1 + p_2}\right)^2 \Delta\theta^2. \quad (4.1.21)$$

We now equate the residues, plus noise, to the detection threshold:

$$\mu_1 p_1 + \mu_2 p_2 + p_n \geq (1 + u_\beta) p_n$$

or

$$\mu_1 p_1 + \mu_2 p_2 \geq u_\beta p_n \qquad (4.1.22)$$

From (4.1.5) and (4.1.21) we have

$$\mu_1 p_1 + \mu_2 p_2 = 4\pi^2 \mathbf{e}_\phi{}^{\mathrm{T}} \mathbf{M} \mathbf{e}_\phi \Delta\theta^2 \left(\frac{p_2{}^2 p_1}{(p_1 + p_2)^2} + \frac{p_1{}^2 p_2}{(p_1 + p_2)^2}\right) = \left(\frac{p_1 p_2}{p_1 + p_2}\right) 4\pi^2 \mathbf{e}_\phi{}^{\mathrm{T}} \mathbf{M} \mathbf{e}_\phi \Delta\theta^2$$

and putting this into (4.1.22) gives

$$\left(\frac{p_1 p_2}{p_1 + p_2}\right) 4\pi^2 \mathbf{e}_\phi{}^{\mathrm{T}} \mathbf{M} \mathbf{e}_\phi \Delta\theta^2 \geq u_\beta p_n .$$

Using $4\pi^2 \mathbf{e}_\phi{}^{\mathrm{T}} \mathbf{M} \mathbf{e}_\phi \Delta\theta^2 = 2.34\Delta\theta^2 / \theta_B{}^2 = 2.34 / r^2$ from (4.1.8) we have finally

$$r = 1.53\sqrt{\frac{p_1 p_2}{(p_1 + p_2) u_\beta p_n}} = 1.53\sqrt{\frac{\sigma_1 \sigma_2}{(\sigma_1 + \sigma_2) u_\beta}} = 1.53\sqrt{\frac{\rho_1 \rho_2}{(\rho_1 + \rho_2)}} \qquad (4.1.23)$$

where $\sigma_k = p_k / p_n$, the signal to noise value of signal k and $\rho_k = p_k / u_\beta p_n$ is the ratio of signal k power to the minimum detectable signal, to replace (4.1.10). This expression will give little difference in its result compared with (4.1.10) when signal 1 is much larger than signal 2, but when they are closer it will make more difference. In particular, if they are equal (at value ρ) then (4.1.23) gives $r = 1.53\sqrt{(\rho/2)}$ which reduces the resolution figure by a factor $\sqrt{2}$ compared with (4.1.10).

4.2. RESOLUTION IN COHERENT (RADAR) CASE

4.2.1 Detection of a second target

In this case, instead of taking the IMP function as the starting point, we take the (negative, reduced) maximum likelihood function, in the form of equation (2.A.27) for the general case, or (2.A.33) for the coherent case. This function is minimized when the correct signal PSVs are used for the projection matrix $\mathbf{Q_A}$ (projecting into the orthogonal space of $\mathbf{A}$). If there is only one target present then there will be only residual noise left, so if the function is over the threshold corresponding to this noise level (set high enough to keep the false alarm rate at a suitable level), we deduce that there is at least one more target present. In the case of two targets this does not actually qualify as resolving them, as two distinct estimated target positions have not been found, but at least a second target (or more) has been detected. (We note that in the case of MUSIC two signals can be detected, through finding the number of non-trivial eigenvalues, but may not necessarily be resolved, as only one MUSIC peak may be formed if the targets are close enough.)

In the case of a single data frame the received set of values across the array is given by

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n} \tag{4.2.1}$$

where, as before, $\mathbf{A}$ is a $n \times m$ matrix of the PSVs of the m targets, $\mathbf{s}$ is the set of complex amplitudes of the signals (at the sampling instant) and $\mathbf{n}$ is an n-vector of noise samples. This could represent the radar case, using a single pulse. [NB: If we had p samples but the $\mathbf{s}$ vector remained essentially constant over the set of samples, as is effectively the case for radar, as long as the total sampling time is short compared with any Doppler frequency period, then $\mathbf{n}$ is replaced by $\bar{\mathbf{n}}$, the mean value of $\mathbf{n}$ over the p vectors (see §2A.3(b)). The components of $\bar{\mathbf{n}}$ are from a distribution with variance ψ/p if ψ is the variance of the distribution from which the components of $\mathbf{n}$ are samples.] The system sample covariance matrix, $\mathbf{R_Y} = \mathbf{Y}\mathbf{Y}^H$ in the multiple sample case, becomes just the rank one matrix $\mathbf{y}\mathbf{y}^H$ in this case.

Thus the likelihood function is, in the radar case,

$$F(\boldsymbol{\Theta}) = \text{trace}(\mathbf{Q_A}\mathbf{R_Y}) = \text{trace}(\mathbf{Q_A}\mathbf{y}\mathbf{y}^H) = \mathbf{y}^H\mathbf{Q_A}\mathbf{y} = \|\mathbf{Q_A}\mathbf{y}\|^2 \tag{4.2.2}$$

where $\boldsymbol{\Theta}$ is the $m \times q$ array of the parameters which define $\mathbf{A}$ – q parameters for each of the m targets. We have used the result $\text{trace}(\mathbf{U}\mathbf{V}) = \text{trace}(\mathbf{V}\mathbf{U})$ for any matrices of compatible size ($\text{size}(\mathbf{U}) = \text{size}(\mathbf{V}^T)$) and the fact that the trace of a scalar (or 1×1

matrix) is just the scalar value. (The last equality in (4.2.2) results from $\mathbf{Q} = \mathbf{Q}^2 = \mathbf{Q}^H\mathbf{Q}$ for a projection matrix.) The value for the general matrix $\mathbf{A}$ in (4.2.2) which actually minimizes F and maximizes the likelihood is $\mathbf{A}$ as given in (4.2.1) and in this case we note $\mathbf{Q}_\mathbf{A}\mathbf{y} = \mathbf{Q}_\mathbf{A}\mathbf{n}$ as $\mathbf{Q}_\mathbf{A}\mathbf{A} = \mathbf{0}$. We use this result to set a threshold for detecting a second target on the first scan.

If there is only one signal present then the peak on the first scan will occur (in the absence of errors) at the position of this signal, given by parameters $\boldsymbol{\theta}_1$ with PSV $\mathbf{a}_1$. The value of the function F with a null set using these parameters (by the projection matrix $\mathbf{Q}_{\mathbf{a}_1}$) will simply be $\mathbf{n}^H\mathbf{Q}_\mathbf{A}\mathbf{n}$ which will be below the threshold with probability $1 - \varepsilon_{FA}$, if ε_{FA} is the false alarm probability. If there is a second signal present then the peak will not actually be at $\boldsymbol{\theta}_1$ in parameter space, but on the line through $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ (as shown in Appendix 4B). In this case neither target 1 nor target 2 will be removed completely, and the residue will be greater than $\mathbf{Q}_\mathbf{A}\mathbf{n}$. We take the second target to be detected if the noise plus residue exceeds the threshold, as this is the condition that the single target model ($\mathbf{A} = \mathbf{a}_1$) is not adequate to account for the data and maximize the likelihood, and a higher order model is needed. We now need to determine the level of the ML function with a null at the position of the peak of the first IMP scan.

Using $\boldsymbol{\theta}$ for the parameter vector $\begin{bmatrix}\alpha\\ \varepsilon\end{bmatrix}$ in general (and $\boldsymbol{\theta}_k$ for $\begin{bmatrix}\alpha_k\\ \varepsilon_k\end{bmatrix}$, $\Delta\boldsymbol{\theta}$ for $\begin{bmatrix}\Delta\alpha\\ \Delta\varepsilon\end{bmatrix}$ and so on) and a circumflex over a variable to indicate its value at the peak position, we have

$$\hat{\mathbf{Q}} = \mathbf{I} - \hat{\mathbf{a}}\hat{\mathbf{a}}^H \text{ and (from (4.B.16))} \ \hat{\boldsymbol{\theta}} = \frac{S_1\boldsymbol{\theta}_1 + S_2\boldsymbol{\theta}_2}{S_1 + S_2} \tag{4.2.3}$$

with $S_k = \sigma_k + r$ as in (4.B.15). If there are two signals present then

$$\hat{F} = \mathbf{y}^H\hat{\mathbf{Q}}\mathbf{y} = (\mathbf{s}^H\mathbf{A}^H + \mathbf{n}^H)\hat{\mathbf{Q}}(\mathbf{A}\mathbf{s} + \mathbf{n}) = \left\|\hat{\mathbf{Q}}\mathbf{A}\mathbf{s}\right\|^2 + 2\,\mathrm{Re}(\mathbf{n}^H\hat{\mathbf{Q}}\mathbf{A}\mathbf{s}) + \mathbf{n}^H\hat{\mathbf{Q}}\mathbf{n}\,. \tag{4.2.4}$$

We neglect the signal-noise cross term (this will be more justifiable with large n) and approximate the last term to the expectation value $(n - 1)\psi$. If we compare $\hat{F}$ with the threshold given in Appendix 4A (eq. (4.A.8)) we see that we detect a second target if

$$\left\|\hat{\mathbf{Q}}\mathbf{A}\mathbf{s}\right\|^2 + (n-1)\psi \geq (n-1)\psi + t_\delta\psi\sqrt{2(n-1)}$$

or

$$\left\|\hat{\mathbf{Q}}\mathbf{A}\mathbf{s}\right\|^2 \geq t_\delta \psi\sqrt{2(n-1)} = u_\delta \psi \tag{4.2.5}$$

where

$$u_\delta = t_\delta\sqrt{2(n-1)}\,, \tag{4.2.6}$$

and we take the resolution limit to be when the equality holds.

Now we have

$$\left\|\hat{\mathbf{Q}}\mathbf{A}\mathbf{s}\right\|^2 = \left\|\hat{\mathbf{Q}}\mathbf{a}_1 s_1 + \hat{\mathbf{Q}}\mathbf{a}_2 s_2\right\|^2 = \left\|\hat{\mathbf{Q}}\mathbf{a}_1\right\|^2 |s_1|^2 + 2\,\mathrm{Re}\left((\hat{\mathbf{Q}}\mathbf{a}_1)^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}_2 s_1 * s_2\right) + \left\|\hat{\mathbf{Q}}\mathbf{a}_2\right\|^2 |s_2|^2 \tag{4.2.7}$$

and

$$\left\|\hat{\mathbf{Q}}\mathbf{a}_k\right\|^2 = \mathbf{a}_k{}^{\mathrm{H}}\mathbf{Q}_{\hat{\mathbf{a}}}\mathbf{a}_k = 1 - \left|\mathbf{a}_k{}^{\mathrm{H}}\hat{\mathbf{a}}\right|^2 = 1 - \left|\hat{\gamma}_k\right|^2 = \hat{\mu}_k \tag{4.2.8}$$

where

$$\hat{\gamma}_k = \mathbf{a}_k{}^{\mathrm{H}}\hat{\mathbf{a}} \text{ and } \hat{\mu}_k = 1 - \left|\hat{\gamma}_k\right|^2. \tag{4.2.9}$$

Also

$$(\hat{\mathbf{Q}}\mathbf{a}_1)^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}_2 = \mathbf{a}_1{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}_2 = \mathbf{a}_1{}^{\mathrm{H}}\mathbf{a}_2 - \mathbf{a}_1{}^{\mathrm{H}}\hat{\mathbf{a}}\hat{\mathbf{a}}^{\mathrm{H}}\mathbf{a}_2 = \gamma - \hat{\gamma}_1\hat{\gamma}_2\,*. \tag{4.2.10}$$

Now the general result in (4.1.4),

$$\gamma = 1 - 2\pi^2\mathbf{e}_\phi{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_\phi\Delta\theta^2$$

(to lowest order approximation) gives γ in terms of the distance between two points in parameter space, $\Delta\theta$, along the direction ϕ (see the diagram above (4.1.6)). We have already noted that $\hat{\boldsymbol{\theta}}$ lies on the line (in parameter space) joining $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ so $\Delta\hat{\boldsymbol{\theta}}_1$ and $\Delta\hat{\boldsymbol{\theta}}_2$ have the same orientation as $\Delta\boldsymbol{\theta}$, and so this result is applicable to $\hat{\gamma}_k$, with the change of scale, giving

$$\hat{\gamma}_k = 1 - 2\pi^2\mathbf{e}_\phi{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_\phi\Delta\hat{\theta}_k{}^2$$

where $\Delta\hat{\theta}_k = \left\|\Delta\hat{\boldsymbol{\theta}}_k\right\| = \left\|\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}\right\|$. For more compact expressions we put

$$m = 4\pi^2\mathbf{e}_\phi{}^{\mathrm{T}}\mathbf{M}\mathbf{e}_\phi \tag{4.2.11}$$

to give

$$\hat{\gamma}_k = 1 - m\Delta\hat{\theta}_k{}^2/2 \text{ and } \hat{\mu}_k = m\Delta\hat{\theta}_k{}^2 \tag{4.2.12}$$

to the same approximation.

