MLE AND RBF FOR AOA ESTIMATION IN A MULTIPATH ENVIRONMENT

By

TITUS KWOK-YEUNG LO

B.A.Sc. (The University of British Columbia)

M.Eng. (McMaster University)

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Ph.D.

McMaster University

Dec. 1995

It is a subtle virtue for one to be constantly aware of the models.

— Lao Tzu (circa 600 BC)

AOA ESTIMATION IN A MULTIPATH ENVIRONMENT

PH.D. (1995) MCMASTER UNIVERSITY

(Electrical and Computer Engineering) Hamilton, Ontario

TITLE: **MLE and RBF for AOA Estimation in A Multipath Environment**

AUTHOR: Titus Kwok-Yeung Lo

B.A.Sc. (The University of British Columbia)

M.Eng. (McMaster University)

SUPERVISOR(S): Dr. John Litva

Professor, Department of Electrical and Computer Engineering

NUMBER OF PAGES: xvii, 116

ABSTRACT

The problem of estimation of angle-of-arrival (AOA) in multipath environments is addressed in this thesis. In particular, two new estimation techniques are developed. The first technique is based on the maximum likelihood estimation (MLE). This algorithm is unique in that a highly deterministic multipath signal model is used when formulating the likelihood function, which is then maximised with respect to the AOA. The deterministic multipath signal model that has been developed to describe the physics underlying the propagation of signals from a signal radio source to a receiver is much more complete than the general AOA model commonly used in other maximum likelihood formulations. This model makes use of the geometrical information and a priori knowledge of a number of physical parameters. By using the deterministic multipath signal model with the MLE estimator, one is essentially making more information available to the estimation process. The net result is that the estimator's performance can be greatly enhanced. The Cramer-Rao bounds that apply specifically to this model have been derived to provide a performance measure for the mean-squared errors (MSE) in the estimated AOAs.

Although the MLE method is optimum in a statistical sense, the computational load of the nonlinear optimisation procedure inherently required by the MLE method is too heavy for real-time processing. Accordingly, we propose a novel approach to the AOA estimation problem, which is based on the use of an associative memory. The functionality of an associative memory is identical to that of the inverse mapping network. This provides a more comprehensive explanation for the rationale of exploiting

the inverse mapping concept in the AOA estimation problem. In particular, the AOA problem is considered as a mapping from the space of AOA to the space of the sensor output. A nonlinear associative memory is used to form the inverse mapping from the space of sensor output to the space of AOA and this memory is realised using the generalised radial basis function (RBF) neural network. In the actual implementation of the RBF network for AOA estimation, the second order statistics of the signals are used as the input vector of the network. The use of second order statistics eliminates the need to deal with the initial phase of the signal. Furthermore, it is suitable for the application to the minimum redundant array. The RBF network is much more efficient in terms of computation than the MLE algorithm. This makes the RBF network attractive for real-time implementation.

Simulations are carried out to understand the efficiency of the RBF neural network approach. The learning and estimation performance is inversely proportional to the number of learning samples and the number of hidden units. At relatively low SNR, the estimation performance of the RBF network becomes insensitive to both the number of learning samples and the number of hidden units. The estimation performance of both the MLE technique and the RBF network is also evaluated as functions of the number of snapshots and SNR. The performance of the MLE algorithm is consistent with the Cramer-Rao bound. The MLE method is more efficient in terms of estimation than a RBF network, provided that the search resolution used in the MLE method is sufficiently high. For equivalent computational complexity, the RBF network gives much better performance than the MLE method. In terms of estimation, with the same computational complexity, the MSE produced by the RBF network is much less than that produced by the MLE method. In summary, for the same performance, the computational complexity required by MLE method is much higher than that required by the RBF network. It follows that the advantage to be gained by using a RBF network for AOA estimation is a considerable reduction in computational complexity. Finally, both the MLE technique and the RBF network

are validated using real data, which were collected using a 32-element sampled aperture antenna. The results obtained using the RBF network are very similar to those obtained using the MLE method.

ACKNOWLEDGEMENT

The author wishes to express his sincere gratitude to Dr. John Litva for his constant encouragement, continued assistance and expert guidance, and supervision throughout the course of this work.

He would also like to thank Drs. P. Yip and S. Haykin, the members of his supervisory committee, for their useful comments and invaluable suggestions.

He is grateful to the staff and students of CRL who have helped him in many ways.

Last, but not least, he would like to express his deep appreciation to his family, including his parents, his wife and daughters, for their support, understanding, encouragement, patience, and love during his study. The smiles from Janice, Anita, and Emily constantly inspire him in both happy and difficult times.

GLOSSARY

Conventions

1. Column vectors are denoted by lower case, bold face letters.
2. Matrices are denoted by upper case, bold face letters.
3. A hat on a variable denotes the estimate of that variable.

List of Symbols

$\boldsymbol{\Phi}$	key matrix
ψ	grazing angle
λ	radio wavelength
λ_m	the m^{th} singular value of a matrix
$\rho e^{j\varphi}$	complex reflection coefficient
σ	width of a basis function
σ_c	conductivity
σ_H	rms wave height
σ_n^2	noise power

θ^d	angle-of-arrival of the direct signal
θ^i	angle-of-arrival of the indirect signal
$\mathcal{C}^k$	k-dimensional complex space
$E[\cdot]$	expected value
h	source height
$\boldsymbol{I}$	identity matrix
K	number of array elements
L	number of units in the output layer
M	number of training points for RBF or number of searching points for MLE
M^d	number of discrete values of θ^d
M^i	number of discrete values of θ^i
N	number of centres in the RBF network
$\Re[\cdot]$	real value
$\Re^n$	n-dimensional real space
V^d	interval at which the discrete values of θ^d are taken
V^i	interval at which the discrete values of θ^i are taken
$\boldsymbol{W}$	weight matrix
z_k	k^{th} antenna's height

Abbreviations

AOA	angle-of-arrival
ASIC	application specific integrated circuits
BW	beamwidth
MLE	maximum likelihood estimation
MSE	mean squared error
RBF	radial basis function
SNR	signal-to-noise ratio
VLSI	very large scale integration

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Background

In many microwave system applications, such as navigation, air traffic control (ATC) surveillance, and remote sensing, determining the angles-of-arrival (AOA) of incoming signals is a significant problem.

In a microwave landing system (MLS), accurate angular measurement of the aeroplanes in the vicinity of the landing path is critical to air traffic control. However, multipath has been found to be a major measurement error source in most MLS's [1]. Fig. 1.1 shows a typical multipath environment in an MLS. The International Civil Aviation Oganization (ICAO) has identified that novel array signal processing techniques could be used to overcome the difficulty due to multipath [2]. In shipborne electronic passive surveillance, one of the functions required to be performed by the electronic system is to estimate the elevation of radio sources located above a sea surface, as depicted in Fig. 1.2. This problem is also closely related to the low-angle radar tracking problem [3]. The difference is that surveillance is a passive operation where the surveillance system operates only in a receiving mode, whereas radar tracking is an active operation where the radar system transmits and receives

signals. Both electronic passive surveillance and low-angle radar tracking are traditionally military applications where low-elevation flying objects such as missles are to be detected and tracked. However, they also find many civilian applications. An example is the detection and tracking of small low flying planes used by smugglers for illegal drug trafficking along coast lines.

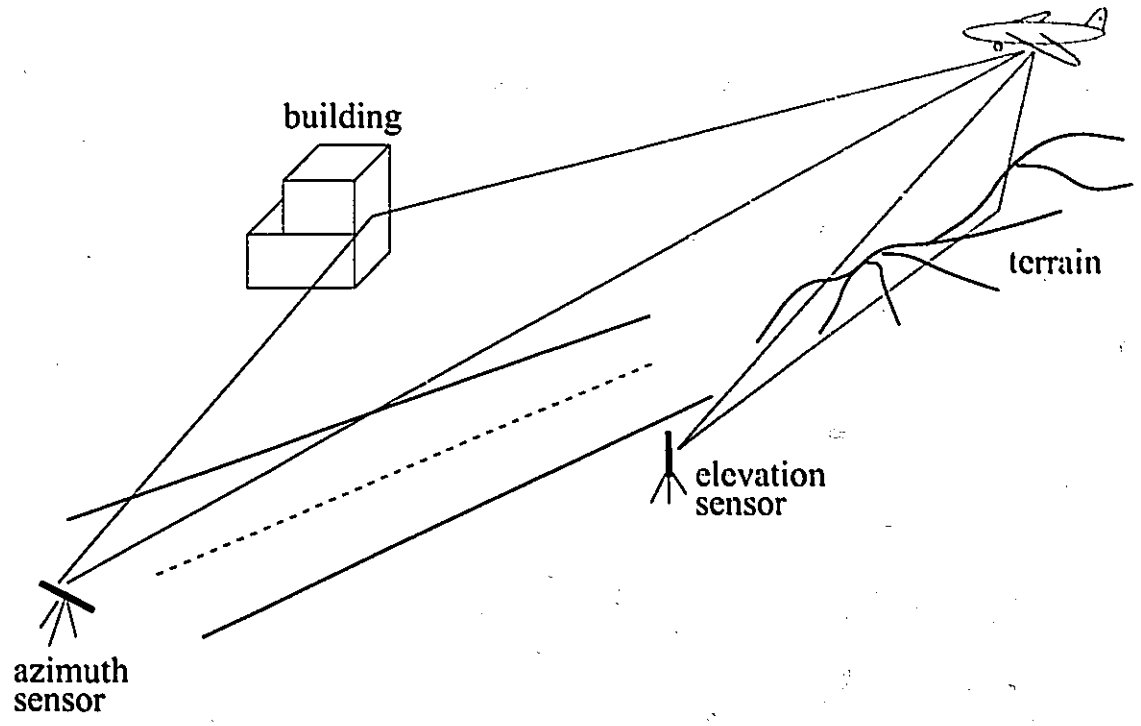

Figure 1.1: A typical multipath environment in a microwave landing system.

The AOA estimation in the above applications is primarily a multipath problem. The propagation of electromagnetic waves takes place via direct and indirect paths. When a plane wave propagates from a source in close proximity to a surface, the received signal consists of two components. The first component propagates along the direct path between the signal source and the receiving antenna, while the other arrives via reflection from the sea surface. The surface-reflected signal is commonly called the multipath signal, which usually consists of two components, a coherent

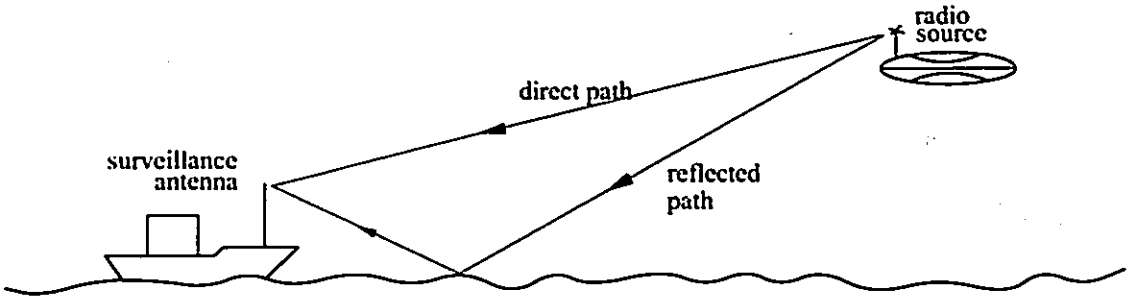

Figure 1.2: Typical situation in shipborne electronic surveillance, where the electronic system is required to estimate the elevation of radio sources located above a sea surface.

(indirect) component and a noise-like (diffuse) component. The problem at hand is to estimate the angles-of-arrival of the signals. At low grazing angles, the situation becomes particularly difficult because the indirect component dominates and is correlated with the direct component. Furthermore, the angular separation between the direct and indirect components is usually less than an antenna beamwidth. When the height of a radio source is such that the angular separation between the direct and indirect paths, along which the radio signal travels from the source to the antenna, is less than 0.8 antenna beamwidths (BWs), conventional monopulse techniques will produce large estimation errors [3].

Many signal processing techniques have been developed for AOA estimation. Much of the research work to date has been concentrated in the area of array signal processing, wherein estimation algorithms have been developed and tested to determine how effective and robust they are in separating a signal source from its image and then providing an accurate estimate of the AOA of the signal source. These estimation algorithms are basically categorised into two groups: (1) wavenumber estimation techniques and (2) parametric estimation techniques.

The first group consists of various algorithms which are solely based on the decomposition of a covariance matrix whose terms consist of estimates of the correlation

between the signals at the elements of an array antenna. The covariance matrix is eigen-decomposed to obtain a vector space which consists of a signal subspace and noise subspace. The magnitudes of the eigenvalues associated with the eigenvectors of the covariance matrix can be used as the criterion for defining and differentiating between these subspaces. Due to the orthogonality of the two subspaces, the solutions are derived from the projection of one subspace onto the other. Some examples are the multiple signal classification (MUSIC) algorithm [4], the modified forward-backward linear prediction (FBLP) algorithm [5], and the Principal Eigenvector Gram-Schmidt (PEGS) algorithm [6]. However, all of these algorithms fail, to different extents, for coherent signals. This is because signal coherence causes the covariance matrix to become ill-conditioned [7]. Although the spatial smoothing technique [8] can be used to alleviate this problem, it degrades the performance of these algorithms by introducing biases to the estimates and reducing the subaperture antenna gain and resolution.

The second group mainly comprises a variety of maximum likelihood estimation (MLE) techniques [9, 10, 11, 12, 13]. In the MLE approach, a particular likelihood function is formulated for the given radio signals. The MLE estimates of desired parameters, such as angles-of-arrival, are the ones for which the likelihood function is maximised. In some models, a priori information is introduced to the ML process, thereby enhancing the performance of the ML estimator [12, 13, 14]. In general, MLE techniques are much less affected by the signal coherence and therefore, these algorithms usually offer better performance for estimation of AOA in multipath. However, the implementation of the MLE techniques requires the minimisation of a highly nonlinear, multimodal, and multidimensional cost function. When a coarse initial estimate is available, the ML cost function can be minimised using the Gaussian-Newton algorithm. However, in the absence of a reliable first estimate, computationally intensive, heuristic algorithms such as simulated annealing [15] have been used to minimise the cost function. Although some methods for decreasing the computational complexity of the ML criterion have been proposed, it is still difficult to achieve the desired

real-time performance. To handle this computational problem, neural networks have been proposed recently as a possible solution [16, 17, 18, 19, 20].

There is a resurgence of interest in the use of neural networks to solve some of the problems arising in signal processing applications in recent years [21, 22]. This stems from

1. the fact that a neural network is a massively parallel interconnected network of simple elements, and

2. the fact that the hierarchical organisation of a neural network is intended to interact, in the same way as a biological nervous system does, with the objects of the real world.

More specifically, a neural network may consist of a layer or multilayer nonlinear network that tries to reproduce human perceptual process through a "learning" procedure in which the error between the input-output pair is minimised by varying the synaptic weights of the interconnections between the input and output layers. Amongst many applications of signal/image processing in which neural networks are found to be useful, the problem of AOA estimation is considered as one potential application.

There are two main trends in the application of neural networks to signal processing problems. In the first trend, a signal processing problem is represented in terms of optimisation where a cost function matches the energy function of a particular neural network [23]. The neural network arrives at a solution of the problem by minimising its energy function. Recurrent networks are usually suitable for this type of processing, where the output of a neuron is fed back as inputs to other neurons and/or to itself. An exciting realisation in this approach is that analogue computations with binary stable outputs can be used to solve such problems, similar to the way biological neural networks do their computations. This is one reason why the input and the output signals in neural networks are often represented in binary codes. The second

trend is the application of neural networks to signal recognition problems, especially speech/image recognition and vision [24] in which the information processing operation is an approximation of a mathematical mapping. A feedforward network with the full capability of approximation is found to be an efficient model for this application.

The application of neural networks to the AOA estimation problem in the literature mainly belongs to the first category. Following the neural network approach to combinatorial optimisation proposed by Hopfield and Tank [25], different types of AOA cost functions, such as the MUSIC algorithm [26] and the MLE [16, 17], have been mapped onto the quadratic energy function of the Hopfield model network in a minimum mean squares sense. For Gaussian noise and a linear equispaced sensor array, this estimator can be shown to be equivalent to the maximum likelihood estimator [17]. These neural network approaches to the AOA problem are all based on the optimisation of the highly nonlinear cost function and suffer inherently from the existence of local minima and slow convergence.

The other approach of using neural networks is to consider the estimation problem as a mathematical mapping and to apply a feedforward neural network to approximate the mapping of the underlying process. In particular, the AOA estimation problem is considered as an inverse mapping problem [19]. That is, the outputs of the antenna elements in a receiving array can be viewed as a transformation or mapping $H : \Re^L \rightarrow \mathcal{C}^K$ from the space of AOA, $\boldsymbol{\theta} \in \Re^L$, to the space of sensor outputs, $\boldsymbol{s} \in \mathcal{C}^K$. Determining $\boldsymbol{\theta}$ from the sensor outputs $\boldsymbol{s}$ can be viewed as finding the corresponding inverse mapping $F : \mathcal{C}^K \rightarrow \Re^L$ of H, so that whenever the sensor outputs are given, the AOA can be obtained through this inverse mapping directly. Since the inverse mapping is highly nonlinear, a nonlinear feedforward neural network is naturally the candidate for the realisation of the nonlinear transformation.

One of these nonlinear networks is the radial-basis function (RBF) network, which has been studied extensively. Radial-basis functions were first introduced as a means of carrying out real multivariate interpolation. It has been shown that the RBF has

the following properties: (1) it can be derived from the regularisation theory [27]: (2) it has the universal approximation ability [28]; and (3) it has the best approximation ability [29]. The radial basis function approach can be implemented using a network. Bloomhead and Lowe [30] were the first to exploit the use of RBF in the design of neural networks. Other major contributions to the theory, design, application of the RBF networks include papers by Moody and Darken [31] and Renals [32]. The RBF network in its most basic form consists of three entirely different layers, namely the input layer, the hidden layer, and the output layer. The input layer is made up of source nodes (sensory units). The hidden layer is usually of higher dimension than the input layer, i.e., the number of hidden units is larger than the number of source units. The output layer supplies the response of the network to the activation patterns applied to the input layer. The transformation from the input space to the hidden-unit space is nonlinear, whereas the transformation from the hidden-unit space to the output space is linear. The RBF networks have been applied to a wide variety of signal/image processing problems. The applications of the RBF networks include image processing [33], speech recognition [34, 35], adaptive equalisation [36, 37], nonlinear beamforming [38], and low-angle radar tracking [39, 40]. The application of the RBF network to AOA estimation [19, 20] is simply a manifestation of the fact that the RBF networks are finding more and more application areas.

1.2 Overview of this Thesis

We have briefly reviewed the problem of AOA estimation in a multipath environment and the techniques that have been proposed to deal with the problem. It is an undisputed fact that the ML method has an optimum performance in the statistical sense and is insensitive to the coherence of the incoming signals. To further exploit these advantageous properties, we will show in this thesis that by introducing a priori information to the MLE process, the performance the MLE estimator can be enhanced.

The main thrust for using the a priori information is based on Litva's low-angle tracking technique where a deterministic multipath model was used to improve tracking performance [41]. Based on this framework, Bosse et al. modified the technique using the ML criterion [12]. The MLE technique is developed for estimating AOA of radio signals propagating above a sea surface. The key feature of this approach is the use of a highly deterministic multipath signal model that we have developed. This model, which is similar to that in [41], makes use of the geometrical information and a priori knowledge of a number of physical parameters. These parameters include: (1) the refractivity gradient, (2) the reflection coefficient, (3) the specular and diffuse scattering coefficient, and (4) the divergence factor. In this model, however, the signals are modelled in terms of the AOA's rather than the source height. The Cramer-Rao bounds will be derived specifically for this estimator. They provide a statistical measure for the mean-squared errors (MSE) in the estimated angles-of-arrival. The bounds will be used as a reference for the performance analysis of the MLE technique based on computer simulations.

It is common knowledge that the computational load of the nonlinear optimisation procedure inherently required by the MLE method is too heavy for real-time processing. For example, a common way to implement the MLE method is to use the grid-search technique to obtain the maximum value of the likelihood function. This is, of course, a slow process and is not suitable for real-time applications. Alternatively, we will propose the matched-filter implementation where the process can be carried out in a parallel fashion. The drawback with this implementation is the high degree of complexity in hardware, which diminishes its practicality and feasibility.

Accordingly, we have set out to search for a practical and feasible alternative. In light of the parallelism in the matched-filter implementation, we propose a novel approach to the AOA estimation problem, which is based on the use of an associative memory. The parallel structure of the matched-filter implementation provides some insight into the approach of using an associative memory. The connection between

the matched-filter implementation and the associative memory is that they resemble each other. The functionality of an associative memory is identical to that of an inverse mapping network. This provides a more comprehensive explanation for the rationale of exploiting the inverse mapping concept for the AOA estimation [19]. A neural network is suitable for forming the associative memory by approximating the underlying inverse mapping F. Another reason for using a neural network to form such an inverse associative mapping is that a closed-form inverse formula for the mapping would be difficult to derive because F is highly nonlinear. Therefore a neural network with nonlinearity should be used for forming the associative memory. The neural network that we will use is the RBF neural network.

In the traditional estimation approaches such as MLE, a mathematical model is constructed for the underlying physical mechanism that is responsible for generating the input data. Based on this model, an estimation technique is then developed, as shown in Fig. 1.3(a). Thus, the success of this approach depends on how closely the model describes the realities of the pertinent environment. The neural network approach, in direct contrast, offers an entirely different solution to the underlying problem. It is well known that one of the strengths of a neural network is that it does not require an explicit mathematical model. A neural network is able to provide itself, based on real-life data, a statistical model of the environment, in which it operates, to carry out the estimation (Fig. 1.3(b)). The success of this approach depends on how representative the training data set is of the environment. The use of real-life data for training is a blessing to this approach. However, it is also a curse in some cases where it may take a very long time or may be impractical to obtain sufficient quantity of data for training. Therefore, of the above two approaches, one is not absolutely superior to the other. Rather, they are complementary to each other. In the case where an estimation technique is impractical to implement even though a mathematical model is constructed and where it is impractical to obtain sufficient quantity of data for training, another approach is needed as an alternative to the

above two. A neural network can be trained using synthetic data derived from the mathematical model to carry out the estimation (Fig. 1.3(c)). The success of this hybrid approach, of course, depends on how closely the model describes the realities of the pertinent environment. The latter approach is the one adopted in this thesis because it is suitable for the applications.

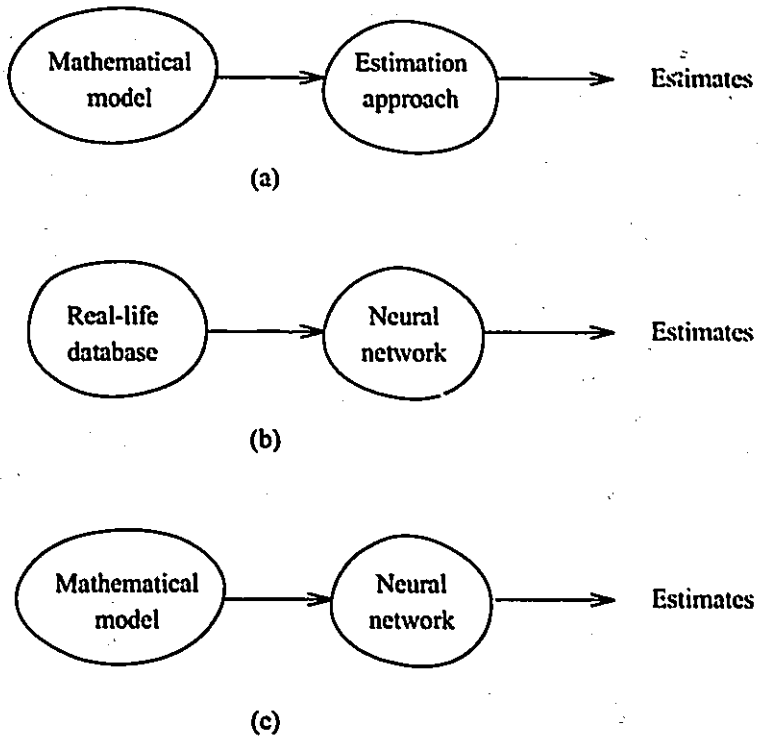

Figure 1.3: (a) Traditional estimation approach; (b) neural network modelling approach; (c) hybrid approach.

This thesis is organised as follows. Chapter 2 gives the problem definition of AOA estimation in a multipath environment. This is followed by a description of the use of antenna arrays for AOA estimation and by the development of the signal models that will be used in this thesis. In Chapter 3, the maximum likelihood method will be derived, as well as the Cramer-Rao bounds (CRB) specifically for this estimator. Consideration will be given to implementation issues. Chapter 4 gives a discussion

on the principles of the radial-basis function network for AOA estimation, whereas Chapter 5 gives the description of the applications of the RBF network to AOA estimation. In Chapter 6, computer simulations and real data tests are presented. The performance of both the MLE method and RBF network approach will be evaluated through simulations. Finally, conclusions and suggestions for future work will be given in Chapter 7.

1.3 Contributions of this Thesis

The contributions made by this thesis include:

1. the development of the deterministic multipath model for AOA estimation,
2. the derivation of the MLE solution with *a priori* information,
3. the derivation of the CRB specifically for AOA estimation in a multipath environment,
4. the introduction of the inverse mapping concept to the AOA estimation problem,
5. the development of the principles of the RBF networks for AOA estimation, which includes the use of the associative memory model for inverse mapping, analysis on interpolation and approximation, and analysis on effects of using the generalised RBF network to implement the nonlinear associative memory,
6. the application of the RBF associative memory network to approximate the inverse mapping for AOA estimation, and the validation of the principles and analyses through computer simulations and real data tests.

Chapter 2

AOA ESTIMATION IN A MULTIPATH ENVIRONMENT

2.1 Problem Definition

In Fig. 2.4 is illustrated the geometry for the two-ray model over a flat surface. A point source is at a distance of r^d from the receiver. If the source is assumed to be a narrow-band signal, it can be represented by

$$y(t) = a \cdot e^{j(\omega t + \xi)} \tag{2.1}$$

where a is the amplitude, ω is the angular frequency, and ξ is the initial phase. The use of complex representation is justifiable on the grounds that the signal is converted to complex form on reception. In the presence of multipath, the signal received by the receiver consists of two components, namely the direct and indirect signals. The direct signal is given by

$$y^d(t) = y(t) \cdot e^{-j\kappa r^d} \tag{2.2}$$

whereas for a simple multipath model of a flat earth, the indirect signal is given by

$$y^i(t) = y(t) \cdot \rho e^{j\varphi} \cdot e^{-j\kappa r^i} \tag{2.3}$$

where $\rho e^{j\varphi}$ is the complex reflection coefficient, $\kappa = \frac{2\pi}{\lambda}$ is the wavenumber, and r^i is the total length of the indirect path. Thus, the total received signal is given by

$$\begin{aligned} y^r(t) &= y^d(t) + y^i(t) \\ &= y(t) \cdot e^{-j\kappa r^d} \left(1 + \rho e^{j(\varphi+\gamma)}\right) \end{aligned} \tag{2.4}$$

where $\gamma = -\kappa(r^i - r^d)$. If coherent detection is used to demodulate the received signal, the final form of the received signal is then

$$\begin{aligned} s(t) &= y^r(t) a_o \cdot e^{-j\omega t} \\ &= A \cdot e^{j\xi} \cdot e^{-j\kappa r^d} \left(1 + \rho e^{j(\varphi+\gamma)}\right) \end{aligned} \tag{2.5}$$

where $A = a \cdot a_o$ and a_o is the amplitude of a local oscillator signal.