From (4.C.11) we have

$$\Delta\hat{\boldsymbol{\theta}}_1 = \boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}} = \frac{-S_2\Delta\boldsymbol{\theta}}{S_1 + S_2} \text{ and } \Delta\hat{\boldsymbol{\theta}}_2 = \boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}} = \frac{S_1\Delta\boldsymbol{\theta}}{S_1 + S_2}$$

where $\Delta\boldsymbol{\theta} = \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1$. Using these results

$$\hat{\mu}_1 = 1 - \left|\hat{\gamma}_1\right|^2 = m\Delta\hat{\theta}_1{}^2 = \frac{mS_2{}^2\Delta\theta^2}{(S_1 + S_2)^2} \text{ and } \hat{\mu}_2 = \frac{mS_1{}^2\Delta\theta^2}{(S_1 + S_2)^2}. \tag{4.2.13}$$

Also, from (4.2.10) and (4.2.12) we have, to second order,

$$(\hat{\mathbf{Q}}\mathbf{a_1})^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}_2 = (1 - \frac{m\Delta\theta^2}{2}) - (1 - \frac{mS_2{}^2\Delta\theta^2}{2(S_1 + S_2)^2})(1 - \frac{mS_1{}^2\Delta\theta^2}{2(S_1 + S_2)^2})$$

$$= \frac{m\Delta\theta^2}{2}\left(\frac{S_1{}^2 + S_2{}^2}{(S_1 + S_2)^2} - 1\right) = -\frac{mS_1S_2\Delta\theta^2}{(S_1 + S_2)^2}. \tag{4.2.14}$$

Putting results (4.2.8), (4.2.13) and (4.2.14) into (4.2.7) we have

$$\left\|\hat{\mathbf{Q}}\mathbf{As}\right\|^2 = \frac{m\Delta\theta^2}{(S_1 + S_2)^2}\left(S_2{}^2\left|s_1\right|^2 - 2S_1S_2\,\mathrm{Re}\left(s_1 * s_2\right) + S_1{}^2\left|s_2\right|^2\right). \tag{4.2.15}$$

Now only the relative phase factor of s_1 and s_2 is significant, rather than the individual phase values, so let ϕ be the phase difference between them. Then for $\mathrm{Re}(s_1{*}s_2)$ we now put $|s_1||s_2|\cos\phi$ and $S_k = |s_k|^2 + |s_1||s_2|\cos\phi$, then, with some rearranging, (4.2.15) becomes

$$\left\|\hat{\mathbf{Q}}\mathbf{As}\right\|^2 = \frac{m\Delta\theta^2\left|s_1\right|^2\left|s_2\right|^2\sin^2\phi}{\left|s_1\right|^2 + 2\left|s_1\right|\left|s_2\right|\cos\phi + \left|s_2\right|^2}. \tag{4.2.16}$$

This shows that if the amplitudes are in phase ($\phi = 0$) or antiphase ($\phi = \pi$) then the target residue is zero to second order in $\Delta\theta$, and a higher order approximation is required. However for relative phases not close to these values the residue is proportional to $\Delta\theta^2$.

From (4.2.5) and (4.2.16) a second signal is detected when

$$\frac{m\Delta\theta^2\left|s_1\right|^2\left|s_2\right|^2\sin^2\phi}{\left|s_1\right|^2 + 2\left|s_1\right|\left|s_2\right|\cos\phi + \left|s_2\right|^2} \geq u_\delta\psi$$

with u_δ given in (4.2.6), and so the resolution threshold is given by

$$\Delta\theta^2 = \frac{|s_1|^2 + 2|s_1||s_2|\cos\phi + |s_2|^2}{m|s_1|^2|s_2|^2\sin^2\phi} u_\delta \psi .$$

In terms of the target signal to noise ratios $\sigma_k = |s_k|^2/\psi$ or the signal level relative to the minimum detectable level at this point, $\rho_k = |s_k|^2/\psi u_\varepsilon$ we have

$$\Delta\theta = \sqrt{\frac{\sigma_1 + 2\sqrt{\sigma_1\sigma_2}\cos\phi + \sigma_2}{m\sigma_1\sigma_2\sin^2\phi} u_\delta} = \sqrt{\frac{\rho_1 + 2\sqrt{\rho_1\rho_2}\cos\phi + \rho_2}{m\rho_1\rho_2\sin^2\phi}} . \quad (4.2.17)$$

With m given in (4.2.11). In the case of quadrature echoes ($\phi = \pm\pi/2$) we have

$$\Delta\theta = \sqrt{\frac{\sigma_1 + \sigma_2}{m\sigma_1\sigma_2} u_\delta} = \sqrt{\frac{\rho_1 + \rho_2}{m\rho_1\rho_2}} . \quad (4.2.18)$$

We see that this is as (4.1.6) with $\rho_1\rho_2/(\rho_1 + \rho_2)$ replacing ρ, and (4.2.18) also leads to the same result as (4.1.23), for the non-coherent case. Thus the case of coherent echoes *in quadrature* is exactly the same, in terms of resolution, as the non-coherent case.

4.2.2 Resolution of a second target

In order to find the second target in the way that IMP operates we need to find the peak value on the second scan, with a single null at the peak position, given by (4.B.16) in Appendix B. We will use a circumflex over all functions evaluated at this point, given by $\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\alpha} \\ \hat{\varepsilon} \end{bmatrix}$. A problem with evaluating the function when a null is included is that the function cannot be evaluated at the null point as both the numerator and the denominator of the function are zero (as $\hat{\mathbf{Q}}\hat{\mathbf{a}} = \mathbf{Q}_{\hat{\mathbf{a}}}\hat{\mathbf{a}} = \mathbf{0}$ where $\hat{\mathbf{a}} = \mathbf{a}(\hat{\boldsymbol{\theta}})$). However, the function does have a finite value at this point, determined using l'Hôpital's Rule as shown in Appendix 4C. Thus we have

$$f(\hat{\boldsymbol{\theta}}) = \frac{\hat{\mathbf{a}}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{R}\hat{\mathbf{Q}}\hat{\mathbf{a}}}{\hat{\mathbf{a}}^{\mathrm{H}}\hat{\mathbf{Q}}\hat{\mathbf{a}}} = \frac{\hat{\mathbf{a}}'^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{R}\hat{\mathbf{Q}}\hat{\mathbf{a}}'}{\hat{\mathbf{a}}'^{\mathrm{H}}\hat{\mathbf{Q}}\hat{\mathbf{a}}'}$$

in general, or, for the single data frame case,

$$f(\hat{\boldsymbol{\theta}}) = \frac{\hat{\mathbf{a}}'^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{y}\mathbf{y}^{\mathrm{H}}\hat{\mathbf{Q}}\hat{\mathbf{a}}'}{\hat{\mathbf{a}}'^{\mathrm{H}}\hat{\mathbf{Q}}\hat{\mathbf{a}}'} . \quad (4.2.19)$$

where $\hat{\mathbf{a}}'$ is the derivative of $\hat{\mathbf{a}}$ with respect to any suitable parameter. In this section we assume that the two targets at $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, at the limit of resolution, are close enough

to $\hat{\boldsymbol{\theta}}$ to allow the use of this form for $f(\boldsymbol{\theta})$ in the region of interest, including the peak of the second scan, yet to be found. We also assume that the peak of the second scan will occur on the line through $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, as is $\hat{\boldsymbol{\theta}}$, the position of the peak of the first scan, and let all $\boldsymbol{\theta}$ values be of the form $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + x\Delta\boldsymbol{\theta}$ (and $\boldsymbol{\theta}_k = \hat{\boldsymbol{\theta}} + x_k \Delta\boldsymbol{\theta}$).

With $\mathbf{y} = \mathbf{As} + \mathbf{n}$ the function at $\boldsymbol{\theta}$ becomes

$$f(\boldsymbol{\theta}) = \frac{\mathbf{a}'^{\mathrm{H}}\hat{\mathbf{Q}}(\mathbf{As}+\mathbf{n})(\mathbf{s}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}+\mathbf{n}^{\mathrm{H}})\hat{\mathbf{Q}}\mathbf{a}'}{\mathbf{a}'^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'} = \frac{\mathbf{a}'^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{Ass}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'}{\mathbf{a}'^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'} + \frac{\mathbf{a}'^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{nn}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'}{\mathbf{a}'^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'}$$

on ignoring signal-noise cross terms. Putting $\mathbf{u} = \hat{\mathbf{Q}}\mathbf{a}' \big/ \left\|\hat{\mathbf{Q}}\mathbf{a}'\right\|$ (with $\mathbf{a}'^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}' = \left\|\hat{\mathbf{Q}}\mathbf{a}'\right\|^2$), the noise term becomes $|\mathbf{u}^{\mathrm{H}}\mathbf{n}|^2$. If the noise terms are from zero-mean normal distributions with variance ψ (~N[0,ψ]) and $\mathbf{u}$ is a unit norm vector, then $\mathbf{u}^{\mathrm{H}}\mathbf{n}$ is also a sample from this normal distribution (as shown for the components of $\mathbf{v}$ in Appendix 4A). Thus $|\mathbf{u}^{\mathrm{H}}\mathbf{n}|^2$ has a Rayleigh distribution (or, following Appendix 4A, a χ^2 distribution with only one degree of freedom (p = 1)). This distribution determines the threshold for any required false alarm probability β, which we define as $(\tau_\beta + 1)\,\psi_n$, relating the threshold to the noise level so that we require, for detection of a second signal on the second scan that the signal contribution to f above, f_s, should exceed $\tau_\beta\,\psi_n$, i.e.

$$f_s(\boldsymbol{\theta})/\psi = \frac{\left|\mathbf{s}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'\right|^2}{\psi\mathbf{a}'^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'} \geq \tau_\beta \qquad (4.2.20)$$

with equality at the resolution limit.

Expanding the numerator, putting $\mathbf{s}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}} = s_1{}^*\mathbf{a}_1{}^{\mathrm{H}} + s_2{}^*\mathbf{a}_2{}^{\mathrm{H}}$ we have

$$\left|\mathbf{s}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'\right|^2 = \left|s_1{}^*\mathbf{a}_1{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}' + s_2{}^*\mathbf{a}_2{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'\right|^2$$

$$= |s_1|^2\left|\mathbf{a}_1{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'\right|^2 + 2\,\mathrm{Re}\left(s_1{}^*s_2\mathbf{a}_1{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'(\mathbf{a}_2{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}')^*\right) + |s_2|^2\left|\mathbf{a}_2{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'\right|^2$$

It is found in (4.2.22) below that the products $\mathbf{a}_k{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'$ are in fact real, so we have

$$\left|\mathbf{s}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'\right|^2 = |s_1|^2\,\mathbf{a}_1{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'^2 + 2\,\mathrm{Re}\left(s_1{}^*s_2\right)\mathbf{a}_1{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'\mathbf{a}_2{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}' + |s_2|^2\,\mathbf{a}_2{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'^2$$

$$= \psi\left(\sigma_1\mathbf{a}_1{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'^2 + 2r\mathbf{a}_1{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'\mathbf{a}_2{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}' + \sigma_2\mathbf{a}_2{}^{\mathrm{H}}\hat{\mathbf{Q}}\mathbf{a}'^2\right) \qquad (4.2.21)$$

where $\sigma_k = |s_k|^2 / \psi$ and $r = \mathrm{Re}(s_1 s_2^*)/\psi$ as in (4.B.15).