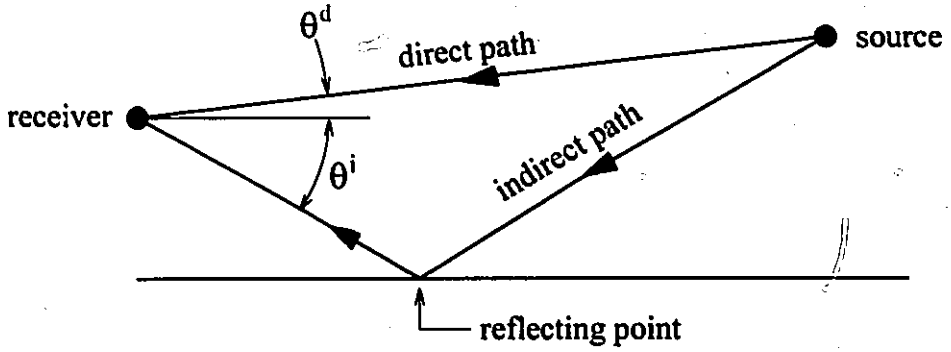

Figure 2.4: Two-ray propagation model for a flat earth.

Given a received signal represented by (2.5), which is the mixture of the direct component and the reflected component, the objective is to estimate the elevation angle at which the direct component, as well as the reflected component, arrives at the receiver, i.e. to estimate the AOA.

2.2 Antenna Array for AOA Estimation

2.2.1 Use of Antenna Arrays

It is impractical, if not impossible, to estimate the AOA based on the phasor signal given by (2.5). The analogy is that we try to solve multiple unknowns in one equation. A feasible and practical means is to use an antenna array, which is able to resolve the modes of propagation arriving at the receiver from different directions within a relatively short time interval.

In an array system, the received signals are obtained using a group of antenna elements at different known spatial locations in the field of interest. The function of the array can be understood as spatially sampling the electromagnetic waves. For example, Fig. 2.5 depicts a uniformly spaced linear array with K identical antenna elements. The spacing between the elements is denoted by d. The wave can be considered as a plane wave if the signal source is within the far-field (Fraunhofer) region of the receiver antenna. The Fraunhofer region R_{ff} is defined by [42]

$$R_{ff} \geq \frac{2D^2}{\lambda} \tag{2.6}$$

where D denotes the aperture of the antenna array and λ represents the wavelength. In the worst case where $R_{ff} = \frac{2D^2}{\lambda}$, the phase difference between the array centre and the end element is 22.5° if the signal source is at the boresight of the array.

If the plane wave impinges upon the array at an angle θ with respect to the array normal, the wavefront arrives at Element 2 sooner than at Element 1 since the differential distance along the two ray paths is $d\sin\theta$. The phase difference between the two elements is

$$\Delta\phi = \frac{2\pi}{\lambda} d\sin\theta. \tag{2.7}$$

Therefore, by determining the phase difference between antenna elements in the array, one is able to estimate the AOA of a radio signal.

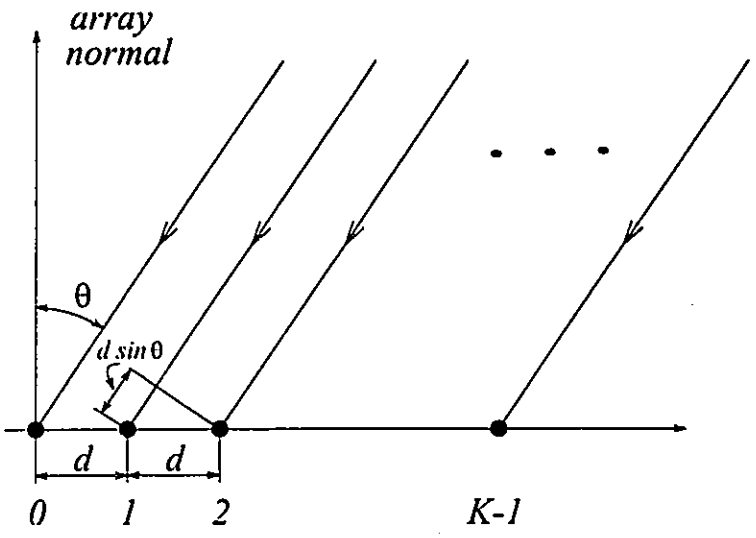

Figure 2.5: Linear array with K elements uniformly spaced by a distance d.

2.2.2 Linear Sparse Array

The most commonly used array geometric structures are those with regular geometrical shapes such as linear, circular, and planar arrays. For the applications considered in this thesis, a vertical linear array is a suitable candidate. The locations of the antenna elements in the linear array can be equally spaced or sparsely spaced.

The obvious effect of a linear sparse array is that the number of array elements is reduced, thereby reducing the cost of the array, while maintaining a certain array aperture size. However, in choosing the positions of the array elements, certain criteria must be satisfied so that the optimal performance of the array can be achieved.

In simple terms, an array may be viewed as a grating or compound interferometer. The array is utilised by considering the outputs from all the possible parings of elements, with each pair acting as an interferometer. Given K elements, there are $K(K-1)/2$ possible pairings. In order to determine their spatial-frequency sensitivity, it is necessary to determine the separation between the elements of each pair. For a uniform structure having unit spacing, there are $K-1$ pairs having a separation of

one unit, $K-2$ pairs having a separation of two units, etc., ending with one pair of elements separated by $K-1$ units.

This type of structure analysis can be presented by the use of a coarray. The coarray is described by considering the values of the separations which exist for each pairing of elements as an interferometer. Fig. 2.6(a) presents an example of this type of characterisation for a four-element uniform array. As well, depicted in the figure are the four pairings obtained by considering elements paired with itself. In this context, there are $K(K+1)/2$ possible pairings. The redundancy concept can be seen in Fig. 2.6(a) as there are three pairings separated by one unit compared to only one pair separated by three units. Reducing the redundancy involves spreading out the coarray, thereby, increasing the number of distinct separations that may be obtained.

The alternate rationale for reducing the redundancy comes from the spectral estimation problem directly. It is well know that the power density is the Fourier transform of the autocorrelation function. A sequence obtained by sampling a process at discrete points can also be described by an autocorrelation function which is also defined at discrete points. It is also known that there is a direct analogy between temporal sampling and spatial sampling if the temporal sampling intervals correspond to the spacing of the antenna elements. Therefore, in this case, the wavenumber spectrum is the Fourier transform of the autocorrelation function obtained from the array sensor outputs. The coarray represents the points at which the autocorrelation function is known. Thus, it is the coarray, not the initial data snapshot, for which uniform sampling is desired.

The minimum redundancy structure attempts to arrange the sampling points such that the coarray is approximately uniformly spaced and is extended beyond the conventional aperture by "spreading out" the redundant estimates of the correlation function. Fig. 2.6(b) depicts the four-element non-redundant array. Each spacing for the elements provides an estimate of the autocorrelation function at a unique

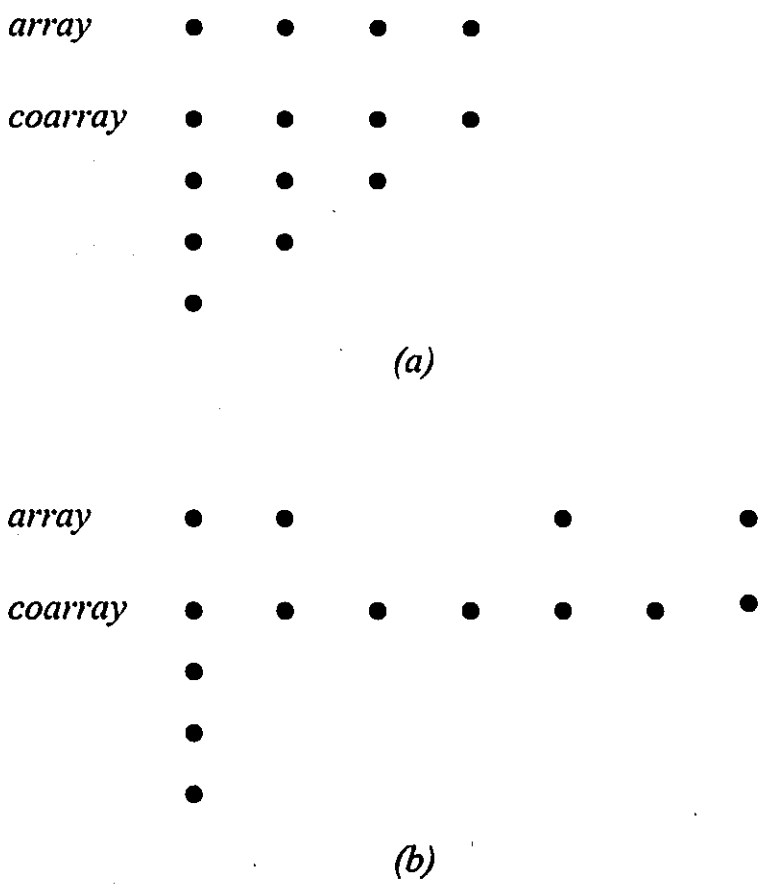

Figure 2.6: Comparison of four-element arrays and their associated coarrays (a) uniform, (b) optimum non-redundant.

sampling point. The coarray contains all positions between zero and six, and only the value at zero is repeated. Comparing the results to the four-element uniform array as described in Fig. 2.6(a), the increased aperture is immediately apparent. The three redundant estimates of the uniform array effectively have been added to the end of the aperture.

Unfortunately, for $K > 4$, there exist no ideal non-redundant arrays. An ideal non-redundant array is one for which its coarray contains exactly one estimate of a particular autocorrelation sample and the set of samples are contiguous. Nonetheless, there are minimum redundant arrays [43], which are defined to be the structures that minimise the redundancy in the coarray.

For these reasons, we assume a minimum redundancy linear array in the study. In particular, a nine-element minimum redundant linear array is used with its elements located at [0 d 4d 10d 16d 22d 24d 27d 29d], where d represents one unit of inter-element spacing. The value of d will be set at 0.05715 metres in this study. As a result, the antenna aperture is 1.66 metres. Since we assume the frequency of the signal to be 10.2 GHz, the beamwidth of the array is 0.9°. The height of the bottom element (defined as the first element) of the array with respect to the sea surface is 7.8 metres. The above specific values are used for the consistency with the experimental set-up, which will be presented in Chapter 6.

2.3 Array Multipath Signal Model

The geometry for an antenna array and a radio source is illustrated in Fig. 2.7. At this point in the discussion, a simple flat earth model is considered. Based on the geometry, all of the signal paths, including both the direct and indirect paths, between the source and the elements of the array vary in length. In other words, r^d and r^i are functions of the element index, k. From (2.5), the received signal at the k^{th} element

is

$$s_k = A\, e^{j\xi} e^{-j\kappa r_k^d} \left[1 + \rho\, e^{j(\varphi+\gamma_k)}\right] \tag{2.8}$$

where

$$\gamma_k = \kappa \left(r_k^i - r_k^d\right).$$

Based on the geometry of Fig. 2.7, r_k^d and r_k^i are given by

$$r_k^d = [r^2 + (h - z_k)^2]^{1/2} \tag{2.9}$$

and

$$r_k^i = [r^2 + (h + z_k)^2]^{1/2}, \tag{2.10}$$

where r is the distance between the source and the array and h is the height of the source.

In order to model the signal more accurately, the curvature of the signal path due to refraction in the troposphere, in addition to the curvature of the earth itself, must now be taken into account.

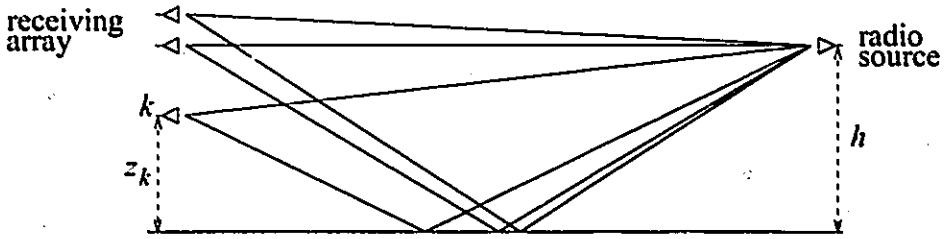

Figure 2.7: Flat-earth multipath geometry for an antenna array.

Radio waves travel in straight lines in free space. However, they do not travel in straight lines when travelling in the earth's atmosphere. They are bent or refracted. Refraction of radio waves in the atmosphere is caused by the variation with altitude of the velocity of propagation. The variation is usually quantified by the index of refraction, which is defined as the ratio of the velocity of propagation in free space to that in the medium in question.