Putting $\hat{\mathbf{Q}} = \mathbf{I} - \hat{\mathbf{a}}\hat{\mathbf{a}}^{\mathrm{H}}$ we can now evaluate the signal term, requiring a number of inner products. In general, if $\mathbf{a}_k = \mathbf{a}(\boldsymbol{\theta}_k) = \mathbf{a}(\hat{\boldsymbol{\theta}} + x_k \Delta\boldsymbol{\theta})$ and $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{a}(\hat{\boldsymbol{\theta}} + x\Delta\boldsymbol{\theta})$ the inner products are given in (4.C.6) – (4.C.10) (assuming μx^2 is small for all values of x used). Using these results we have

$$\mathbf{a}_k{}^{\mathrm{H}} \hat{\mathbf{Q}} \mathbf{a}' = -\mu(x - x_k) - (1 - \mu x_k{}^2/2)(-\mu x) = \mu x_k (1 - \mu x_k x/2) \qquad (4.2.22)$$

Thus (4.2.21) becomes

$$\left|\mathbf{s}^{\mathrm{H}} \mathbf{A}^{\mathrm{H}} \hat{\mathbf{Q}} \mathbf{a}'\right|^2 = \psi\mu^2 \left(\sigma_1 \left(x_1(1 - \mu x_1 x/2)\right)^2 + 2r x_1 (1 - \mu x_1 x/2) x_2 (1 - \mu x_2 x/2) + \right.$$

$$\left. + \sigma_2 \left(x_2 (1 - \mu x_2 x/2)\right)^2 \right). \qquad (4.2.23)$$

The numerator of the target contribution in (4.2.20) is thus $N(x)$, given, from (4.2.23) by

$$N(x) = \psi\mu^2 (A - 2Bx + Cx^2) \qquad (4.2.24)$$

where

$$A = \sigma_1 x_1{}^2 + 2r x_1 x_2 + \sigma_2 x_2{}^2 \qquad (4.2.25a)$$

$$B = (\mu/2)\left(\sigma_1 x_1{}^3 + 2r x_1 x_2 (x_1 + x_2) + \sigma_2 x_2{}^3\right) \qquad (4.2.25b)$$

$$C = (\mu/2)^2 \left(\sigma_1 x_1{}^4 + 2r x_1{}^2 x_2{}^2 + \sigma_2 x_2{}^4\right). \qquad (4.2.25c)$$

Using (4.C.11) for x_1 and x_2, equations (4.2.25) can be put in the form

$$A(S_1 + S_2)^2 = \sigma_1 S_2{}^2 2r S_1 S_2 + \sigma_2 S_1{}^2 \qquad (4.2.26a)$$

$$B(S_1 + S_2)^3 = (\mu/2)\left(-\sigma_1 S_2{}^3 + 2r S_1 S_2 (S_1 + S_2) + \sigma_2 S_1{}^3\right) \qquad (4.2.26b)$$

$$C(S_1 + S_2)^4 = (\mu/2)^2 \left(\sigma_1 S_2{}^4 + 2r S_1{}^2 S_2{}^2 + \sigma_2 S_1{}^4\right). \qquad (4.2.26c)$$

The denominator is $D(x)$, given by

$$D(x)/\psi = \mathbf{a}'^{\mathrm{H}} \hat{\mathbf{Q}} \mathbf{a}' = \|\mathbf{a}'\|^2 - \left|\hat{\mathbf{a}}^{\mathrm{H}} \mathbf{a}'\right|^2 = \mu - \mu^2 x^2 = \mu(1 - \mu x^2) \qquad (4.2.27)$$

using (4.C.9) and (4.C.10). Thus (4.2.20) becomes

$$f_s(x)/\psi = \frac{N(x)}{D(x)} = \mu \frac{A - 2Bx + Cx^2}{1 - \mu x^2}. \qquad (4.2.28)$$

Again assuming μx^2 is small we can put, approximating to second order,

$$f_s(x)/\psi\mu \approx (A - 2Bx + Cx^2)(1 + \mu x^2) \approx A - 2Bx + (C + \mu A)x^2 . \qquad (4.2.29)$$

This is a quadratic function of x and we find that the peak value is given by

$$f_s(\hat{x})/\psi\mu = A - \frac{B^2}{C + \mu A} \qquad (4.2.30)$$

where

$$\hat{x} = \frac{B}{C + \mu A} \qquad (4.2.31)$$

is the value of x at the peak position of the second scan.

On putting in the expressions for A, B and C from (4.2.26) and S_k as in (4.B.15), we obtain the peak value in terms of σ_1, σ_2, ϕ and μ only. As μ contains $\Delta\theta$, we can obtain an expression for the limiting separation in terms of the signal powers and the array parameters, or, using (4.1.9), we can obtain an expression for the resolution improvement factor r.

Expressions for A, B and C are obtained in Appendix C, but these show that A and B are both proportional to $\sin^2\phi$, where ϕ is the relative phase of the two signal amplitudes. Thus both these have value zero in the in-phase ($\phi = 0$) and antiphase ($\phi = \pi$) cases. This means (from (4.2.30)) that the apparent peak signal residue is zero, which is not a realistic result. In fact we have been approximating the inner products to lowest order (as $\Delta\theta^2$) and we deduce that the peak value, if expressed as a power series in $\Delta\theta$, is of higher order than two. We do not expect a significant response of order three, but rather of order four. Setting the peak value as $c(\sigma_1,\sigma_2,\phi)\Delta\theta^4$ and equating this to some threshold value τ we have $\Delta\theta = (\tau/c)^{1/4}$ – the separation at the limit of resolution, in the in-phase and antiphase cases, is the fourth root of a function of the signal powers (taken, from noting the other results obtained to be linear, or of the form $\sigma_1\sigma_2/(\sigma_1+\sigma_2)$). The implication is that the resolution only improves slowly as the powers are increased. For example if both powers are increased by 10dB the resolution improvement ratio will only increase by about 1.8 times.

However we can use the results for the quadrature case ($\phi = \pm\pi/2$) and for a range of angles for which the second order approximation dominates. (Unfortunately, until the fourth order term is determined this range cannot be found, but it is likely that the

second order result will be applicable over most of the angular range.) Here we investigate the strictly quadrature case.

From equations (4.C.16), (4.C.17) and (4.C.18) with $\phi = \pm\pi/2$ (so $r = \sqrt{\sigma_1\sigma_2}\cos\phi = 0$, $S_k = \sigma_k$, $\sin^2\phi = 1$) we have

$$A = \frac{S_1 S_2}{S_1 + S_2} = \frac{\sigma_1\sigma_2}{\sigma_1 + \sigma_2} = \sigma_{12}, \tag{4.2.32}$$

$$B = \frac{\mu}{2}\frac{(S_1^2 - S_2^2)S_1S_2}{(S_1 + S_2)^3} = \frac{\mu}{2}\frac{S_1 - S_2}{S_1 + S_2}\frac{S_1S_2}{S_1 + S_2} = \frac{\mu}{2}\delta_{12}\sigma_{12}, \tag{4.2.33}$$

where

$$\delta_{12} = \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2}, \tag{4.2.34}$$

and

$$C = \left(\frac{\mu}{2}\right)^2\left(\frac{(S_1 - S_2)^2 S_1S_2}{(S_1 + S_2)^3} + \frac{S_1^2S_2^2}{(S_1 + S_2)^3}\right) = \left(\frac{\mu}{2}\right)^2\sigma_{12}\left(\frac{(S_1 - S_2)^2}{(S_1 + S_2)^2} + \frac{S_1S_2}{(S_1 + S_2)^2}\right)$$

$$= \left(\frac{\mu}{2}\right)^2\sigma_{12}\left(\frac{(\sigma_1 - \sigma_2)^2}{(\sigma_1 + \sigma_2)^2} + \frac{\sigma_1\sigma_2}{(\sigma_1 + \sigma_2)^2}\right) = \left(\frac{\mu}{2}\right)^2\sigma_{12}\varepsilon_{12}. \tag{4.2.35}$$

where

$$\varepsilon_{12} = 1 - 3\sigma_1\sigma_2/(\sigma_1 + \sigma_2)^2. \tag{4.2.36}$$

We note that $\sigma_1\sigma_2/(\sigma_1 + \sigma_2)^2 \le 1/4$ (from $(\sigma_1 - \sigma_2)^2 \ge 0$, so that $(\sigma_1 + \sigma_2)^2 \ge 4\sigma_1\sigma_2$, with equality when $\sigma_1 = \sigma_2$) so $1/4 \le \varepsilon_{12} < 1$.

We take two extreme conditions, (a) $\sigma_1 >> \sigma_2$ and (b) $\sigma_1 = \sigma_2$.

(a) In this case, as $\sigma_1/\sigma_2 \to \infty$, from (4.2.32) and (4.2.36) we have $\delta_{12} \to 1$ and $\varepsilon_{12} \to 1$, and so, from (4.2.32), (4.2.33) and (4.2.35) we have A $\to \sigma_2$, B $\to \mu\sigma_2/2$ and C $\to (\mu/2)^2\sigma_2$. Using these results, the peak position, from (4.2.31), is at $2/(\mu+4)$ which is slightly under ½. (As, from (4.C.11), we have x_1 near zero in this case (i.e the first scan peak is near to x_1 as expected, this moves the second scan peak to close to the midpoint between the targets. We continue to make the assumption that μ is small if conditions are such as to give good resolution.) The peak value is given, from (4.2.30), by

$$f_s(\hat{x})/\psi\mu = \sigma_2\left(1-\frac{(\mu/2)^2}{(\mu/2)^2+\mu}\right)=\frac{\sigma_2}{1+\mu/4}$$

which is slightly less than σ_2. If we approximate this to σ_2 the resolution limit, from (4.2.20), is when

$$\mu\sigma_2 = \tau_\delta$$

and, putting $\mu = 2.34/r^2$, where *r* is the resolution improvement factor, we obtain

$$r^2 = 2.34\sigma_2/\tau_\delta \qquad (4.2.37)$$

where σ_2 is the signal to noise ratio in the array for signal 2 and signal 1 is much greater than signal 2. We see that the resolution improvement rises as the square root of the weaker signal strength, without bound, until in practice limited by system errors.

(b) Let the equal signal strengths (signal to noise ratios) be σ, then in this case we have $\sigma_{12} = \sigma/2$, $\delta_{12} = 0$ and $\varepsilon_{12} = 1/4$, giving A = $\sigma/2$, B = 0 and C = $\mu^2\sigma/32$. The peak position is at $\hat{x}=0$, coinciding with the first peak position, which is the midpoint between the targets from (4.C.11), as expected, and the peak value is given by

$$f_s(\hat{x})/\psi\mu = \sigma/2.$$

The resolution improvement factor is thus given by

$$r^2 = 2.34\sigma/2\tau_\delta = 1.17\sigma/\tau_\delta \qquad (4.2.38)$$

and we see that this gives a result for the resolution measure which is a factor of $\sqrt{2}$ lower than for the case where one signal dominates strongly.

[NB: We note that the second scan peak, in the equal signal case coincides with the first. In this case continuing with the basic IMP algorithm is not possible as the second null will simply be the same as the first. In fact, having determined that there is a second signal present, the procedure is to form two new null positions separated by perhaps ½ beamwidth, and proceed with the tweaking convergence.]

APPENDIX 4A: DETECTION THRESHOLDS

4A.1 IMP scan threshold

The IMP function is given, at stage *k* + 1 (i.e. after finding *k* targets), in Chapter 2 (eq. (2.2.27)) by

$$f_k(\boldsymbol{\theta}) = \frac{\mathbf{b}(\boldsymbol{\theta})^{\mathrm{H}} \mathbf{R}_{\mathbf{Y}} \mathbf{b}(\boldsymbol{\theta})}{\left\|\mathbf{b}(\boldsymbol{\theta})\right\|^2} \tag{4.A.1}$$

where $\mathbf{b}(\boldsymbol{\theta}) = \mathbf{Q}_{\mathbf{A}_k} \mathbf{a}(\boldsymbol{\theta})$ and $\mathbf{A}_k$ contains the PSVs of the signals found so far. $\mathbf{R}_{\mathbf{Y}} = \mathbf{Y}\mathbf{Y}^{\mathrm{H}}/p$ is the (estimated) system covariance matrix (taken over p data frames). If all m targets have been found the final scan becomes (putting $\mathbf{Y} = \mathbf{AS} + \mathbf{N}$)

$$f_m(\boldsymbol{\theta}) = \frac{\mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} (\mathbf{AS} + \mathbf{N})(\mathbf{S}^{\mathrm{H}} \mathbf{A}^{\mathrm{H}} + \mathbf{N}^{\mathrm{H}}) \mathbf{Q}_{\mathbf{A}} \mathbf{a}(\boldsymbol{\theta})}{p \left\|\mathbf{b}(\boldsymbol{\theta})\right\|^2} = \frac{\mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \mathbf{N}\mathbf{N}^{\mathrm{H}} \mathbf{Q}_{\mathbf{A}} \mathbf{a}(\boldsymbol{\theta})}{p \left\|\mathbf{b}(\boldsymbol{\theta})\right\|^2}$$

$$= \frac{\mathbf{b}(\boldsymbol{\theta})^{\mathrm{H}} \mathbf{N}\mathbf{N}^{\mathrm{H}} \mathbf{b}(\boldsymbol{\theta})}{p \left\|\mathbf{b}(\boldsymbol{\theta})\right\|^2} \tag{4.A.2}$$

as $\mathbf{Q}_{\mathbf{A}} \mathbf{A} = \mathbf{0}$. The expectation value of $\mathbf{N}\mathbf{N}^{\mathrm{H}}/p$ is $\psi_n \mathbf{I}$ where ψ_n is the noise level in each channel and $\mathbf{I}$ is the $n \times n$ identity matrix so the expectation value of the function is ψ_n. However, we need to know the statistics of f_m in order to set a detection threshold giving a suitably low false alarm rate.

We note that $\mathbf{u} = \mathbf{b}(\boldsymbol{\theta}) / \left\|\mathbf{b}(\boldsymbol{\theta})\right\|$ is a unit norm vector and we can put

$$f_m = \mathbf{u}^{\mathrm{H}} \mathbf{N}\mathbf{N}^{\mathrm{H}} \mathbf{u} / p = \mathbf{v}^{\mathrm{H}} \mathbf{v} = \sum_{k=1}^{p} \left| v_k \right|^2 . \tag{4.A.3}$$

where we omit the parameter vector $\boldsymbol{\theta}$ for clarity, and we have put $\mathbf{v} = \mathbf{N}^{\mathrm{H}} \mathbf{u} / \sqrt{p}$.

Now $v_k = \sum_{j=1}^{n} n_{jk} * u_j \Big/ \sqrt{p}$ where each component of $\mathbf{N}$ (e.g. n_{jk}) is from a zero mean normal distribution with variance ψ_n (i.e. $n_{jk} \sim \mathrm{N}[0, \psi_n]$) and it follows (see [30], for example) that

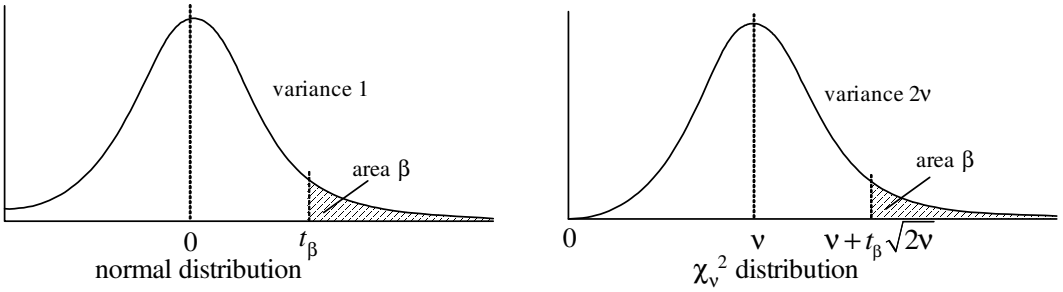


Figure 4.1 Probability density functions of normal and χ^2 distributions

$$v_k \sim \mathrm{N}[0, \sum_{j=1}^{n} |u_k|^2 \psi_n \Big/ p] = \mathrm{N}[0, \psi_n / p] \qquad (4.A.4)$$

as **u** is of norm unity. Thus we see that **v** is a p-vector with components from a zero mean normal distribution of variance ψ_n/p. This result is independent of $\boldsymbol{\theta}$ so we see that at all points over the IMP scan (i.e. over the parameter domain) the function is given by (4.A.3) and (4.A.4). If we look at the expectation value of f_m we have $\langle f_m \rangle = \mathbf{u}^{\mathrm{H}} \langle \mathbf{N}\mathbf{N}^{\mathrm{H}} \rangle \mathbf{u}/p = \mathbf{u}^{\mathrm{H}} \psi_n \mathbf{I} \mathbf{u} = \psi_n$, as expected.

A random variable y obtained by summing the squares of ν random variables x_k (i.e. $y = \Sigma x_k^2$) all with a zero mean, unit variance normal distribution (i.e. $x_k \sim \mathrm{N}[0,1]$ for k = 1 to ν) has a χ_ν^2 distribution, with a mean ν and variance 2ν. For a moderately large number of degrees of freedom ν, the χ_ν^2 distribution starts to approach a normal distribution as shown (in diagram form) in Figure 1. The main differences are that it has a non-zero mean and has no probability of negative values for y (from the definition of y) so is not quite symmetrical. Various approximations for the value of the cumulative χ_ν^2 distribution in terms of $\Phi(u)$ (the cumulative probability for the normal distribution – the area under the curve up to the point u) are available (see p.176 of [18], for example) but the simplest is to model the χ_ν^2 distribution as normal i.e. to take $y \sim \mathrm{N}[\nu, 2\nu]$ (see equation (26) on p.176 of [18]). This is relatively poor by comparison with the more complicated forms, but should be quite adequate for the application here. In this case there are p degrees of freedom, and the normal distribution has variance ψ_n/p so that the approximating distribution function for f is $\mathrm{N}[p \times \psi_n/p, 2p \times (\psi_n/p)^2]$, or $\mathrm{N}[\psi_n, 2\psi_n^2/p]$. The mean level of this distribution, ψ_n, agrees with the expectation level for f_m, obtained above.

To keep the false alarm rate down – i.e. to avoid deciding there is a signal present on a peak value due to noise alone (on the last scan, when all the signals have been nulled), we set a threshold such that this probability is quite low. If the rate is to be $\beta << 1$ then a threshold t_β is required at a few standard deviations from the mean, as shown in Figure 4.1. Because the χ_ν^2 distribution is similar to a normal distribution we can equate the threshold relative to the mean level to the equivalent threshold t_β for the normal distribution, where the area under the curve for this distribution, above this point, is also β. If $\Phi(u)$ is the cumulative probability for the normal distribution then t_β, is given by

$$\Pr[U < t_\beta] = \Phi(t_\beta) = 1 - \beta. \qquad (4.A.5)$$

In this case we have, for the normal distribution with mean ψ_n and variance $2\psi_n^2/p$, the required threshold is

$$T_\beta = \psi_n + t_\beta \sqrt{2\psi_n^{\,2}/p} = \psi_n(1 + t_\beta\sqrt{2/p}) = \psi_n(1+u_\beta) \qquad (4.A.6)$$

with t_β defined by (4.A.5) for a given β (or false alarm rate) and obtained from tables of Φ, and putting $u_\beta = t_\beta\sqrt{2/p}$. [NB: f_m is not flat, with a single value from the χ^2 distribution over the scan, as different **u** vectors sum noise samples differently so f_m varies over the scan region with values from this distribution. If f_m is evaluated at K points it might appear that for a false alarm rate over the whole scan δ to be some reasonably low value, where the false alarm rate on each point is β, then we have

$$\delta = 1 - (1-\beta)^K \approx 1 - e^{-K\beta} \approx K\beta$$

for $K\beta$ small. However, this assumes the noise sums differently at every sample point, however close the points may be. This will not be the case, as two close **u** vectors will sum the noise samples to two close values. The noise samples will be effectively independent at intervals of the beamwidth in each parameter domain, so we should choose

$$K = \prod_{k=1}^{q} (range_k / beamwidth_k)$$

where $range_k$ is the range of parameter k and $beamwidth_k$ is the natural beamwidth of the scan, given by the aperture in the observation domain. As example values, if we take $t_\beta = 4$ then $\beta = 3.2\times10^{-5}$ and if $K = 1000$ we have $\delta = 0.032$.]

4A.2 ML threshold for detection of second target

If only one target is present, the minimum of the ML function (i.e. the reduced ML function used for IMP) is given in the radar case, (i.e. with coherent targets) from (2.A.33) and (2.2.2) with the following discussion, by $F_{\min} = \left\| \mathbf{Q}_{\mathbf{a}_1} \mathbf{n} \right\|^2$, where $\mathbf{a}_1$ is the PSV of target 1, and **n** is the vector of noise samples across the n receiver channels. If the samples are from a zero mean normal distribution with variance ψ then $F_{\min}$ will have the a χ^2 distribution with $n-1$ degrees of freedom – not n (as $\|\mathbf{n}\|^2$ would have) as $\mathbf{Q}_{\mathbf{a}_1}$ projects the n-vector **n** into a space of $n-1$ dimensions. We can show this by noting that in general $\mathbf{Q}_\mathbf{A}$, where **A** is $n\times m$, has $n-m$ eigenvalues of unity and m of zero. Also, as **Q** is Hermitian its eigenvectors are orthogonal, so we can find a unitary

matrix **U** (i.e. $\mathbf{U}\mathbf{U}^{\mathrm{H}} = \mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{I}$) such that $\mathbf{U}\mathbf{Q}\mathbf{U}^{\mathrm{H}} = \mathbf{\Lambda} = \mathrm{diag}[1 \ldots 1\ 0 \ldots 0]$, containing *m* zeros. First we note that (for any *n*×*n* unitary matrix and *n*-vector **v**)

$$\|\mathbf{U}\mathbf{v}\|^2 = \mathbf{v}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{v} = \mathbf{v}^{\mathrm{H}}\mathbf{v} = \|\mathbf{v}\|^2.$$

Thus we can put

$$\|\mathbf{Q}_{\mathbf{a}_1}\mathbf{n}\|^2 = \|\mathbf{U}\mathbf{Q}_{\mathbf{a}_1}\mathbf{n}\|^2 = \|\mathbf{U}\mathbf{Q}_{\mathbf{a}_1}\mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{n}\|^2 = \|\mathbf{\Lambda}\mathbf{U}\mathbf{n}\|^2 = \sum_{k=1}^{n-1} n_k'^2$$

where $\mathbf{n}'$ (with components n_k', all from a distribution with variance ψ) is given by $\mathbf{U}\mathbf{n}$. As $\mathbf{\Lambda}\mathbf{n}'$ has only $n - 1$ non-zero components it follows that $\|\mathbf{Q}_{\mathbf{a}_1}\mathbf{n}\|^2$ has a χ^2 distribution with $n - 1$ degrees of freedom.

Thus $\|\mathbf{Q}_{\mathbf{a}_1}\mathbf{n}\|^2$ has a mean of $(n - 1)\psi$ and variance of $2(n - 1)\psi^2$. Again if *n* is not too small we approximate this χ^2 distribution by a normal distribution with the same mean and variance. As in §4A.1 above, if we set a false alarm probability δ and the threshold for this is t_δ standard deviations from the mean, then the required threshold is

$$(n-1)\psi + t_\delta\sqrt{2(n-1)\psi^2} = \psi(n-1+t_\delta\sqrt{2(n-1)}).$$

(We use δ here, rather than β as in §4A.1 above, as this test is at a single point, rather than over a whole scan, so if we want a false alarm rate δ on this set of data, as in §4A.1, we set a threshold t_δ.)

APPENDIX 4B: PEAK POSITIONS ON FIRST SCAN

4B.1 Non-coherent case

The IMP function of the first scan (with no nulls) is given, in the non-coherent case with multiple data frames, by

$$f_0(\boldsymbol{\theta}) = \mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}}\mathbf{R}\mathbf{a}(\boldsymbol{\theta}) \qquad (4.B.1)$$

where the system covariance matrix (in practice this is estimated over a finite number *p* of data frames) is

$$\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^{\mathrm{H}} + \psi\mathbf{I} \qquad (4.B.2)$$

where $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2]$, $\mathbf{P} = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix}$, ψ is the noise variance and $\mathbf{I}$ is the identity matrix of

order n. $\mathbf{a}_k$ and p_k are the PSV and power level of signal k. Putting (4.B.2) into (4.B.1) gives (omitting $\boldsymbol{\theta}$ for clarity)

$$f_0 = \begin{bmatrix} \mathbf{a}^{\mathrm{H}}\mathbf{a}_1 & \mathbf{a}^{\mathrm{H}}\mathbf{a}_2 \end{bmatrix} \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1{}^{\mathrm{H}}\mathbf{a} \\ \mathbf{a}_2{}^{\mathrm{H}}\mathbf{a} \end{bmatrix} + \psi\mathbf{I} = |\gamma_1|^2 \, p_1 + |\gamma_2|^2 \, p_2 + \psi\mathbf{I} \qquad (4.B.3)$$

where $\gamma_j = \mathbf{a}_j{}^{\mathrm{H}}\mathbf{a}$ (j = 1,2).