Ray-path curvature is usually dealt with by replacing the earth with an imaginary earth whose effective radius is [44],

$$R_e = R \cdot \left(1 + 6.37 \times 10^{-3} \frac{dN}{dh}\right)^{-1} \tag{2.11}$$

where R is the radius of the real earth and $\frac{dN}{dh}$ is the refractivity gradient. The ray paths for the imaginary earth turn out to be straight lines for this choice of R_e, and therefore much easier to deal with. From the curved earth geometry given in Fig. 2.8, the distance from the receiver to the point of reflection, r_1, can be found by solving the cubic equation [44]

$$2r_1^3 - 3rr_1^2 + [r^2 - 2R_e(z_k + h)]r_1 + 2R_e z_k r = 0. \tag{2.12}$$

The signal paths and their difference are then computed as follows:

$$r_k^d = [r^2 + (h' - z_k')^2]^{1/2}, \tag{2.13}$$

and

$$r_k^i = [r^2 + (h' + z_k')^2]^{1/2}, \tag{2.14}$$

where h' and z_k' are the effective values for h and z_k. Specifically, the effective height of the source is approximated by

$$h' \simeq h - \frac{(r - r_1)^2}{2R_e} \tag{2.15}$$

and the effective height of the array elements by

$$z_k' \simeq z_k - \frac{{r_1}^2}{2R_e}. \tag{2.16}$$

Expanding (2.13) and (2.14) using the binomial theorem gives

$$r_k^d = r \cdot \left(1 + \frac{(h' - z_k')^2}{2r^2} - \frac{(h' - z_k')^4}{8r^4} + \ldots\right) \tag{2.17}$$

and

$$r_k^i = r \cdot \left(1 + \frac{(h' + z_k')^2}{2r^2} - \frac{(h' + z_k')^4}{8r^4} + \ldots\right). \tag{2.18}$$

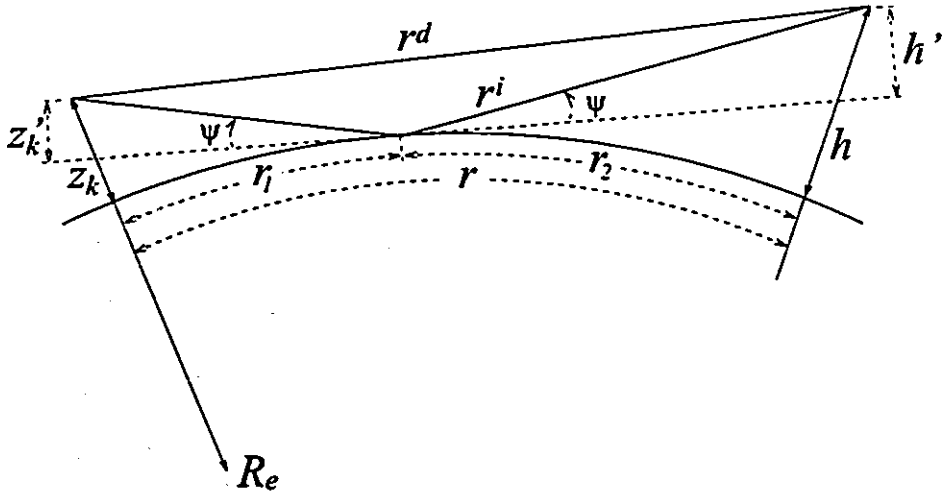

Figure 2.8: Geometry of a multipath model for a curved earth.

In the low elevation angle region $r >> h'$ and $r >> z'_k$, therefore the fourth and higher order terms are small compared to the first two terms and can therefore be neglected. It follows that

$$r_k^d \simeq r + \frac{(h' - z'_k)^2}{2r} \tag{2.19}$$

and

$$r_k^i \simeq r + \frac{(h' + z'_k)^2}{2r}. \tag{2.20}$$

The path difference is then given by

$$\begin{aligned} \Delta r_k &= r_k^i - r_k^d \\ &\simeq \frac{2h'z'_k}{r}. \end{aligned} \tag{2.21}$$

This approximation is based on the fact that in the low-angle region, the distance along the tangent plate from the respective terminal points to the reflection point may be taken to be r_1 or $r - r_2$. The composite signal given by (2.8) becomes

$$s_k = A \cdot e^{j\xi} e^{-j\kappa(r+h'^2/2r)} e^{j\kappa(2h'z'_k - z'^2_k)/2r} \left(1 + \rho e^{j(\varphi+\gamma'_k)}\right) \tag{2.22}$$

where

$$\gamma'_k = -\kappa \cdot \frac{2h'z'_k}{r}. \tag{2.23}$$

2.4 AOA Signal Model

When estimating angles-of-arrival, it is assumed that the direct and indirect signals arrive at angles which remain unchanged from element to element. In other words, the signals are considered to be two plane waves arriving at angles θ^d and θ^i with respect to the array normal. The output at the k^{th} element of the array is then

$$\begin{aligned} s_k &\simeq A \cdot e^{j\xi}\left(e^{-j\kappa d_k \sin\theta^d} + \rho e^{j\varphi} \cdot e^{-j\kappa d_k \sin\theta^i}\right) \\ &= A \cdot e^{j\xi} e^{-j\kappa d_k \sin\theta^d}\left(1 + \rho e^{j\alpha_k}\right) \end{aligned} \tag{2.24}$$

where α_k is the total phase difference between the direct and indirect signals and d_k is the distance of the k^{th} element with respect to the reference element, say the middle element. By comparing (2.22) and (2.24), the phase difference α_k in (2.22) is given by

$$\alpha_k = \varphi + \gamma_k' \tag{2.25}$$

where γ_k' is given by (2.23). Although h' and r in (2.23) are unknown, they can be expressed in terms of θ^d, θ^i, and the height of the reference element, z_r. Their relations are derived, based on the geometrical model given in Fig. 2.8 , as follows [13],

$$r = \frac{r_1 \cdot \tan\theta^i - z_r(1 - 3S^2)}{\tan\theta^d + \tan\theta^i + 4r_1/2R_e} \tag{2.26}$$

and

$$h' = r \cdot \tan\theta^d + z_r - \frac{r_1(r - r_1)}{R_e}\,. \tag{2.27}$$

In (2.26), S is the compensation factor for adjusting the flat earth expressions to apply to the curved-earth signal path, which is given by

$$S^2 = \frac{{r_1}^2}{2R_e \cdot z_r} \tag{2.28}$$

and r_1 is the distance from the receiver to the reflection point, which is given by

$$r_1 = -R_e \cdot \tan\theta^i - \sqrt{R_e(R_e \cdot \tan^2\theta_i - 2z_r)}. \tag{2.29}$$

The grazing angle is given by

$$\psi = \tan^{-1}(-\tan\theta^i - \frac{r_1}{R_e}) \,. \tag{2.30}$$

All unknown parameters in (2.26) to (2.30) are represented in terms of θ^d and θ^i. In other words, the output at the k^{th} element of the array s_k is a complex nonlinear function H of θ^d and θ^i, i.e.,

$$s_k = H(\theta^d,\ \theta^l). \tag{2.31}$$

If the receiver noise and diffuse noise from the reflected signal are included, the signal at the k^{th} antenna element is given by

$$x_k = s_k + n_k^R + n_k^D \tag{2.32}$$

where superscript R denotes the receiver noise and D the diffuse noise. If both n_k^R and n_k^D are assumed to have a normal distribution, they can be combined to give an effective noise term, *i*.e.

$$x_k = s_k + n_k. \tag{2.33}$$

2.5 The Reflection Coefficient

In (2.22) and (2.24) the term $\rho e^{j\varphi}$ is the overall reflection coefficient. It generally consists of three factors, namely the reflection coefficient for a plane surface, the divergence factor due to a curved surface, and the surface roughness factor.

The approximate expressions for the Fresnel reflection coefficient for a smooth plane surface are given by

$$\Gamma_v \simeq \frac{\sqrt{\epsilon_c}\cdot\sin\psi - 1}{\sqrt{\epsilon_c}\cdot\sin\psi + 1} \tag{2.34}$$

for vertical polarisation and

$$\Gamma_h \simeq \frac{\sin\psi - \sqrt{\epsilon_c}}{\sin\psi + \sqrt{\epsilon_c}} \tag{2.35}$$

for horizontal polarisation, where ψ is the grazing angle and ϵ_c is the complex dielectric constant which is given by

$$\epsilon_c = \frac{\epsilon}{\epsilon_o} - j\ 60\ \lambda \cdot \sigma_c. \tag{2.36}$$

$\frac{\epsilon}{\epsilon_o}$ is the relative dielectric constant of the reflecting medium and σ_c is its conductivity.

When a bundle of parallel rays is incident on the convex side of a curved surface, the rays diverge after reflection because each has a slightly different angle of incidence. Because of this divergence, the reflected signal is defocused and the power density is reduced. The divergence factor can be derived solely from geometrical considerations [45] and is given here in a form which incorporates the usual approximations which are applicable in the low-angle region, namely

$$D \simeq \left(1 + \frac{2r_1 r_2}{R_e r \psi}\right)^{-\frac{1}{2}}. \tag{2.37}$$

In addition to the divergence factor, the effect of reflection by rough surfaces is also considered here. In general, when parallel rays are reflected by a rough surface, they undergo changes in path length and amplitude because of the deviations from a smooth reflecting surface. These changes result in a portion of the reflected signal becoming diffuse. Where previously it was totally coherent, it now consists of two components: (1) a diffuse component and (2) a coherent component with reduced magnitude. The reduction in the magnitude of the coherent component brought about by reflections from a rough surface is related to the grazing angle and the signal wave length. It is best expressed in terms of the root-mean-squared (rms) specular scattering coefficient S which is given by [45]

$$S = e^{-\mu} \tag{2.38}$$

where

$$\mu = \begin{cases} 2(2\pi\eta)^2 & \eta \leq 0.1 \text{ rad.} \\ 0.16\eta^2 + 7.42\eta + 0.0468 & \text{otherwise} \end{cases} \tag{2.39}$$

and η is the surface roughness factor given by

$$\eta = \frac{\sigma_H \psi}{\lambda} . \tag{2.40}$$

In (2.40), σ_H is the rms value of the sea wave height and λ is the wavelength of the microwave signal. In Fig. 2.9 are shown the specular and diffuse scattering coefficients plotted against the roughness factor. It can be noticed that the specular scattering coefficient curve has both a theoretical branch and an experimental branch. In fact, for $\eta > 0.1$ radians, (2.39) is the best fit to the experimental branch.

The parameter used for quantifying the roughness of an ocean surface is called sea-state. It is related to the rms wave height, σ_H, through the significant wave height, H_s. The significant wave height is defined as the average of the highest third of the waves in a wave group and is related to σ_H by

$$\sigma_H = \frac{H_s}{4}. \tag{2.41}$$

Combining the above three factors, the overall reflection coefficient is

$$\rho e^{j\varphi} = \Gamma \cdot D \cdot S. \tag{2.42}$$

The diffuse component is also incorporated into the model. For simplicity, it is treated as incoherent white Gaussian noise. The diffuse noise power is proportional to the diffuse scattering coefficient given in Fig. 2.9(b).

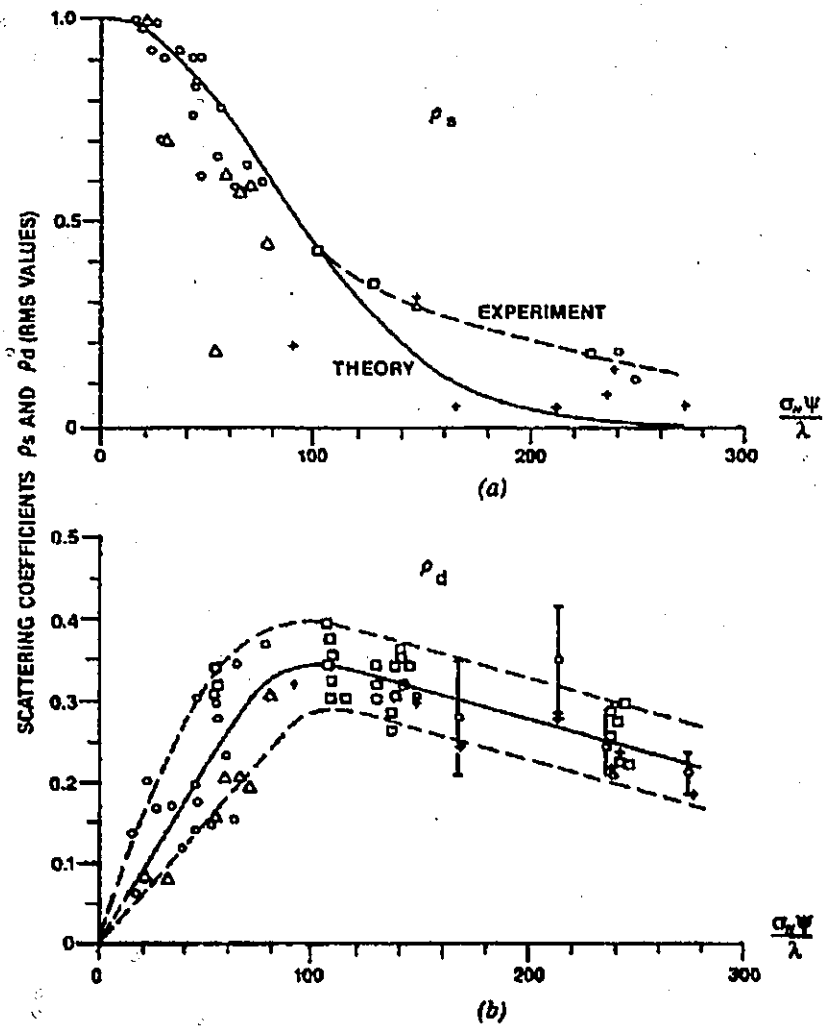

Figure 2.9: Specular and diffuse scattering coefficients compared to microwave measurements (from: Beckmann and Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surface*, Artech House, Norwood, MA, 1987.). (a) Specular scattering coefficients; (b) diffuse scattering coefficients. Note that the specular curve has a theoretical branch and an experimental branch.