To find the peak of f_0 we need to set its derivatives with respect to α and ε to zero, i.e. we need

$$\frac{\partial f_0}{\partial \alpha} = 2\,\mathrm{Re}\left(\gamma_1 * \frac{\partial \gamma_1}{\partial \alpha} \right) p_1 + 2\,\mathrm{Re}\left(\gamma_2 * \frac{\partial \gamma_2}{\partial \alpha} \right) p_2 = 0 \qquad (4.B.4)$$

and similarly for the derivative with respect to ε. Now (from (4.1.4) for example)

$$\gamma_j \approx 1 - \frac{2\pi^2}{n} \sum\nolimits_k \left(\mathbf{r}_k{}^{\mathrm{T}} \Delta\mathbf{e}_j \right)^2 \qquad (4.B.5)$$

to lowest order, where (for small target separations $\Delta\alpha = \alpha_2 - \alpha_1$ and $\Delta\varepsilon = \varepsilon_2 - \varepsilon_1$)

$$\Delta\mathbf{e} = \mathbf{e}_\alpha \Delta\alpha + \mathbf{e}_\varepsilon \Delta\varepsilon = \mathbf{E} \begin{bmatrix} \Delta\alpha \\ \Delta\varepsilon \end{bmatrix} \qquad (4.B.6)$$

and $\mathbf{E} = \begin{bmatrix} \mathbf{e}_\alpha & \mathbf{e}_\varepsilon \end{bmatrix}$, the derivatives of the unit direction vector in the immediate target area. Thus we have

$$\gamma_j = 1 - 2\pi^2 \begin{bmatrix} \Delta\alpha_j & \Delta\varepsilon_j \end{bmatrix} \mathbf{E}^{\mathrm{T}} \mathbf{M} \mathbf{E} \begin{bmatrix} \Delta\alpha_j \\ \Delta\varepsilon_j \end{bmatrix} \qquad (4.B.7)$$

using (**3**.3A.6) from Chapter 3, with $\Delta\alpha_j = \alpha - \alpha_j$ and $\Delta\varepsilon_j = \varepsilon - \varepsilon_j$. The derivatives of γ_j are thus given by

$$\frac{\partial \gamma_j}{\partial \alpha} = -4\pi^2 \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{K} \begin{bmatrix} \Delta\alpha_j \\ \Delta\varepsilon_j \end{bmatrix} \text{ and } \frac{\partial \gamma_j}{\partial \varepsilon} = -4\pi^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \mathbf{K} \begin{bmatrix} \Delta\alpha_j \\ \Delta\varepsilon_j \end{bmatrix} \qquad (4.B.8)$$

where $\mathbf{K} = \mathbf{E}^{\mathrm{T}}\mathbf{M}\mathbf{E}$ Then we have, to first order (so, setting $\gamma_j \approx 1$, and noting that γ_j and its derivative are real), using (4.B.8) in (4.B.4),

$$\frac{\partial f_0}{\partial \alpha} = -8\pi^2 \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{K} \begin{bmatrix} \Delta\alpha_1 \\ \Delta\varepsilon_1 \end{bmatrix} p_1 - 8\pi^2 \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{K} \begin{bmatrix} \Delta\alpha_2 \\ \Delta\varepsilon_2 \end{bmatrix} p_2 = 0$$

and

$$\frac{\partial f_0}{\partial \varepsilon} = -8\pi^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \mathbf{K} \begin{bmatrix} \Delta\alpha_1 \\ \Delta\varepsilon_1 \end{bmatrix} p_1 - 8\pi^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \mathbf{K} \begin{bmatrix} \Delta\alpha_2 \\ \Delta\varepsilon_2 \end{bmatrix} p_2 = 0 .$$

Combining these we have (as $\begin{bmatrix} [1 & 0] \\ [0 & 1] \end{bmatrix} = \mathbf{I}$)

$$\mathbf{K} \begin{bmatrix} \Delta\alpha_1 \\ \Delta\varepsilon_1 \end{bmatrix} p_1 + \mathbf{K} \begin{bmatrix} \Delta\alpha_2 \\ \Delta\varepsilon_2 \end{bmatrix} p_2 = \mathbf{0} \quad \text{or} \quad \mathbf{K}\Delta\boldsymbol{\theta}_1 p_1 + \mathbf{K}\Delta\boldsymbol{\theta}_2 p_2 = \mathbf{0} . \qquad (4.B.9)$$

Putting $\Delta\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{pk} - \boldsymbol{\theta}_1$, etc. (as this equation defines the peak position) and rearranging, we have

$$(p_1 + p_2)\mathbf{K}\boldsymbol{\theta}_{pk} = \mathbf{K}\left(\boldsymbol{\theta}_1 p_1 + \boldsymbol{\theta}_2 p_2\right)$$

or

$$\boldsymbol{\theta}_{pk} = \frac{p_1}{(p_1 + p_2)}\boldsymbol{\theta}_1 + \frac{p_2}{(p_1 + p_2)}\boldsymbol{\theta}_2 + \mathbf{k} \qquad (4.B.10)$$

where **k** is in the null space of **K**. However, **K** is of full rank (two) in general so **k** is zero. The result given in (4.B.10) is not surprising, given a lowest order approximation (on the basis that the targets will be quite close, in terms of beamwidth, when at the limit of resolution); it states that the peak position is on the line (in parameter space) between the two targets and is at their centroid, when weighted by the powers of the targets.

4B.2 Coherent case

We take the case of a single data frame with two large signals present and ignore the contribution of noise. The argument follows that of §B1.1 above, with some significant differences.

Let the received data be **y** where $\mathbf{a}_k$ and s_k are the PSV and signal amplitude for signal k. (s_k is complex in general, and its phase is that at the array element position reference point, generally the array centroid, and its amplitude is that which would be received in the array when steered with (normalized) PSV $\mathbf{a}_k$.) Then we have

$$\mathbf{y} = \mathbf{a}_1 s_1 + \mathbf{a}_2 s_2 \qquad (4.B.11)$$

and the first scan function is given by

$$f_0(\boldsymbol{\theta}) = \mathbf{a}(\boldsymbol{\theta})^{\mathrm{H}} \mathbf{y}\mathbf{y}^{\mathrm{H}} \mathbf{a}(\boldsymbol{\theta}) = \mathbf{a}(\boldsymbol{\theta})\left(\left|s_1\right|^2 \mathbf{a}_1 \mathbf{a}_1^{\mathrm{H}} + 2\,\mathrm{Re}(s_1 s_2 * \mathbf{a}_1 \mathbf{a}_2^{\mathrm{H}}) + \left|s_2\right|^2 \mathbf{a}_2 \mathbf{a}_2^{\mathrm{H}}\right)\mathbf{a}(\boldsymbol{\theta})$$

or, dropping $\boldsymbol{\theta}$ for clarity

$$f_0 = |s_1|^2 |\gamma_1|^2 + 2\,\mathrm{Re}(s_1 s_2 * \gamma_1 * \gamma_2) + |s_2|^2 |\gamma_2|^2 \tag{4.B.12}$$

with γ_j as in (4.B.3). This is the equation corresponding to (4.B.3) for the non-correlated case. The equivalent of (4.B.4) is

$$\frac{\partial f_0}{\partial \alpha} = 2\,\mathrm{Re}\left(\gamma_1 * \frac{\partial \gamma_1}{\partial \alpha}\right)|s_1|^2 + 2\,\mathrm{Re}\left(s_1 s_2 * \left(\frac{\partial \gamma_1}{\partial \alpha}^{*} \gamma_2 + \gamma_1 * \frac{\partial \gamma_2}{\partial \alpha}\right)\right) +$$

$$+2\,\mathrm{Re}\left(\gamma_2 * \frac{\partial \gamma_2}{\partial \alpha}\right)|s_2|^2 = 0 \tag{4.B.13}$$

and similarly for the derivative with respect to ε. Using (4.B.8) and approximating γ_j to unity as before we obtain

$$\frac{\partial f_0}{\partial \alpha} = -8\pi^2 \left(\begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{K} \begin{bmatrix} \Delta\alpha_1 \\ \Delta\varepsilon_1 \end{bmatrix} |s_1|^2 + \mathrm{Re}\left(s_1 s_2 *\right) \left(\begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{K} \begin{bmatrix} \Delta\alpha_1 \\ \Delta\varepsilon_1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{K} \begin{bmatrix} \Delta\alpha_2 \\ \Delta\varepsilon_2 \end{bmatrix} \right) + \right.$$

$$\left. + \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{K} \begin{bmatrix} \Delta\alpha_2 \\ \Delta\varepsilon_2 \end{bmatrix} |s_2|^2 \right) = 0$$

where $\Delta\mathbf{e}_j = \mathbf{e}(\boldsymbol{\theta}) - \mathbf{e}(\boldsymbol{\theta}_j) \approx \mathbf{e}_\alpha(\alpha - \alpha_j) + \mathbf{e}_\varepsilon(\varepsilon - \varepsilon_j) = \mathbf{E} \begin{bmatrix} \Delta\alpha_j \\ \Delta\varepsilon_j \end{bmatrix}$ and $\mathbf{K} = \mathbf{E}^{\mathrm{T}}\mathbf{M}\mathbf{E}$. With the equivalent for the derivative of f_0 with respect to ε, we have, combining these,

$$\mathbf{K} \begin{bmatrix} \Delta\alpha_1 \\ \Delta\varepsilon_1 \end{bmatrix} |s_1|^2 + \mathrm{Re}\left(s_1 s_2 *\right) \left(\mathbf{K} \begin{bmatrix} \Delta\alpha_1 \\ \Delta\varepsilon_1 \end{bmatrix} + \mathbf{K} \begin{bmatrix} \Delta\alpha_2 \\ \Delta\varepsilon_2 \end{bmatrix} \right) + \mathbf{K} \begin{bmatrix} \Delta\alpha_2 \\ \Delta\varepsilon_2 \end{bmatrix} |s_2|^2 = \mathbf{0}$$

or

$$S_1 \mathbf{K} \begin{bmatrix} \Delta\alpha_1 \\ \Delta\varepsilon_1 \end{bmatrix} + S_2 \mathbf{K} \begin{bmatrix} \Delta\alpha_2 \\ \Delta\varepsilon_2 \end{bmatrix} = \mathbf{0} \tag{4.B.14}$$

where, after dividing by the noise level, ψ, we define S_1 and S_2 by

$$S_k = \sigma_k + r \text{ with } \sigma_k = |s_2|^2 / \psi \text{ and } r = \mathrm{Re}\left(s_1 s_2 *\right) / \psi . \tag{4.B.15}$$

Putting $\Delta\alpha_1 = \alpha_{\mathrm{pk}} - \alpha_1$, etc., we obtain

$$(S_1 + S_2)\mathbf{K} \begin{bmatrix} \alpha_{\mathrm{pk}} \\ \varepsilon_{\mathrm{pk}} \end{bmatrix} = S_1 \mathbf{K} \begin{bmatrix} \alpha_1 \\ \varepsilon_1 \end{bmatrix} + S_2 \mathbf{K} \begin{bmatrix} \alpha_2 \\ \varepsilon_2 \end{bmatrix}$$

or

$$\begin{bmatrix} \alpha_{\text{pk}} \\ \varepsilon_{\text{pk}} \end{bmatrix} = \frac{S_1}{S_1 + S_2} \begin{bmatrix} \alpha_1 \\ \varepsilon_1 \end{bmatrix} + \frac{S_2}{S_1 + S_2} \begin{bmatrix} \alpha_2 \\ \varepsilon_2 \end{bmatrix} \tag{4.B.16}$$

(taking **K** as non-singular). This is similar in structure to (4.B.10) for the non-correlated case, but the coefficients are significantly different. We can put

$$S_1 + S_2 = \left(|s_1|^2 + |s_2|^2 + 2|s_1||s_2|\cos\phi \right) \Big/ \psi = |s_1 + s_2|^2 \Big/ \psi$$

where ϕ is the phase difference between the signal amplitudes ($\phi = \arg(s_2) - \arg(s_1)$). In particular, if their magnitudes are close and their phases differ by π then $s_1 + s_2$ is near zero and we can get some solutions for the peak position which are far outside the interval between the target positions.

We can also put (4.B.16) in the form

$$\boldsymbol{\theta}_{\text{pk}} = \frac{S_1 \boldsymbol{\theta}_1 + S_2 \boldsymbol{\theta}_2}{S_1 + S_2} \tag{4.B.17}$$

where $\boldsymbol{\theta}_k = \begin{bmatrix} \alpha_k \\ \varepsilon_k \end{bmatrix}$ is the parameter vector for points 1, 2 and pk, represented by k, as in §4B.1.

If we put $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2)/2$ and $\Delta\boldsymbol{\theta} = \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1$, where $\boldsymbol{\theta}_0$ is the midpoint between the two signal positions, and $\Delta\boldsymbol{\theta}$ is their vector separation (and so $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 - \Delta\boldsymbol{\theta}/2$ and $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_0 + \Delta\boldsymbol{\theta}/2$) then we find that the position of the peak is given by

$$\boldsymbol{\theta}_{\text{pk}} = \boldsymbol{\theta}_0 + t\Delta\boldsymbol{\theta}$$

where the factor t is given by

$$t = \frac{S_2 - S_1}{2(S_2 + S_1)} = \frac{|s_2|^2 - |s_1|^2}{2|s_2 + s_1|^2} = \frac{\text{Re}\left((s_2 - s_1)^*(s_2 + s_1)\right)}{(s_2 + s_1)^*(s_2 + s_1)}.$$

We see that the solution for the peak position is on the line joining the points ($\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$). In most cases it will be between the points but this is not necessarily so. We see that if one of the signals is dominant then the solution is close to the position of that signal. For example, if $|s_2| >> |s_1|$ then we have t close to ½ and so $\boldsymbol{\theta}$ is close to $\boldsymbol{\theta}_2$, as expected. However, if the amplitudes are close and the signal amplitudes are in antiphase ($s_2 \approx -$

s_1) then the denominator of t is close to zero and the peak may be anywhere on the line, including at points a long way beyond either signal position. (We note that as $s_2 \to -s_1$ both the numerator and the denominator approach zero, but the numerator does so as, $|\delta s|$ where $s_2 = -s_1 + \delta s$, while the denominator does so as $|\delta s|^2$, so at this point $t \to \infty$.) In fact the extreme cases will arise only rarely – even if the echoes are essentially the same in amplitude, the relative phase will not generally also be near 180º.

However, we note that we can be well outside the points, i.e. many times the length of the interval $\Delta\boldsymbol{\theta}$, but, if $\Delta\boldsymbol{\theta}$ is small enough, the peak could still be within a beamwidth, in which case IMP could still converge. If it is more than a beamwidth then the convergence, if at all, will be to the wrong solution.

APPENDIX 4C: AUXILIARY RESULTS FOR SECOND SCAN PEAK VALUE.

4C.1. Function value at a null point

An expression of the form

$$f(\boldsymbol{\theta}) = \frac{\mathbf{a}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{R}\mathbf{Q}_{\mathbf{a}}\mathbf{a}}{\mathbf{a}^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{a}}$$

where $\mathbf{Q}_{\mathbf{a}}\mathbf{a} = \mathbf{0}$, is clearly indeterminate as it stands, but may in fact have a finite value. By l'Hôpital's Rule we evaluate the function by taking differentials of both numerator and denominator. Thus we can put

$$f(\boldsymbol{\theta}) = \frac{2\,\mathrm{Re}(\mathbf{a}'^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{R}\mathbf{Q}_{\mathbf{a}}\mathbf{a})}{2\,\mathrm{Re}(\mathbf{a}'^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{a})}$$

but we still have the factor $\mathbf{Q}_{\mathbf{a}}\mathbf{a}$ in both numerator and denominator, so we need a second derivative, to obtain

$$f(\boldsymbol{\theta}) = \frac{2\,\mathrm{Re}(\mathbf{a}'^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{R}\mathbf{Q}_{\mathbf{a}}\mathbf{a})}{2\,\mathrm{Re}(\mathbf{a}'^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{a})} = \frac{2\,\mathrm{Re}(\mathbf{a}'^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{R}\mathbf{Q}_{\mathbf{a}}\mathbf{a}' + \mathbf{a}''^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{R}\mathbf{Q}_{\mathbf{a}}\mathbf{a})}{2\,\mathrm{Re}(\mathbf{a}'^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{a}' + \mathbf{a}''^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{a})}$$

or

$$f(\boldsymbol{\theta}) = \frac{\mathbf{a}'^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{R}\mathbf{Q}_{\mathbf{a}}\mathbf{a}'}{\mathbf{a}'^{\mathrm{H}}\mathbf{Q}_{\mathbf{a}}\mathbf{a}'} \qquad (4.C.1)$$

using $\mathbf{Q}_{\mathbf{a}}\mathbf{a} = \mathbf{0}$ and the fact that the non-zero terms are real.

4C.2. Inner products

Let $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ be two points at distances x_1 and x_2 from $\hat{\boldsymbol{\theta}}$ along the line at angle ϕ in parameter space (i.e. at points $\hat{\boldsymbol{\theta}} + x_k \Delta\boldsymbol{\theta}$) then component j of $\mathbf{a}_k$ is given by $a_{kj} = (1/\sqrt{n}) \exp(2\pi i \mathbf{r}_j^{\mathrm{T}} \mathbf{e}(\hat{\boldsymbol{\theta}} + x_k \Delta\boldsymbol{\theta}))$. Also $\mathbf{e}(\hat{\boldsymbol{\theta}} + x_k \Delta\boldsymbol{\theta}) \approx \hat{\mathbf{e}}(\hat{\boldsymbol{\theta}}) + x_k \Delta\theta \mathbf{e}_\phi$ where $\hat{\mathbf{e}} = \mathbf{e}(\hat{\boldsymbol{\theta}})$, $\mathbf{e}_\phi$ is the derivative of $\mathbf{e}$ along the direction ϕ (see below (4.1.5)) and $\Delta\theta$ is the magnitude of $\Delta\boldsymbol{\theta}$. Also let the general point along this line be given by $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + x\Delta\boldsymbol{\theta}$, then the components of $\mathbf{a}(\boldsymbol{\theta})$ are given (dropping $\boldsymbol{\theta}$ or $\hat{\boldsymbol{\theta}}$ as understood) by

$$a_j = (1/\sqrt{n}) \exp 2\pi i \mathbf{r}_j^{\mathrm{T}} (\hat{\mathbf{e}} + x\Delta\theta \mathbf{e}_\phi)$$

$$= \hat{a}_j \exp 2\pi i \mathbf{r}_j^{\mathrm{T}} x\Delta\theta \mathbf{e}_\phi = \hat{a}_j \exp ixq_j \qquad (4.C.2)$$

where $q_j = 2\pi \mathbf{r}_j^{\mathrm{T}} \mathbf{e}_\phi \Delta\theta$. Also $|a_j|^2 = |\hat{a}_j|^2 = 1/n$.

Taking the array element positions to be given relative to the centroid we have (as in (3.3A.6) for example)

$$\sum_j q_j = 2\pi\Delta\theta \sum_j \mathbf{r}_j^{\mathrm{T}} \mathbf{e}_\phi = 0 \qquad (4.C.3)$$

and also we have

$$(1/n) \sum_j q_j^2 = (2\pi\Delta\theta)^2 \sum_j (\mathbf{r}_j^{\mathrm{T}} \mathbf{e}_\phi)^2 / n = 4\pi^2 \mathbf{e}_\phi \mathbf{M} \mathbf{e}_\phi \Delta\theta^2 = \mu. \qquad (4.C.4)$$

Differentiating (4.C.2) with respect to x we have

$$a_j' = iq_j \hat{a}_j \exp ixq_j. \qquad (4.C.5)$$

From (4.1.9) we note that if r is greater than 3 or 4 then μ is small and so is μx^2 if x is of order unity, which will be the case for x (and x_1 and x_2) in most cases, particularly if the targets are of considerably different strengths. Assuming the conditions are such that we have a moderately high resolution improvement factor then we take μ, μx^2 and μx_k^2 etc. to be small and we use this fact to expand the exponentials. With these preliminary results we can easily obtain the inner products required. From (4.C.2), (4.C.3) and (4.C.4)

$$\mathbf{a}_1^{\mathrm{H}} \mathbf{a}_2 = |\hat{a}_j|^2 \sum_j \exp iq_j (x_2 - x_1) \approx (1/n) \sum_j \left(1 + iq_j (x_2 - x_1) - q_j^2 (x_2 - x_1)^2 / 2\right)$$

$$= 1 - \mu (x_2 - x_1)^2 / 2 = 1 - \mu/2 \qquad (4.C.6)$$

as $x_2 - x_1 = 1$ from (4.C.11) below,

and similarly

$$\mathbf{a}_k{}^{\mathrm{H}}\mathbf{a} = 1 - \mu(x - x_k)^2/2 \text{ and } \mathbf{a}_k{}^{\mathrm{H}}\hat{\mathbf{a}} = 1 - \mu x_k{}^2/2 . \tag{4.C.7}$$

Also

$$\mathbf{a}_k{}^{\mathrm{H}}\mathbf{a}' = (1/n)\sum_j iq_j \exp iq_j(x - x_k) \approx (1/n)\sum_j iq_j \left(1 + iq_j(x - x_k) - \ \ldots\right)$$

$$= -\mu(x - x_k) . \tag{4.C.8}$$

and (as $x = 0$ for $\hat{\mathbf{a}}$, replacing x_k)

$$\hat{\mathbf{a}}^{\mathrm{H}}\mathbf{a}' = -\mu x . \tag{4.C.9}$$

From (4.C.5)

$$\mathbf{a}'^{\mathrm{H}}\mathbf{a}' = \sum_j q_j{}^2 \left|\hat{a}_j\right|^2 = \sum_j q_j{}^2/n = \mu . \tag{4.C.10}$$

For the actual points $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ we have

$$\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}} = x_1\Delta\boldsymbol{\theta} = \boldsymbol{\theta}_1 - \frac{S_1\boldsymbol{\theta}_1 + S_2\boldsymbol{\theta}_2}{S_1 + S_2} = \frac{S_2(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)}{S_1 + S_2} = -\frac{S_2\Delta\boldsymbol{\theta}}{S_1 + S_2}$$

and

$$\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}} = x_2\Delta\boldsymbol{\theta} = \boldsymbol{\theta}_2 - \frac{S_1\boldsymbol{\theta}_1 + S_2\boldsymbol{\theta}_2}{S_1 + S_2} = \frac{S_1(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)}{S_1 + S_2} = \frac{S_1\Delta\boldsymbol{\theta}}{S_1 + S_2}$$

so we see that

$$x_1 = -\frac{S_2}{S_1 + S_2},\ x_2 = \frac{S_1}{S_1 + S_2} \text{ and } x_2 - x_1 = 1 . \tag{4.C.11}$$

4C.3 Evaluation of constants A, B and C

Here we evaluate A, B and C, given in (4.2.25), with

$$x_1 = -S_2/(S_1 + S_2) \quad x_2 = S_1/(S_1 + S_2) \quad S_k = \sigma_k + r \quad r = \sqrt{\sigma_1\sigma_2}\cos\phi \tag{4.C.12}$$

and ϕ is the difference in phase between the signal amplitudes ($\phi = \arg(s_2) - \arg(s_1)$). Then we have

$$A(S_1 + S_2)^2 = \sigma_1 S_2{}^2 - 2rS_2S_1 + \sigma_2 S_1{}^2 = (S_1 - r)S_2{}^2 - 2rS_1S_2 + (S_2 - r)S_1{}^2$$

$$= S_1S_2(S_1 + S_2) - r(S_1{}^2 + 2S_1S_2 + S_2{}^2) .$$

or

$$A(S_1 + S_2) = S_1S_2 - r(S_1 + S_2)\,. \tag{4.C.13}$$

Now, using (4.C.12),

$$S_1S_2 - r(S_1 + S_2) = (\sigma_1 + r)(\sigma_2 + r) - r(\sigma_1 + \sigma_2 + 2r)$$

$$= \sigma_1\sigma_2 - r^2 = \sigma_1\sigma_2(1 - \cos^2\phi) = \sigma_1\sigma_2 \sin^2\phi \tag{4.C.14}$$

and

$$S_1 + S_2 = \sigma_1 + \sigma_2 + 2\sqrt{\sigma_1\sigma_2}\cos\phi\,. \tag{4.C.15}$$

Thus A is given by

$$A = \frac{\sigma_1\sigma_2 \sin^2\phi}{\sigma_1 + \sigma_2 + 2\sqrt{\sigma_1\sigma_2}\cos\phi}\,, \tag{4.C.16}$$

though the denominator could be put in the alternative form $S_1 + S_2$ if convenient.