Chapter 3

THE MLE TECHNIQUE AND CR BOUNDS

3.1 Introduction

In the previous chapter, a signal model is expressly designed for estimating the vertical direction of a low-elevation radio source, located above the ocean surface. The key feature of this model is the use of a highly deterministic multipath signal model described in the previous chapter. This model makes use of the geometrical information and a priori knowledge of a number of physical parameters. In this chapter, our objective is to show that by introducing the a priori information, which is incorporated into the signal model, to the MLE process, the performance of the ML estimator can be greatly enhanced. In contrast to the standard angle-of-arrival model, wherein there are two unknown angles, two unknown signal amplitudes, and two unknown signal phases, this signal model has only four unknown parameters, i.e. the total signal amplitude, total signal phase and angles-of-arrival of the direct and indirect signals. Therefore, unlike other MLE techniques in which the distance between the source and the array is a required parameter in their signal models [12], this algorithm does not require such information. This makes the technique particularly useful in the case

where a passive electronic measurement system is used to estimate the elevation of an incoming signal wherein no range information is available. In this chapter, we also derive the Cramer-Rao bounds specifically for this estimator. They provide a statistical measure for the mean-squared errors (MSE) in the estimated angles-of-arrival. The bounds will be used later on as a measure of performance when evaluating the RBF neural network approach.

3.2 The Maximum Likelihood Technique

The formulation of the likelihood function is based on the assumption that the receiver noise is Gaussian and independent from channel to channel, but identically distributed, with variance σ^2 and zero mean. The likelihood function for the unknown parameters, which can be represented by a vector, $\chi(A, \xi, \theta^d, \theta^i)$, is given by the conditional probability density of the observations,

$$\begin{aligned} \mathcal{L} &= P(\boldsymbol{x}|\chi) \\ &= \frac{1}{(2\pi)^{K/2}[\det(\boldsymbol{R}_n)]^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{s})^H \boldsymbol{R}_n^{-1}(\boldsymbol{x}-\boldsymbol{s})} \end{aligned} \tag{3.43}$$

where $\boldsymbol{x}$ is the sampled data vector or the simulated data, the elements of vector $\boldsymbol{s}$ is given by (2.24), and the superscript H denotes the Hermitian transpose operation. $\boldsymbol{R}_n$ is the noise covariance matrix, given by

$$\boldsymbol{R}_n = \sigma_n^2 \boldsymbol{I}, \tag{3.44}$$

where $\boldsymbol{I}$ is the identity matrix, and thus (3.43) becomes

$$\mathcal{L} = (\frac{1}{2\pi\sigma_n^2})^{K/2} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{s})^H \boldsymbol{R}_n^{-1}(\boldsymbol{x}-\boldsymbol{s})} \tag{3.45}$$

Taking the logarithm of $\mathcal{L}$ yields

$$\begin{aligned} l &= \ln(\mathcal{L}) \\ &= \frac{K}{2}\ln(\frac{1}{2\pi\sigma_n^2}) - \frac{(\boldsymbol{x}-\boldsymbol{s})^H(\boldsymbol{x}-\boldsymbol{s})}{2\sigma_n^2} \end{aligned} \tag{3.46}$$

It is clear that l is maximised if

$$\nu = \frac{(\boldsymbol{x}-\boldsymbol{s})^H(\boldsymbol{x}-\boldsymbol{s})}{\sigma_n^2} \tag{3.47}$$

is minimised. By using $\boldsymbol{s} = Ae^{j\xi}\boldsymbol{f}$ from (2.24), where $\boldsymbol{f}$ is independent of A and ξ, and by setting $\frac{\partial \nu}{\partial A} = 0$ and $\frac{\partial \nu}{\partial \xi} = 0$, it follows that if

$$Ae^{j\xi} = \frac{\boldsymbol{x}^H\boldsymbol{f}}{\|\boldsymbol{f}\|^2} \tag{3.48}$$

ν is minimised. Substituting (3.48) into (3.47) gives the optimum value of ν,

$$\nu_{opt} = \frac{\|\boldsymbol{x}\|^2}{\sigma_n^2}\left(1 - \frac{\|\boldsymbol{x}^H\boldsymbol{f}\|^2}{\|\boldsymbol{x}\|^2\|\boldsymbol{f}\|^2}\right) . \tag{3.49}$$

It follows that the ML estimates of θ^d and θ^i are the pair which corresponds to the minimum value of ν_{opt}. Therefore, a two-dimensional search is required in implementing this algorithm.

It should be pointed out that the model is based on the concept of specular reflections. Thus, the performance of the technique may be degraded if the variations in surface wave heights within the first Fresnel zone exceeds the Rayleigh roughness criterion

$$\sigma_H < \frac{\lambda}{8\sin\psi}. \tag{3.50}$$

It must be noted, though, that in those cases where the grazing angle is small, the roughness factor, η, remains small even for relatively rough surface, thereby keeping the diffuse component small.

3.3 Cramer-Rao Bounds for AOA Estimation

The use of the Cramer-Rao Bound to determine the performance limitations of wavenumber estimations is well described in the literature [46, 47]. The bound is also clearly documented in [48]. Therefore, only a short discussion is given for derivation of the bounds that applies specifically to the model that was discussed previously.

It is well known that the Cramer-Rao bounds are represented by the inequality,

$$E\{(\hat{\chi}_i - \chi_i)^2\} \geq \frac{\text{cofactor } J_{ii}}{\det(\boldsymbol{J})} \tag{3.51}$$

where J_{ii} is the i^{th} diagonal element of $\boldsymbol{J}$ which is the Fisher information matrix. Its elements are given by

$$J_{ij} = -E\left\{\frac{\partial^2 \log \mathcal{L}}{\partial \chi_i \, \partial \chi_j}\right\} , \tag{3.52}$$

where χ_i's are the unknown parameters to be estimated. It can be shown that

$$J_{ij} = -E\left\{\frac{\partial \boldsymbol{s}^H}{\partial \chi_i} \boldsymbol{R}_n^{-1} \frac{\partial \boldsymbol{s}}{\partial \chi_j}\right\} . \tag{3.53}$$

The complex signal vector, $\boldsymbol{s}$, can be decomposed into real and imaginary components, i.e.

$$\boldsymbol{s} = \boldsymbol{u} + j\boldsymbol{v} . \tag{3.54}$$

Since

$$\boldsymbol{R}_n^{-1} = \frac{1}{\sigma_n^2} \boldsymbol{I} , \tag{3.55}$$

(3.53) can be further expressed as

$$J_{ij} = -E\left\{\frac{1}{\sigma_n^2} \sum_{k=0}^{K-1} \left(\frac{\partial u_k}{\partial \chi_i}\frac{\partial u_k}{\partial \chi_j} + \frac{\partial v_k}{\partial \chi_i}\frac{\partial v_k}{\partial \chi_j}\right)\right\} . \tag{3.56}$$

By letting

$$\omega^d = \kappa \sin \theta^d \tag{3.57}$$

and

$$\omega^i = \kappa \sin \theta^i , \tag{3.58}$$

it follows from (2.24) that

$$u_k = A \cdot \cos(\xi - d_k \omega^d) + A \cdot \rho \cdot \cos(\xi + \varphi - d_k \omega^i) \tag{3.59}$$

and

$$v_k = A \cdot \sin(\xi - d_k \omega^d) + A \cdot \rho \cdot \sin(\xi + \varphi - d_k \omega^i) . \tag{3.60}$$

By evaluating the partial derivatives with respect to ω^d, ω^i, A, and ξ and rearranging them, the bound for ω^d is expressed by the inequality

$$E\{(\hat{\omega}^d - \omega^d)^2\} \geq \frac{\sigma_n^2}{A^2} \cdot \frac{\text{cofactor } Q_{11}}{\det(\boldsymbol{Q})} \tag{3.61}$$

and the bound for ω^i by

$$E\{(\hat{\omega}^i - \omega^i)^2\} \geq \frac{\sigma_n^2}{A^2} \cdot \frac{\text{cofactor } Q_{22}}{\det(\boldsymbol{Q})}. \tag{3.62}$$

In (3.61) and (3.62), $\boldsymbol{Q}$ is a four-by-four matrix and its elements are given by [13]

$$Q_{11} = \sum_{k=0}^{K-1} d_k^2, \tag{3.63}$$

$$Q_{22} = \rho^2 Q_{11}, \tag{3.64}$$

$$Q_{33} = Q_{44} = -K(1+\rho^2) + \sum_{k=0}^{K-1} \cos\beta_k, \tag{3.65}$$

$$Q_{12} = Q_{21} = -\rho \sum_{k=0}^{K-1} d_k{}^2 \cos\beta_k, \tag{3.66}$$

$$Q_{13} = Q_{31} = -\rho \sum_{k=0}^{K-1} d_k \sin\beta_k, \tag{3.67}$$

$$Q_{14} = Q_{41} = \sum_{k=0}^{K-1} d_k(1+\rho\cos\beta_k), \tag{3.68}$$

$$Q_{23} = Q_{32} = \rho \sum_{k=0}^{K-1} d_k \sin\beta_k, \tag{3.69}$$

$$Q_{24} = Q_{42} = \rho \sum_{k=0}^{K-1} d_k(\rho+\cos\beta_k), \tag{3.70}$$

and

$$Q_{34} = Q_{43} = 0, \tag{3.71}$$

where,

$$\beta_k = d_k(\omega^i - \omega^d) - \varphi\ . \tag{3.72}$$

Examples of the bounds for the nine-element minimum redundant array are given in Figs. 3.10 and 3.11. A smooth sea surface (SWH=0.01) is assumed and the refractivity gradient, $\frac{dN}{dh}$, is set equal to -39 $Nunits/km$. The SNR is 20 dB. For the results

shown in Fig. 3.10, the range of the source is kept fixed at 4.61 km and the height of the source is varied from 3 to 20 metres. The curves shown in Fig. 3.10 represent the bounds on MSE for both the direct signal and indirect signal as well as the overall MSE, which is defined as

$$\mathcal{E}_{overall} = \frac{1}{\kappa^2}\left[(\hat{\omega}^d - \omega^d)^2 + (\hat{\omega}^i - \omega^i)^2\right] \tag{3.73}$$

It should pointed out that the only reason why the bounds are plotted as a function of source height or range is that source height and range are the physical parameters that are controllable in the signal model. Varying these two parameters is equivalent to varying the AOA of the direct and indirect signals. As is expected, in general the errors are greater for the lower source heights. In Fig. 3.11, the bounds are obtained by keeping the source height at 15 m and varying the range from 1000 m to 6000 m. Changes in signal level brought about by spatial attenuation due to changes in range is represented by the variation of the SNR. In order to study the effect of the range on the CRB, it is assumed that the SNR is constant. Normally, errors increase as the range increases. However, as the source moves towards the array, the bound for the indirect signal rises. In this case, the amplitude of the indirect signal decreases monotonically because the specular reflection coefficient decreases with increasing grazing angle. One of the characteristics of the CRB that should be noted is its periodic behaviour. It can be observed that at certain ranges and heights large errors occur. The reason is that at these ranges and heights, the phase difference between the direct and indirect signals at the centre of the array aperture is π, causing the maximum cancellation of the received signals.

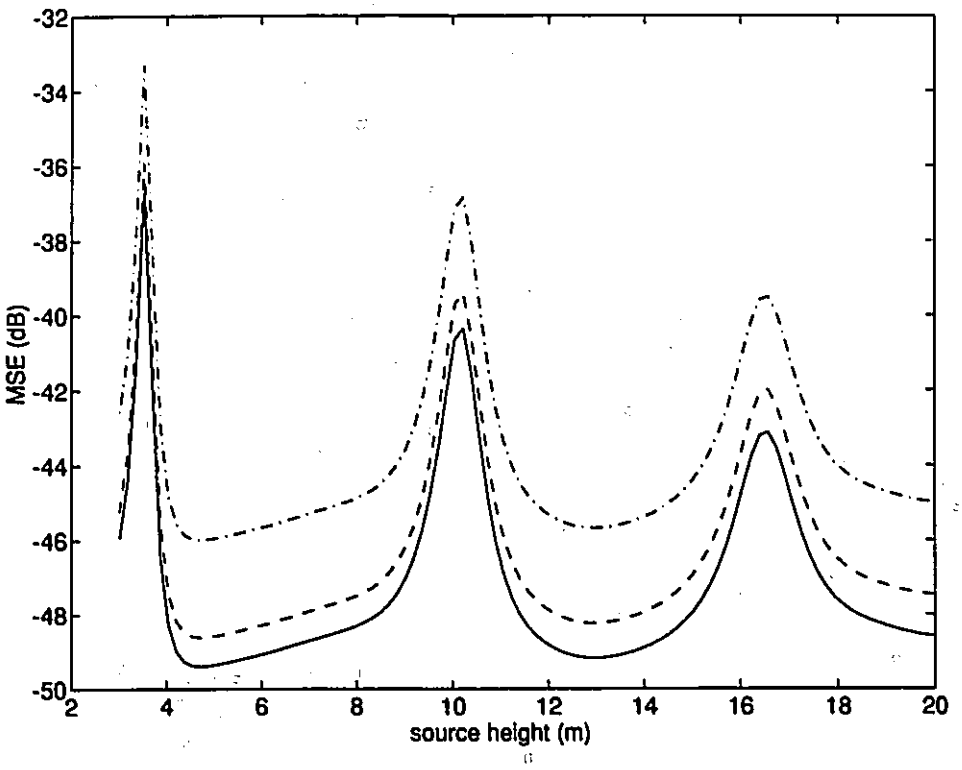

Figure 3.10: Cramer-Rao bounds as a function of source height for $r = 4.61km$ and $z_1 = 7.8m$; — lower bound for the direct signal, - - lower bound for the indirect signal, $-\cdot-$ lower bound for the overall MSE.

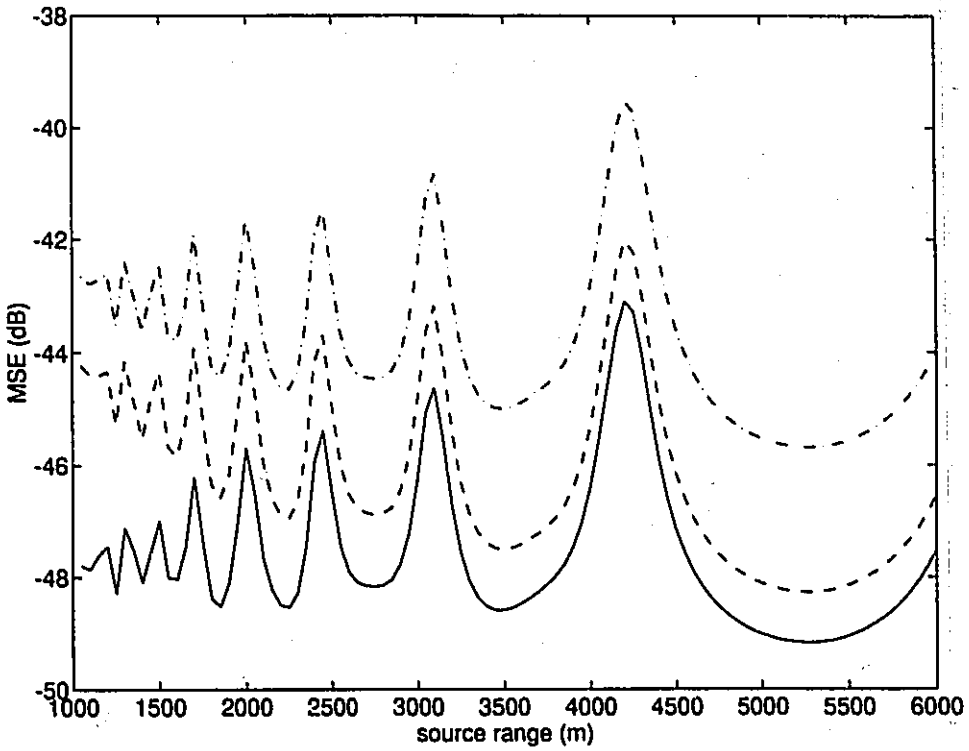

Figure 3.11: Cramer-Rao bounds as a function of source range for $h = 15m$ and $z_1 = 7.8m$; — lower bound for the direct signal, - - lower bound for the indirect signal, — · — lower bound for the overall MSE.

3.4 Implementation Considerations

3.4.1 Grid Search

As we have mentioned above, the optimum value of (3.49) can be determined using a two-dimensional search. In particular, we consider an $M = M^d \times M^i$ two-dimensional lattice of AOA values, where M^d is the number of discrete values of θ^d in an interval V^d and M^i is the number of discrete values of θ^i in an interval V^i. A grid on the lattice represents one of the combinations of $\{(\theta^d_m,\ \theta^i_n),\ m = 1, ..., M^d,\ n = 1, ..., M^i\}$. For the $(m, n)^{th}$ grid on the lattice, the corresponding array vectors $\boldsymbol{f}(\theta^d_m, \theta^i_n)$ and then the values of $\nu_{opt}(m, n)$ in (3.49) are evaluated. The estimate $\hat{\boldsymbol{\theta}}$ would be $[\theta^d_m\ \theta^i_n]^T$ which has the corresponding maximum value of ν_{opt}. To avoid repeated computations of the array vectors $\{\boldsymbol{f}(\theta^d_m, \theta^i_n),\ m = 1, ..., M^d,\ n = 1, ..., M^i\}$, they can be precalculated and stored in memory. When needed, they can be retrieved by the table-look-up method. For each estimate $\hat{\boldsymbol{\theta}}$, the required number of complex inner product operations is M. The accuracy of the estimate depends, of course, on how large the value of M is. That is, there is a trade-off between the number of calculations and the estimation accuracy.

3.4.2 Matched-Filter Implementation

In (3.49), the value of the norm $\|\boldsymbol{x}\|$ is fixed for a given $\boldsymbol{x}$. Therefore, ν is minimized if the following term

$$C = \frac{\|\boldsymbol{x}^H \boldsymbol{f}\|^2}{\|\boldsymbol{x}\|^2 \|\boldsymbol{f}\|^2} \tag{3.74}$$

is maximized. In fact, C is the cross-correlation coefficient between $\boldsymbol{x}$ and $\boldsymbol{f}$ and it is clear that $0 \leq C \leq 1$. Accordingly, the MLE technique that has been described in this chapter can be implemented using a family of correlating matched-filters, as shown in Fig. 3.12. The normalized input vector $\boldsymbol{x}' = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ is correlated with each of the normalized vectors $\{\boldsymbol{f}'(m) = \frac{\boldsymbol{f}(m)}{\|\boldsymbol{f}(m)\|}, m = 1, ..., M\}$ to form a set of correlation

coefficients $\{C(m) = \|\boldsymbol{x}'^H \boldsymbol{f}'(m)\|^2, m = 1, ..., M\}$. Each correlation coefficient is then operated on by a nonlinear function $Z[C(m)]$. One example for $Z(\cdot)$ is

$$Z(x) = \begin{cases} 1 & \text{if } x > t \\ 0 & \text{otherwise} \end{cases} \tag{3.75}$$

where t is a threshold value. The output of the nonlinear function is then used to weight the corresponding AOA vector. The weighted AOA vectors are summed to produce an AOA estimate, i.e.

$$\hat{\boldsymbol{\theta}} = \sum_{m=1}^{M} Z[C(m)]\boldsymbol{\theta}(m). \tag{3.76}$$

In matrix form, $\hat{\boldsymbol{\theta}}$ can be expressed as

$$\hat{\boldsymbol{\theta}} = \boldsymbol{T}\boldsymbol{\Theta} \tag{3.77}$$

where

$$\boldsymbol{T} = \begin{bmatrix} Z[C(1)] & 0 & \cdots & 0 \\ 0 & Z[C(2)] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & Z[C(M))] \end{bmatrix} \tag{3.78}$$

and

$$\boldsymbol{\Theta} = \begin{bmatrix} \theta^d(1) & \theta^i(1) \\ \vdots & \vdots \\ \theta^d(M) & \theta^i(M) \end{bmatrix}. \tag{3.79}$$

It should be noted that grid search and matched-filter implementation are mathematically equivalent to each other because they are based on the same objective function. However, with this structural implementation, the cross-correlation operation is carried out in a parallel fashion, which can shorten the computation time required by

the sequential two-dimensional search. A large M may be needed for a required accuracy, but may also increase the hardware complexity of the filter to a point where the implementation is no longer practical nor feasible. In addition, it is difficult to determine the threshold t. Nevertheless, this matched-filter structure does shed some light on the neural network approach, as we shall see in the next chapter.

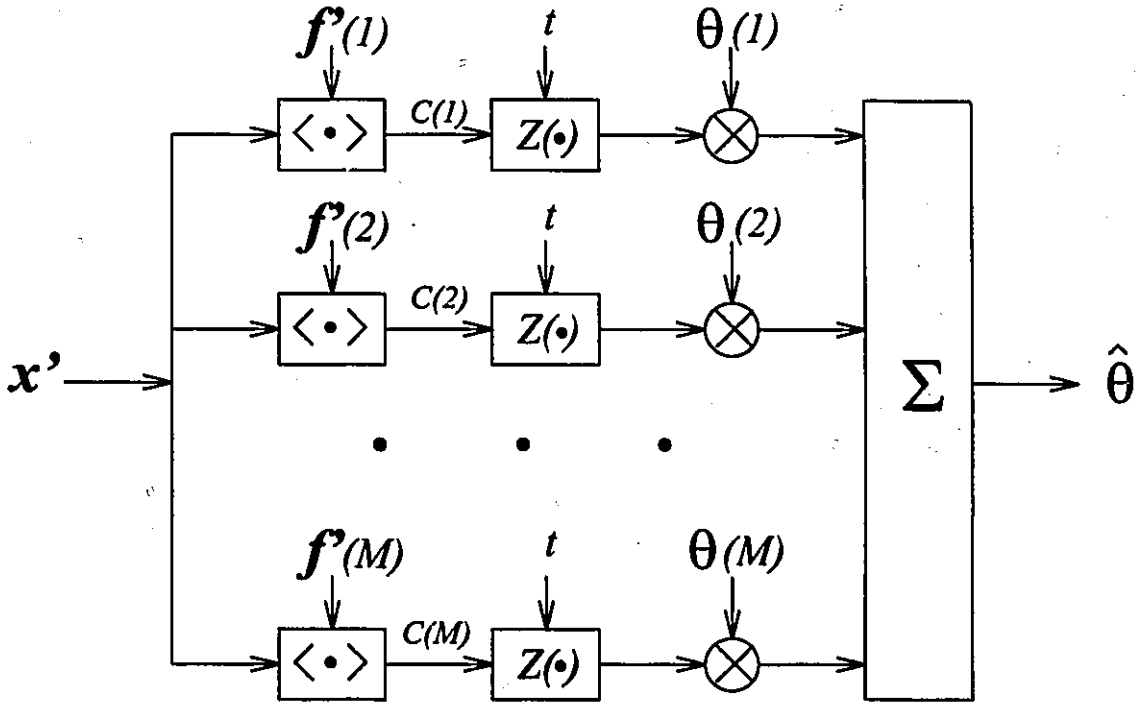

Figure 3.12: Correlating matched-filter implementation of the maximum likelihood method; $< \cdot >$ denotes the vector inner product operation.

Chapter 4

PRINCIPLES OF RBF NETWORK FOR AOA ESTIMATION

4.1 Introduction

The MLE technique is typically formulated as a nonlinear optimisation problem. Based on knowledge of the probability distribution of the measurement noise $\boldsymbol{n}$ where Gaussianity is usually employed, the AOA, $\boldsymbol{\theta}$, is estimated by optimising the likelihood function when an actual sensor array output $\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n}$ is presented. Although the MLE method has an optimum performance in the statistical sense, the computational load of the optimisation procedure is too heavy for real-time processing. Even though a parallel matched-filter implementation may be used as proposed in the previous chapter, we have already recognised that there exist a number of drawbacks associated with this implementation, such as the choice of threshold and the hardware complexity required for high resolution (small size grids). However, the correlating matched-filter implementation exhibits a special characteristic similar to that of an

emerging technology – neural networks, which inherently provide a parallel computation capability, in addition to other important properties. The connection between the matched-filter and a neural network is that the structure of a matched-filter resembles an important class of neural network model, namely the associative memory model. Because of this resemblance, we postulate that an associative memory may easily be adopted for the AOA estimation. In this chapter, we will show how this can be done.

The neural network to which we are referring here can be defined, in a formal engineering context, as a highly parallel dynamic system with the topology of direct graph that can carry out information processing by means of its state response to initial input [23]. Therefore, within this framework, we are able to study more abstract logical principles of "neural computing", without devising the detailed implementations of the underlying structure.

4.2 The Associative Memory Model

Let us consider a system shown in Fig. 4.13, which has well defined inputs and outputs. We further denote $[\boldsymbol{s}(m), \boldsymbol{\theta}(m)]$ as the m^{th} associated pair of M input-output patterns, where

$$\boldsymbol{s}(m) = [s_1(m)\ s_2(m)\ \cdots\ s_K(m)]^T \in C^K \tag{4.80}$$

is the m^{th} input pattern and

$$\boldsymbol{\theta}(m) = [\theta_1(m)\ \theta_2(m)\ \cdots\ \theta_L(m)]^T \in C^L \tag{4.81}$$

is the corresponding output pattern. It should be noted that $\boldsymbol{s}$ and $\boldsymbol{\theta}$ are assumed to be complex for a general case. The basic operation of this system is to transform the ordered set of inputs into another ordered set of outputs. When the input vector $\boldsymbol{s}(m)$ is presented to the system as a key, the system reproduces the corresponding

output vector $\boldsymbol{\theta}(m)$ through some mapping F, i.e.

$$\boldsymbol{s}(m) \xrightarrow{F} \boldsymbol{\theta}(m). \tag{4.82}$$

Accordingly, $\boldsymbol{s}(m)$ is usually referred to as the key pattern and $\boldsymbol{\theta}(m)$ the memorised pattern. The underlying operation is commonly known as associative recall and the system is referred to as an associative memory. Three types of associative memories have been distinguished in the literature [49]. They are:

1. Heteroassociative memory: which implements a mapping, F, of $\boldsymbol{s}$ to $\boldsymbol{\theta}$ such that $F(\boldsymbol{s}(m)) = \boldsymbol{\theta}(m)$, and if an arbitrary $\boldsymbol{s}$ is "closer" to $\boldsymbol{s}(m)$ than any other, $m = 1, \cdots, M$, then $F(\boldsymbol{s}) = \boldsymbol{\theta}(m)$.