B is given by

$$\frac{2}{\mu}B(S_1 + S_2)^3 = -(S_1 - r){S_2}^3 + r(-S_2{S_1}^2 + S_1{S_2}^2) + (S_2 - r){S_1}^3$$

$$= S_1S_2({S_1}^2 - {S_2}^2) + r({S_2}^3 - {S_1}^3 + S_1{S_2}^2 - S_2{S_1}^2) = ({S_1}^2 - {S_2}^2)\big(S_1S_2 - r(S_1 + S_2)\big)\,,$$

so

$$\frac{2}{\mu}\frac{B(S_1 + S_2)^3}{({S_1}^2 - {S_2}^2)} = S_1S_2 - r(S_1 + S_2)\,.$$

Cancelling a factor of $(S_1 + S_2)$ and using (4.C.14) and (4.1.12) we have

$$B = \frac{\mu}{2}\frac{(S_1 - S_2)\sigma_1\sigma_2 \sin^2\phi}{(S_1 + S_2)^2} = \frac{\mu}{2}\frac{(\sigma_1 - \sigma_2)\sigma_1\sigma_2 \sin^2\phi}{(\sigma_1 + \sigma_2 + \sqrt{\sigma_1\sigma_2}\cos\phi)^2}\,. \tag{4.C.17}$$

C is given by

$$\left(\frac{2}{\mu}\right)^2 C(S_1 + S_2)^4 = (S_1 - r){S_2}^4 + 2r{S_2}^2{S_1}^2 + (S_2 - r){S_1}^4$$

$$= S_1S_2({S_1}^3 + {S_2}^3) - r({S_1}^2 - {S_2}^2)^2\,.$$

This can be manipulated to give

$$\left(\frac{2}{\mu}\right)^2 C(S_1+S_2)^4 = (S_1-S_2)^2(S_1+S_2)\big(S_1S_2 - r(S_1+S_2)\big) + S_1^{\,2}S_2^{\,2}(S_1+S_2)$$

and using (4.C.14) again we have

$$C = \left(\frac{\mu}{2}\right)^2 \left(\frac{(S_1-S_2)^2\sigma_1\sigma_2\sin^2\phi}{(S_1+S_2)^3} + \frac{S_1^{\,2}S_2^{\,2}}{(S_1+S_2)^3} \right)$$

or

$$C = \left(\frac{\mu}{2}\right)^2 \left(\frac{(\sigma_1-\sigma_2)^2\sigma_1\sigma_2\sin^2\phi}{(\sigma_1+\sigma_2+2\sqrt{\sigma_1\sigma_2}\cos\phi)^3} + \frac{S_1^{\,2}S_2^{\,2}}{(S_1+S_2)^3} \right). \qquad (4.C.18)$$

Chapter 5: Conclusions

5.1 RESULTS ON ACCURACY

The results on accuracy are given concisely in the summary of Chapter 3, §3.5.2, which we recapitulate and enlarge on here, in the list below. We emphasize that these are in general (apart from early results in Chapter 3, where the method used was being established in simpler cases) for the case of two angle parameters, and require a 2D or 3D array. In fact the theory is the same for the 3D array as for the apparently simpler case of the 2D array (which is generally more likely to be found in practice), so all the theoretical results are applicable to both. This is in contrast with most of the previous work on superresolution accuracy, which is for a single angle parameter and a linear array (or in some cases a planar array, but with the sources in the plane of the array, so still being only a single parameter case).

(1) A general observation is that the array moment matrix **M** (the 3×3 matrix of second moments of the matrix elements about the mean, for each pair of element coordinates – x^2, xy, xz etc.) is a very significant factor in determining the accuracy (and indeed also the resolution). This, at least in the 3D form (rather than the scalar form derived in the linear array case), is not a quantity seen in the literature reviewed, but is a relatively simple and clearly important factor.

(2) Although superresolution systems can give parameter estimates of multiple signals, the accuracy in the single signal case may still be of interest. This case is the same for MUSIC and IMP which both reduce to simple beamforming. (The results down to (5) below are for the single signal case.) The analysis gives the parameter estimates covariance matrix as quite a simple expression ((3.3.17)), proportional to the inverse of a scalar function of **M**. (Thus the larger **M** the lower the variance, and the higher the accuracy.) The parameter error variances are proportional to the variance of the system phase errors, but are independent of the system amplitude errors (at least up to any likely practical values). This interesting result (the independence from amplitude errors, confirmed by simulations) was not seen in the literature reviewed, when generally two signals were taken to be present. However this result does not hold when two (or more) signals are present.

(3) Most practical arrays are of similar elements, oriented in parallel (an EPP array) and this work has generally taken this to be the case (as apparently has generally been so in the literature reviewed). However, an investigation into the case where this does not

hold produced a closely related but slightly more complicated expression ((3.3.26) with (3.3.23) and (3.3.25)), requiring two moment matrices, one $\mathbf{M}_1$ with weighting of the element positions by their power gains in the direction of the signal (given in (3.2.24)), and another $\mathbf{M}_2$ with the element positions weighted by the square of these power gains (see below (3.3.25)). As for the other results on accuracy, this expression was confirmed by simulation. (It was checked, in the course of the simulations, that using only one of these matrices in the theory does not give a result matching the simulation results.)

(4) As the parameter estimation error variances are simply proportional to the system error variance, it is possible to define an error sensitivity as the ratio of the standard deviation (s.d., the square root of the variance for unbiased distributions) of the parameter estimates to the s.d. of the system phase errors. The theory enables this sensitivity to be presented in the form of contour plots over a two dimensional angle region, which could be useful when designing or choosing an array for a given application.

(5) It was possible to derive two remarkably simple rules of thumb for the array sensitivity, or ratio of parameter estimate s.d. to the s.d. of the system errors; one gives the sensitivity as approximately $4/n\sqrt{n}$, for a *circular* array of n elements (with half wavelength element spacing), over a substantial angular region (see §3.3.7). (Thus for a 12 element circular array, with 5° phase errors, the s.d. of the estimates of the (single) target azimuth and elevation are about 0.5°, within a factor of 2 over a considerable range of angles.) The other is for more general planar arrays of aperture k wavelengths giving sensitivity of about 1/2k over most of angle space.

(6) For the multiple target case, an elegant expression for the errors (but not their statistics) in the parameter estimates was derived ((3.4.25) with (3.4.26)). However it is difficult to obtain the error variances in this case. On the basis that the case of two targets (particularly two close targets) is of considerable interest, a more complex expression for the error covariance matrix than for the single source was derived though having some structural similarity ((3.4.44)). This shows the parameter errors depending on both the phase and amplitude system errors. If these errors have the same variance then the parameter error covariance matrix becomes much simpler ((3.4.45) with (3.4.36)). However, in both cases we require a modified form $\tilde{\mathbf{M}}$ (in (**3**.4.34)) of the element covariance matrix **M**.

(7) The error analyses have all been for the case of large signals, where errors limit

performance, rather than noise, which was ignored. However, it was found that the parameter estimates covariance matrix was identical in structure to the Cramér-Rao bound (CRB) matrix, enabling an equivalence to be made between the system errors and a corresponding noise level. (This is the case when only one signal is present and also in the two signal case strictly only when the phase and amplitude errors have the same variance.) This interesting result indicates at what signal to noise levels system errors, rather than noise, limit the accuracy obtainable. With this comparison we see that the CRB can be given in terms of the array moment matrix (see (3.4.52) and (3.2.35)), which is a result not clearly shown in other literature.

(8) The theoretical accuracy of IMP in the two signal case is more difficult to derive than for MUSIC, as it uses a different basic scan function, so was investigated only in simulation. In fact this showed that the accuracy of IMP is essentially identical to that of MUSIC in simulation and to the theoretical expressions for MUSIC. As both methods approach the CRB in accuracy when limited by noise, so have similar accuracy in this case, perhaps the result that they have similar accuracy when limited by system errors is not surprising.

(9) Finally, using results required for the accuracy study, a good approximate expression for the beamwidth of a general (irregular 2D or 3D) EPP array was obtained. This may be of interest in itself, but is also of use in defining resolution performance, in relating the target separation at the limit of resolution to the beamwidth.

5.2 RESULTS ON RESOLUTION

The starting point for the resolution study of IMP, in the non-coherent target case, was the paper by Speirs *et al* [41], which suggested that resolution failed when the weaker signal was reduced to the threshold of detection by the null placed on the first signal. This seemed a realistic basis, but the paper did not relate the resolution performance to the array description. This was the initial result of Chapter 4, giving a simple expression ((4.1.6) or (4.1.7)) again using the array moment matrix **M**. The angular resolution limits were found to be inversely proportional to the square root of a scalar function of **M**, so that if the array is scaled by some factor k, then **M** scales as k^2 (containing the second moments) and the resolution as $1/k$ – i.e. the resolution limit is inversely proportional to array size. It is also found to be inversely proportional to the square root of the strength of the weaker signal (its signal to noise power ratio, or to the ratio of the weaker signal power as received to the minimum level for detection – the minimum

detectable signal power). This led to a very simple expression ((4.1.10)) for the resolution improvement factor *r*, defined as the ratio of the array beamwidth to the minimum angular separation for resolution.

These expressions for the resolution limit or the resolution improvement factor show resolution improving indefinitely with the strength of the weaker signal. In practice, of course, this cannot be the case, and at some point errors will limit the performance. In fact errors will lead to the null applied to the large signal being reduced to a finite level (rather than, in principle, to zero) and so to detect a second signal (avoiding false alarms) a suitable raised threshold will have to be set. This leads to a modified result ((4.1.14), (4.1.15) or (4.1.17)) involving the ratio of the two signal powers.

In practice, of course, the peak of the first IMP scan, which is where the null is set for the second scan, searching for the second signal, is not precisely at the first signal position. An expression was found for this peak position ((4.1.18)) and used to refine the initial result ((4.1.23)). This is a more satisfactory expression, using both signal powers, while the first result corresponded to the case of a very large (effectively infinite) first signal. However, the actual figure is not changed greatly, being reduced by a factor of $\sqrt{2}$ at the most, when the targets are equal in strength.

The study so far corresponds to the case of non-coherent signals. The case of coherent targets, including the case of resolving two targets using a single pulse, as in the radar case, was investigated in §4.2. Initially we considered only the limiting condition for the detection of a second target (using the ML function, rather than the second scan IMP function). This is not quite the same as resolution, in the sense of forming an estimate of its parameters. In fact it gave a result ((4.2.18)) for the case of echoes *in quadrature* (i.e. with the arguments of their complex echo amplitudes differing by $\pi/2$ radians) which was the same as for the coherent case ((4.1.23)), interestingly. However, in the in-phase or antiphase cases (phase differences of 0 or π) – the signals being 'relatively' real – there was effectively no residual signal component to the ML function, indicating that the single target model satisfied the condition of maximizing the likelihood function, at least at the level of approximation used, hitherto successfully.

Finally, an analysis of the second IMP scan, finding an analytic expression for the peak value, was attempted in §4.2.2 Again, for the in-phase and antiphase cases the 'peak' signal value was found to be zero, with a realistic non-zero level for the quadrature case. (In general the peak is proportional to $\sin^2\phi$, where ϕ is the phase difference between the

complex signal amplitudes, so there is a non-zero peak in general except at the exact in-phase or antiphase conditions, though not necessarily exceeding the threshold.) For this case the quadrature condition gives essentially the same result as before ((4.2.37) and (4.2.38)) and also the same as the detection condition.

In the course of determining a threshold limit it was necessary to provide a threshold value for the IMP scan function. Rather than assuming some heuristically chosen figure, a theoretical (slightly approximate) figure was found based on a chosen false alarm rate, which may be useful as a basis for practical IMP implementations. Similarly a (different) threshold for the detection decision, using the ML function, was also determined. Although normal IMP does not perform this detection test, it could be incorporated at the end of each scan (and after the subsequent tweaking) to decide whether another signal is present, before carrying out a further scan.

5.3 GENERAL CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

A general conclusion to this work is that some useful progress has been made on the subjects of both accuracy and resolution of superresolution systems, in this case MUSIC and IMP specifically, but the work has also been less successful at some points and a number of loose ends remain.

Firstly, the accuracy of IMP as limited by system errors has not been established theoretically. This is disappointing, rather than disastrous, as it is clear that its performance is close to that of MUSIC, the theoretical performance of which has been found.