2. Interpolative associative memory: which implements a mapping, F, of $\boldsymbol{s}$ to $\boldsymbol{\theta}$ such that $F(\boldsymbol{s}(m)) = \boldsymbol{\theta}(m)$. Furthermore, if the input vector differs from $\boldsymbol{s}(m)$ by the vector $\boldsymbol{n}$, i.e. $\boldsymbol{x} = \boldsymbol{s}(m) + \boldsymbol{n}$, then the output of the associative memory also differs from memorised pattern by some vector $\boldsymbol{\varepsilon}$, i.e. $F(\boldsymbol{x}) = F(\boldsymbol{s}(m) + \boldsymbol{n}) = \boldsymbol{\theta}(m) + \boldsymbol{\varepsilon}$.

3. Autoassociative memory: which assumes $\boldsymbol{\theta}(m) = \boldsymbol{s}(m)$ and implements a mapping, F, of $\boldsymbol{s}$ to $\boldsymbol{s}$ such that $F(\boldsymbol{s}(m)) = \boldsymbol{s}(m)$, and, if some arbitrary $\boldsymbol{s}$ is "closer" to $\boldsymbol{s}(m)$ than any other, $m = 1, \cdots, M$, then $F(\boldsymbol{s}) = \boldsymbol{s}(m)$.

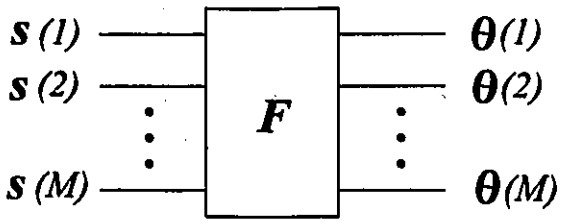

Figure 4.13: A generic input-output system.

An associative memory can be applied to the AOA estimation. In particular, $\boldsymbol{s}$ can be considered as the complex output of the antenna array, which can be used

by the associative memory processor as a key to retrieving the corresponding AOA vector $\boldsymbol{\theta}$, which is real in this particular case. Since $\boldsymbol{\theta}$ and $\boldsymbol{s}$ are different quantities, it is obvious that the autoassociative memory is not suitable for this application. Either the heteroassociative memory or the interpolative associative memory may be used for AOA estimation.

When the heteroassociative memory is used for the estimation, it will retrieve only the discrete AOA values $\{\boldsymbol{\theta}(m), m = 1, 2, \cdots, M\}$ as estimates. Obviously, the value of M should be large if a reasonably accurate estimate is to be obtained. To a large extent, the heteroassociative memory is very similar to the correlating matched-filter implementation of the MLE method, in that they both produce discrete AOA estimates and require a large value of M. Building such a memory is not a difficult task mathematically if we place the restriction that the vectors, $\boldsymbol{s}(m)$, form an orthonormal set, i.e.,

$$\boldsymbol{s}^H(m)\boldsymbol{s}(n) = \delta_{mn} = \begin{cases} 1 & if\ m = n \\ 0 & if\ m \neq n. \end{cases} \tag{4.83}$$

The orthogonal set can be obtained using the Gram-Schmit orthogonalisation procedure. To build a linear associative memory for AOA estimation, we consider

$$F(\boldsymbol{s}) = \left[\boldsymbol{\theta}(1)\boldsymbol{s}^H(1) + \boldsymbol{\theta}(2)\boldsymbol{s}^H(2) + ... + \boldsymbol{\theta}(M)\boldsymbol{s}^H(M)\right]\boldsymbol{s}. \tag{4.84}$$

If $\boldsymbol{s} = \boldsymbol{s}(m)$ is the input vector, then $F(\boldsymbol{s}(m)) = \boldsymbol{\theta}(m)$. That is,

$$\begin{aligned} \boldsymbol{\theta} &= \left[\boldsymbol{\theta}(1)\boldsymbol{s}^H(1) + \boldsymbol{\theta}(2)\boldsymbol{s}^H(2) + ... + \boldsymbol{\theta}(M)\boldsymbol{s}^H(M)\right]\boldsymbol{s}(m) \\ &= \boldsymbol{\theta}(1)\delta_{1m} + \boldsymbol{\theta}(2)\delta_{2m} + ... + \boldsymbol{\theta}(M)\delta_{Mm} \\ &= \boldsymbol{\theta}(m). \end{aligned} \tag{4.85}$$

The result is then a perfect recall of $\boldsymbol{\theta}(m)$. Note that (4.85) can be rewritten as

$$\boldsymbol{\theta}(m) = \boldsymbol{T\Theta} \tag{4.86}$$

where

$$\boldsymbol{T} = \begin{bmatrix} \delta_{1m} & 0 & \cdots & 0 \\ 0 & \delta_{2m} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & \delta_{Mm} \end{bmatrix} \tag{4.87}$$

and

$$\boldsymbol{\Theta} = \begin{bmatrix} \theta^1(1) & \theta^2(1) & \cdots & \theta^L(1) \\ \theta^1(2) & \theta^2(2) & \cdots & \theta^L(2) \\ \vdots & \vdots & \ddots & \vdots \\ \theta^1(M) & \theta^2(M) & \cdots & \theta^L(M) \end{bmatrix}. \tag{4.88}$$

By comparing (4.86), which describes the heteroassociative memory, with (3.77) , which describes the matched-filter implementation of MLE, it is clear that they are of the same form. The only difference is that the diagonal transformation matrix $\boldsymbol{T}$ in (4.86) results from the orthonormality whereas $\boldsymbol{T}$ in (3.77) is the product of a set of thresholding functions. In spite of this difference, (4.86) and (3.77) are equivalent from the functionality point of view. In fact, a number of matched-filter types of associative memories have been proposed and investigated [50, 51]. Furthermore, there was no training involved in the definition of the linear associator. The function that mapped $\boldsymbol{s}$ into $\boldsymbol{\theta}$ is defined by the mathematical expression in (4.84).

A linear associative memory may be regarded as the lowest-order approach to the problem of associative transformations, wherein a mathematical mapping F is to be found to relate arbitrary output patterns to arbitrary input patterns. One of the requirements for the linear associative memory is that the dimension of the input pattern ($\boldsymbol{s}$), M, must be equal to or greater than the number of output patterns, L. When $M < L$, the column vectors in $\boldsymbol{\Theta}$ become linearly dependent and no exact solution exists. As well, it is believed that the size of the training set can

be reduced if the ability to interpolate is added to the memory. For these reasons, there arises a need to generalise the class of associative mappings. Using nonlinear transformations, it is possible to implement selective associative mappings for patterns which on account of their linear dependence would not be distinguishable if subjected to a linear transformation [52]. A mathematical justification for this rationale may also be traced back to the work by Cover [53], where a pattern-classification problem cast in a high-dimensional space nonlinearly is more likely to be linearly separable than in a low-dimensional space. For the AOA estimation application, the need for a nonlinear associative mapping becomes essential. The AOA mapping H given by (2.31) is a highly nonlinear mapping, and hence the inverse mapping F is also expected to be highly nonlinear if F exists. At this point of our discussion, we assume that F exists. The existence of F will be discussed in Chapter 5. A nonlinear neural network is therefore required to form the AOA associative memory as shown in (4.80). In the next section is presented a class of nonlinear neural networks that can be used to achieve the required transformation.

The nonlinear neural network that we will use to form the associative memory is the radial basis function (RBF) neural network, which has been studied extensively. It has been shown that an RBF network can be derived from the regularisation theory [27]. As a result, it can approximate continuous functions arbitrarily well [29]. That is, an RBF network is a universal approximator[28]. Furthermore, An RBF network has the best approximation property [29]. This means that there exits a set of coefficients for the network to approximate a given unknown function better than all other possible sets. These properties make the RBF network an attractive candidate for constructing the inverse mapping F.

4.3 The RBF Networks

A discriminant function $g(\boldsymbol{x})$ can be represented by the weighted summation of a finite number of potential functions [54] as follows

$$g(\boldsymbol{x}) = \sum_{i=1}^{M} w_i P(\boldsymbol{x}, \boldsymbol{x}_i) \tag{4.89}$$

where $P(\boldsymbol{x}, \boldsymbol{x}_i)$ is the i^{th} potential function of $\boldsymbol{x}$, obtained by shifting $P(\boldsymbol{x}, 0)$ by $\boldsymbol{x}_i$, and w_i is a real constant. For instance, the potential function $P(\boldsymbol{x}, \boldsymbol{x}_i)$ in classical physics varies inversely with $\|\boldsymbol{x} - \boldsymbol{x}_i\|$, i.e. $P(\boldsymbol{x}, \boldsymbol{x}_i)$ has a maximum value at $\boldsymbol{x} = \boldsymbol{x}_i$ and decreases monotonically to zero as $\|\boldsymbol{x} - \boldsymbol{x}_i\|$ approaches infinity. The potential function can be selected based on the following condition:

$$P(\boldsymbol{x}, \boldsymbol{x}_i) = \sum_{j=1}^{\infty} \phi_j(\boldsymbol{x})\phi_j(\boldsymbol{x}_i) \quad \text{and} \quad \max_{\boldsymbol{x}} P(\boldsymbol{x}, \boldsymbol{x}_i) \leq b \tag{4.90}$$

where the function system $\{\phi_j(\boldsymbol{x})\}$ is orthonormal in the space of $\boldsymbol{x}$ and b is a bounded scalar. The potential function approach to the expansion of a function described by (4.89) takes a form similar to that of the nonparametric estimation of a probability density function based on the Parzen window [55]. In the Parzen-window approach, a probability density function, $p(\boldsymbol{x})$, is estimated from the observed input samples, $\boldsymbol{x}_i$'s, $i = 1, \cdots, n$, by

$$p_M(\boldsymbol{x}) = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{h_M^K} \phi\left(\frac{\boldsymbol{x} - \boldsymbol{x}_i}{h_M}\right) \tag{4.91}$$

where ϕ represents a bounded nonnegative kernel function of the K-dimensional input vector, $\boldsymbol{x}$, and h_M is a sequence of positive numbers such that $\lim_{M\to\infty} h_M = 0$ and $\lim_{M\to\infty} M h_M^K = \infty$. It can be shown from (4.91) that $p_M(\boldsymbol{x})$ converges to $p(\boldsymbol{x})$ as M approaches infinity.

A more general form for (4.89) or (4.91), incorporating the adjustment of shape parameters, can be expressed as

$$g(\boldsymbol{x}) = \sum_{i=1}^{M} w_i \phi(\boldsymbol{x}, \boldsymbol{c}_i, \sigma_i) \tag{4.92}$$

where M represents the number of kernel functions, w_i represents the summation weight, and $\boldsymbol{c}_i$ and σ_i represent new parameter vectors to include the position shift and the shape parameters of the i^{th} kernel function, respectively. In (4.92), w_i, $\boldsymbol{c}_i$ and σ_i, for $i = 1, \cdots, M$, are subject to adjustment through a learning process. According to Funahashi [56] and Hornit, Stinchcombe, and White [57], (4.92) can approximate a continuous function with a desirable degree of accuracy based on a sufficiently large number of kernels, provided that ϕ is an absolutely integrable or a bounded monotonic function. That is, the function ϕ should satisfy the following conditions.

1. The kernel function is bounded, i.e.

$$L_\phi \leq \phi(\boldsymbol{x}, \boldsymbol{c}, \sigma) \leq U_\phi. \tag{4.93}$$

   where L_ϕ and U_ϕ represent the lower and upper bounds of $\phi(\boldsymbol{x}, \boldsymbol{c}, \sigma)$ respectively.

2. The kernel function is localised, i.e.

$$\int_{-\infty}^{\infty} (\frac{1}{\sigma})^N \phi(\boldsymbol{x}, \boldsymbol{c}, \sigma) d\boldsymbol{x} = C \tag{4.94}$$

$$\lim_{\sigma \to 0} (\frac{1}{\sigma})^N \phi(\boldsymbol{x}, \boldsymbol{c}, \sigma,) = C\delta(\boldsymbol{x}) \tag{4.95}$$

   where N is the dimension of $\boldsymbol{x}$ and C is a positive constant.

Amongst many functions that satisfy the above two conditions, the Gaussian function is the most widely used, i.e.,

$$\phi(\boldsymbol{x}, \boldsymbol{c}, \sigma) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{c}\|^2}{\sigma^2}}. \tag{4.96}$$

The expansion of the function $g(\boldsymbol{x})$ given in (4.92) in terms of radial basis functions $\phi(\boldsymbol{x}, \boldsymbol{c}_i, \sigma_i)$ suggests the network structure shown in Fig. 4.14 as a method for its implementation. It consists of three entirely different layers, namely the input layer, hidden layer, and the output layer. The input layer is made of source nodes whose number is equal to the dimension of the input vector $\boldsymbol{x}$. The hidden layer is composed

of nonlinear units that are connected directly to all of the nodes in the input layer. This layer is usually of higher dimension than the input layer, i.e. the number of hidden units is larger than the number of source units. The output layer consists of linear units, being fully connected to the hidden layer. By "linearity" we mean that the outputs of the network are linearly weighted sums of the outputs of the hidden units. The output layer supplies the response of the network to the activation patterns applied to the input layer.