Secondly, the coherent signal, or single pulse, resolution performance of IMP has not been found for the 'real' target case, where the signal echoes are in phase or in antiphase. It might be suggested that this is because the single target model satisfies the condition of maximizing the likelihood in this 'real' case, but if we consider the case of well separated targets this is not likely to be true. However it may be very nearly true for two very close targets, and the failure is because the approximations used are of too low order in this case, though these approximations worked well in the accuracy study. It would be of interest to determine the actual dependence of the separation on the array element positions and the signal power levels, which may require a fourth order approximation. This would help to define a cross-over level at which the 'real' ($\phi = 0$ or π) case differed from the more general case, centred on the quadrature condition ($\phi = \pm\pi/2$). It may be the case that a different approach would be more fruitful, such as that

of Manikas *et al* [26, 27] based on differential geometry.

Thirdly, although the resolution of MUSIC has been investigated previously, with and without errors, this could be revisited in the approach taken here, with the aim of expressing it in a form using the array moment matrix, and also in simple (perhaps 'rule of thumb') form, even if with some loss of precision. Furthermore, there needs to be a further development of the resolution measure, to give the probability of resolution as a function of signal separation, rather than a simple threshold of separation above or below which the targets are or are not resolved. This has already been done for MUSIC (for the case without errors) [25, 54], albeit giving rise to very complex expressions, and simpler forms would be of interest.

Finally, noting that the approach used has been very successful in the application to accuracy estimation of MUSIC with errors (and also beamwidth estimation for irregular arrays), the work done and techniques used could well be a basis for further work in the areas highlighted above. We note that the aim of finding simple expressions in some cases, and simple rules of thumb), have been achieved, in the areas where results have been achieved at all.

REFERENCES

1. J. F. Böhme, *On parametric methods for array processing*, Signal Processing II: Theories and Applications, p673, ed. H. W. Schlussler, Elsevier, 1983.

2. J. F. Böhme, *Source-parameter estimation by approximate maximum likelihood and non-linear regression*, IEEE J. Oceanic Eng. **OE-10**, 206-212, July 1985.

3. J. P. Burg, *Maximum entropy spectral analysis*, Presented at 37th Annual Meeting, Society of Exploration Physicists, Oklahoma City, OK, 1967

4. D. H. Brandwood, *A complex gradient operator and its application in adaptive array theory*, IEE Proc **130** Pts F and H, 11 – 16, 1983

5. D. H. Brandwood, *High resolution DF using adaptive array techniques*. Proc. Military Microwaves 1986

6. D. H. Brandwood, *Noise space projection: MUSIC without eigenvectors*, IEE Proc. Pt H, **134**, 303-309, 1987

7. D. H. Brandwood, *Fourier Transforms in Radar and Signal Processing*, Artech House, Boston and London, 2003

8. Y. Bresler and A. Makovski, *Exact ML estimation of superimposed exponential signals in noise*, **ASSP-34,** 1081-1089, Oct. 1986 BM86

9. J. Capon, *High resolution frequency-wavenumber resolution spectrum analysis,* Proc. IEEE **57**, 1408-1418, 1969

10. M. Cederval and R. L. Moses, *Efficient Maximum Likelihood DOA estimation for signals with known waveforms in the presence of multipath*, IEEE Trans. Sig. Proc. **SP-45**, March, 808-811, 1997 CM97

11. I. J. Clarke, *The IMP algorithm*, Advances in Spectrum Analysis and Array Processing, ed. Haykin, Chapter 2

12. A. P. Dempster, N. M. Laird and D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm,* Ann. Roy. Stat. Soc., 1-38, Jan 1977

13. D. R. Farrier, D. J. Jeffries, R. Mardani, *Theoretical performance prediction of the MUSIC algorithm*, IEE Proc. **135**, Part F 1988, 216-224 FJM88

14. M. Feder and E. Weinstein, *Parameter estimation of superimposed signals using the EM algorithm,* IEEE Trans. **ASSP-36**, 477-489, 1988

15. B. Friedlander, *A sensitivity analysis of the MUSIC algorithm*, IEEE Trans. Acoustics, Speech, Sig. Proc. **ASSP-38**, 1740-1751, Oct 1990 F90

16. A. G. Jaffer, *Maximum Likelihood direction finding of stochastic sources: a separable solution* ICASSP 88: 1988 International Conference on Acoustics, Speech,

and Signal Processing, **5**, 2893-2896, 1988

17. D. J. Jeffries and D. R. Farrier, *Asymptotic results for eigenvector methods*, IEE Proc.**132**, Part F, 589-594, 1985 JF85

18. N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Univariate Distributions – 1*, Houghton Miflin, Boston, USA, 1970

19. G. F. Hatke and K. W. Forsythe, *A class of polynomial rooting algorithms for joint azimuth-elevation estimation using multidimensional arrays*, Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers 694-699, **1**, 1995. HF95

20. A. Kangas, P. Stoica and T. Soderstrom, *Finite sample and modelling error effects on ESPRIT and MUSIC direction estimators*, IEE Proc. Radar,Sonar and Navigation **141**, 249-255, Oct 1994 KSS94

21. M. Kaveh and A. J. Barabell, *The statistical performance of the MUSIC and Minimum-Norm algorithms in resolving plane waves in noise*, IEEE Trans. Acoustics, Speech, Sig. Proc. **ASSP-34**, 331-341, April 1986 KB86

22. R. Kumaresan and D.W. Tufts, *Estimating the angles of arrival of multiple plane* waves*,* IEEE Trans. 1983, **AES-19**, 134-139.

23. H. B. Lee and J. Jaffner, *Compact CRB expansion for an extended signal model*, Proc. IEEE **SP-42**, 2868-2870 LJ94

24. H. B. Lee and M. S. Wengrovitz, *Resolution threshold of beamspace MUSIC for two closely spaced emitters*, IEEE Trans. Ac. Speech, Sig. Proc. **ASSP-38**, pp1545-1559, Sept 1990 LW90

25. H. B. Lee and M. S. Wengrovitz, *Theoretical resolution-threshold curve for the MUSIC algorithm*, 1991 International Conference on Acoustics, Speech and Signal Processing, 3313-3316, 1991 LW91

26. H. K. Karimi, and A. Manikas, *Manifold of a planar array and its effects on the accuracy of direction finding systems*, IEE Proc. Radar, Sonar, Navigation **143**, 349-357, 1996 KM96

27. A. Manikas, A. Alexiou and H. K. Karimi, *Comparison of the ultimate direction-finding capabilities of a number of planar array geometries*, IEE Proceedings-Radar, Sonar and Navigation, **144**, 321-329, Dec. 1997 MAK97

28. A. Manikas, T. Ratnarajah and J. Lee, *Evaluation of superresolution array techniques as applied to coherent sources*, International Journal Electronics, **82**, 77-105, 1997. MRL97

29. Manikas A N and Turner L F, *Adaptive signal parameter estimation and*

classification technique, IEE Proc. Part F, **138**, 267-277, 1991.

30. K.V. Mardia, J.T. Kent and J. M. Bibby, *Multivariate Analysis*, Academic Press, London, 1979.

31. J. L. Mather, *The Incremental Multi-Parameter algorithm,* Proc. 4th Asilomar Conf., 368-372, Nov. 1990.

32. M. I. Miller and D. R. Fuhrmann *Maximum likelihood narrow-band direction finding and the EM algorithm*, IEEE Trans **ASSP-38**, 1560-1577, 1990

33. S. Nemirovsky and M. A. Doron, *Sensitivity of MUSIC and Root-MUSIC to gain cal. errors of 2D arbitrary array configuration*, 2004 IEEE Sensor Array and Multichannel Signal Processing Workshop. ND04

34. U. Nickel, *Superresolution using an active antenna array,* IEE Proc. Int. Conf. RADAR-82, p87, 1982.

35. B. Ottersten, M. Viberg, P. Stoica and A. Nehorai, *Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing*, Radar Array Processing, Springer-Verlag (Series on Information Sciences, Vol. **25**, 95-151, 1993 OVSN93

36. V. F. Pisarenko, *The retrieval of harmonics from covariance,* Geophys. J Roy. Astron. Soc, **33**, 347-366, 1973

37. S. S. Reddi, *Multiple source location: a digital approach*, IEEE Trans. **AES-15**, 95-105, 1979

38. A. Satish and R. L. Kashyap, *Maximum Likelihood estimation and Cramér-Rao bounds for direction of arrival parameters of a large sensor array*, **AP-44**, 478-491, April 1996. SK96

39(a). R. O. Schmidt, *A signal subspace approach to multiple emitter location and spectral estimation*, PhD dissertation, Stanford University, California, 1979

39(b). R. O. Schmidt, *Multiple emitter location and signal parameter estimation*, Proc. RADC Spectral Estimation Workshop, 243-248, 1979

39(c). R. O. Schmidt, *Multiple emitter location and signal parameter estimation,* IEEE Trans 1986, **AP-34**, pp276-280 (reprint of 1979 paper)

40. L. P. Seidman, *Bearing estimation error with a linear array*, IEEE Trans. Audio and. Electroacoustics, **AU-19**, 147-157, June 1971 S71

41. C. A. Speirs, J. L. Mather, I. J. Clarke and H. D. Rees, *A fundamental bound for super-resolution performance*, NATO Symposium on High Resolution Radar Techniques, Granada, Spain, 22-24 March 1999 SMCR99

42. P. Stoica and A. Nehorai, *MUSIC, Maximum Likelihood and the Cramer-Rao*

bound, IEEE Proc. **ASSP 17**, 720-741, May 1989. SN89

43. P. Stoica and A. Nehorai, *Performance study of conditional and unconditional direction-of-arrival estimation*, **ASSP-38**, 1783-1795, Oct. 1990 SN90

44. P. Stoica P and K. C. Sharman, *Maximum likelihood methods for direction-of-arrival estimation,* IEEE Trans **ASSP-38**, 1132-1143, July 1990.

45. Su Weimin, Ni Jinlin, Liu Guosui and Zhang Guangyi, *The effect of channel mismatch on the spatial spectrum and resolution ability of the MUSIC algorithm*, 1996 CIE International Conference of Radar Proceedings, 96-99, 1996 SNLZ96(1)

46. Su Weimin, Ni Jinlin, Liu Guosui and Zhang Guangyi, *A performance analysis of the MUSIC algorithm in the presence of channel mismatch*, ICSP '96. 1996 3rd International Conf. on Signal Processing Proceedings, 221-224 vol.1, 1996 SNLZ96(2)

47. Su Weimin, Gu Hong, Ni Jinlin, Liu Guosui and Zhang Guangyi, *A performance analysis of the MUSIC Algorithm in the presence of amplitude and phase errors*, National Aerospace and Electronics Conf. 2000, Proc IEEE 2000, 653-660 SGNLZ00

48. A. L. Swindlehurst and T. Kailath, *A performance analysis of subspace methods in the presence of model errors, Part 1: The MUSIC Algorithm*, IEEE Trans. Sig.Proc. **SP-40**, 1758-1774, July, 1992 SK92

49. M. Viberg, B. Ottersten and T. Kailath, *Direction-of-arrival estimation and detection using weighted subspace fitting,* Proc. 3rd Asilomar Conf. 604-608, 1989.

50. M. Viberg, B. Ottersten and T. Kailath, *Detection and estimation and detection in sensor arrays using weighted subspace fitting,* IEEE Trans **ASSP-39**, 2436-2449, 1991.

51. M. Viberg, B. Ottersten and A. Nehorai, *Performance analysis of direction finding with large arrays and finite data*, IEEE Proc. **SP-43**, 469-477, Feb. 1995. VON95

52. A. J. Weiss, B. Friedlander, *Effects of Modeling errors on the resolution of the MUSIC algorithm*, IEEE Trans. Sig.Proc. **SP-42**, 720-740, June, 1994. WF94

53. S. F. Yau and Y. Bresler, *A compact Cramér-Rao bound expression for parametric estimation of superimposed signals*, IEEE Proc. **SP-40**, 1226-1230, May 1992 YB92

54. Q. T. Zhang, *Probability of resolution of the MUSIC algorithm*, Trans. Sig.Proc. **SP-43**, 720-740, April, 1995. Z95

55. C. Zhou, F. Haber, D. L. Jaggard, *A resolution measure for the MUSIC algorithm*, AP-S International Symposium 1989. 1989 International Symposium Digest: Antennas and Propagation, 416-19, vol.1, 1989. ZHJ89

56. I. Ziskind and M. Wax, *Maximum likelihood localization of multiple sources by alternating Projection,* IEEE Trans **ASSP-36**, 1553-1560, 1988.