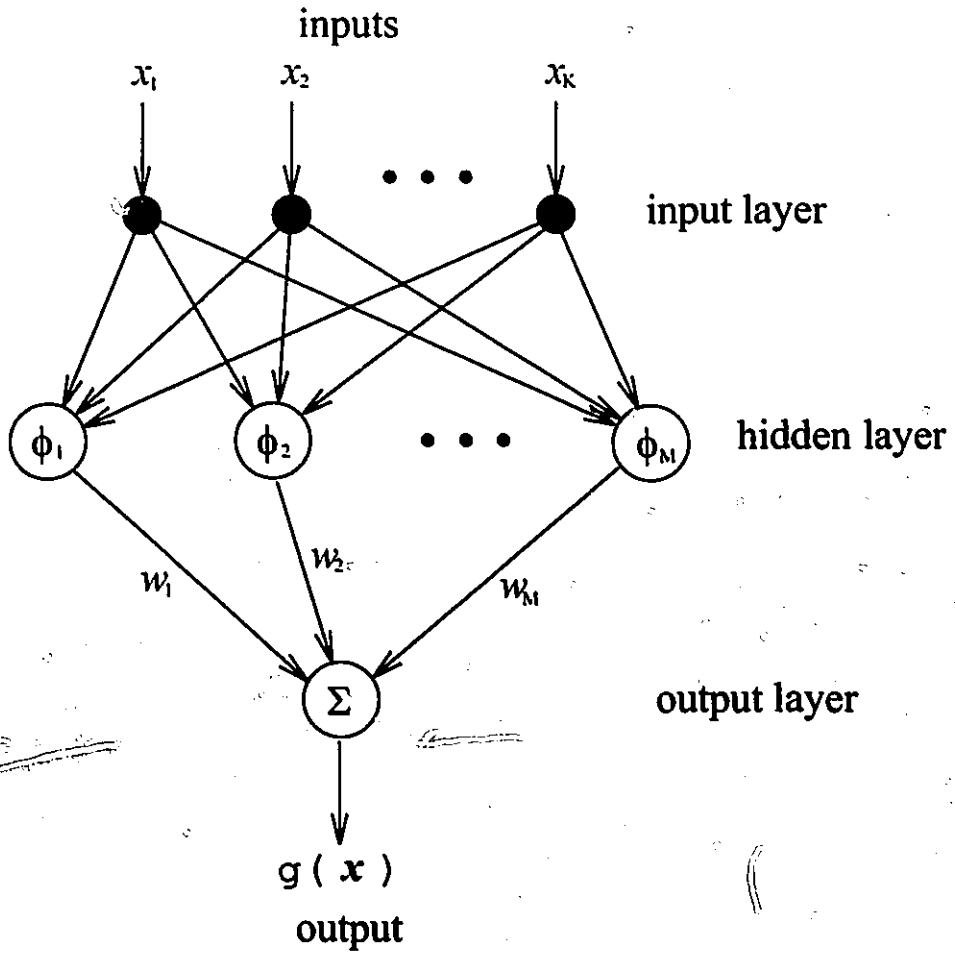

Figure 4.14: Radial basis-function network.

4.4 Associative Memory Using the RBF Network

Having described the RBF network, we proceed to discuss the implementation of an associative memory using the RBF network. Given a set of associated pairs $\{[\boldsymbol{s}(m), \boldsymbol{\theta}(m)] \mid m = 1, \cdots, M\}$, the mapping for these pairs can be represented using the RBF networks. If the mapping must strictly represent all these pairs, the mapping function that is constructed using the RBF network is constrained to pass through every associated pair. This constraint is often termed as the strictly interpolative condition. In this strict sense, all the M pairs are used to generate the mapping function, i.e.

$$\theta^l(j) = F[\boldsymbol{s}(j)] = \sum_{m=1}^{M} w_m^l \phi[\boldsymbol{s}(j), \boldsymbol{s}(m), \sigma_m] \quad l = 1, \cdots, L, \; j = 1, \cdots, M \tag{4.97}$$

where

$$\phi[\boldsymbol{s}(j), \boldsymbol{s}(m), \sigma_m] = e^{-\frac{\|\boldsymbol{s}(j) - \boldsymbol{s}(m)\|^2}{\sigma_m^2}}. \tag{4.98}$$

This means that the RBF is expanded on a finite set of M basis functions, which arise from ϕ being centred on the M data points. The expansion can be implemented using the network structure shown in Fig. 4.15. Defining the matrices $\boldsymbol{\Theta}$, $\boldsymbol{W}$ and the symmetric matrix $\boldsymbol{\Phi}$ as follows

$$\boldsymbol{\Theta} = \begin{bmatrix} \theta^1(1) & \theta^2(1) & \cdots & \theta^L(1) \\ \theta^1(2) & \theta^2(2) & \cdots & \theta^L(2) \\ \vdots & \vdots & \ddots & \vdots \\ \theta^1(M) & \theta^2(M) & \cdots & \theta^L(M) \end{bmatrix}, \tag{4.99}$$

$$\boldsymbol{\Phi} = \begin{bmatrix} \phi(\boldsymbol{s}(1), \boldsymbol{s}(1), \sigma_1) & \phi(\boldsymbol{s}(1), \boldsymbol{s}(2), \sigma_2) & \cdots & \phi(\boldsymbol{s}(1), \boldsymbol{s}(M), \sigma_M) \\ \phi(\boldsymbol{s}(2), \boldsymbol{s}(1), \sigma_1) & \phi(\boldsymbol{s}(2), \boldsymbol{s}(2), \sigma_2) & \cdots & \phi(\boldsymbol{s}(2), \boldsymbol{s}(M), \sigma_M) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(\boldsymbol{s}(M), \boldsymbol{s}(1), \sigma_1) & \phi(\boldsymbol{s}(M), \boldsymbol{s}(2), \sigma_2) & \cdots & \phi(\boldsymbol{s}(M), \boldsymbol{s}(M), \sigma_M) \end{bmatrix}, \tag{4.100}$$

and

$$\boldsymbol{W} = \begin{bmatrix} w_1^1 & w_1^2 & \cdots & w_1^L \\ w_2^1 & w_2^2 & \cdots & w_2^L \\ \vdots & \vdots & \ddots & \vdots \\ w_M^1 & w_M^2 & \cdots & w_M^L \end{bmatrix}, \tag{4.101}$$

we may rewrite (4.97) as

$$\boldsymbol{\Theta} = \boldsymbol{\Phi W}. \tag{4.102}$$

$\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$ are usually referred to as the memorised matrix and the key matrix, respectively.

The Gaussian RBF are nonzero only in a small neighbourhood of the centre $\boldsymbol{s}(m)$ or in other words each hidden unit has its own receptive field. That is, the hidden unit will be activated when the input $\boldsymbol{s}$ lies in the set $\{\boldsymbol{s} : \|\boldsymbol{s} - \boldsymbol{s}(m)\| < \delta\}$, and the size of the receptive field will be proportional to the shape parameter σ_m. In constructing the RBF associative memory, it is tempting to first follow the same strategy for building a linear associative memory, i.e. $w_m^l = \theta^l(m)$. The content $\theta^l(j)$ stored in the memory can be retrieved by

$$\theta^l(j) = \sum_{m=1}^{M} \theta^l(m) e^{-\frac{\|\boldsymbol{s}(j)-\boldsymbol{s}(m)\|^2}{\sigma_m^2}}. \tag{4.103}$$

To be more explicit,

$$\begin{aligned} \theta^l(j) \;=\; & \theta^l(1)e^{-\frac{\|\boldsymbol{s}(j)-\boldsymbol{s}(1)\|^2}{\sigma_1^2}} + \cdots + \theta^l(j-1)e^{-\frac{\|\boldsymbol{s}(j)-\boldsymbol{s}(j-1)\|^2}{\sigma_{j-1}^2}} + \theta^l(j) \\ & + \theta^l(j+1)e^{-\frac{\|\boldsymbol{s}(j)-\boldsymbol{s}(j+1)\|^2}{\sigma_{j+1}^2}} + \cdots + \theta^l(M)e^{-\frac{\|\boldsymbol{s}(j)-\boldsymbol{s}(M)\|^2}{\sigma_M^2}}. \end{aligned} \tag{4.104}$$

When σ_m is set to be very small, all the exponential terms approach zero. A perfect recall is then followed. Therefore, the RBF interpolative associative memory given by (4.103) approaches a perfect recall if the Gaussian function is chosen to have a sufficiently high degree of localization.

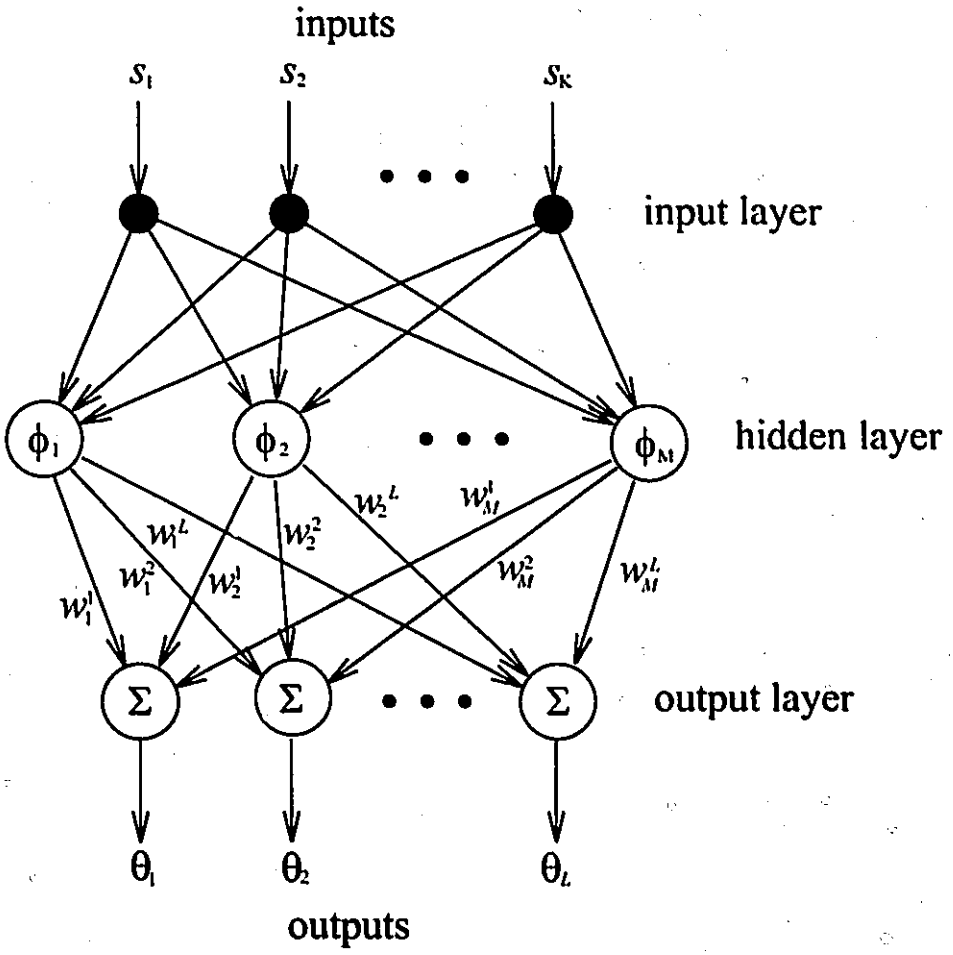

Figure 4.15: Implementation of an associative memory using the radial basis-function network.

Although the above RBF associative memory is plausible, at least in principle, it is essentially a linear associative memory and exhibits the same drawbacks, which we tried to overcome in the first place. In order for the RBF to learn the underlying rules of the inverse AOA mapping F from the training pairs, we have to solve the linear system represented by (4.102) for $\boldsymbol{W}$. This requires the following property [58].

Property 4.1 *Let* $\boldsymbol{s}(1)$, $\boldsymbol{s}(2)$, $\cdots$, $\boldsymbol{s}(M)$ *be* M *distinct points in* $\Re^K$. *Then the* $M \times M$ *interpolation matrix* $\boldsymbol{\Phi}$ *whose* mn^{th} *element is* $\phi_{nm} = \phi(\|\boldsymbol{s}(m) - \boldsymbol{s}(n)\|)$ *is positive definite.*

This property is obviously applicable to the Gaussian function. The implication of the property is that $\boldsymbol{\Phi}$ is invertible. Therefore, we are able to obtain the weight matrix by taking the inverse of the matrix $\boldsymbol{\Phi}$. That is,

$$\boldsymbol{W} = \boldsymbol{\Phi}^{-1}\boldsymbol{\Theta} \tag{4.105}$$

If the input to the associative memory is $\boldsymbol{s} = \boldsymbol{s}(m)$, based on (4.102), the output is given by

$$\boldsymbol{\theta}^T = \left[e^{-\frac{\|\boldsymbol{S}(m)-\boldsymbol{S}(1)\|^2}{\sigma_1^2}}, \cdots, e^{-\frac{\|\boldsymbol{S}(m)-\boldsymbol{S}(M)\|^2}{\sigma_M^2}} \right] \boldsymbol{W} \tag{4.106}$$

Substituting (4.105) into (4.106), we have

$$\boldsymbol{\theta}^T = \left[e^{-\frac{\|\boldsymbol{S}(m)-\boldsymbol{S}(1)\|^2}{\sigma_1^2}}, \cdots, e^{-\frac{\|\boldsymbol{S}(m)-\boldsymbol{S}(M)\|^2}{\sigma_M^2}} \right] \boldsymbol{\Phi}^{-1}\boldsymbol{\Theta} \tag{4.107}$$

Since the row vector in the square bracket on the right hand side of (4.107) is just the m^{th} row of $\boldsymbol{\Phi}$, (4.107) becomes

$$\boldsymbol{\theta}^T = [0, \cdots, 1, 0, \cdots, 0]\boldsymbol{\Theta} = \boldsymbol{\theta}^T(m) \tag{4.108}$$

where 1 in the row vector is at the m^{th} column position, and hence a perfect recall follows.

4.5 Interpolation and Approximation

So far we have only dealt with the perfect recall cases, i.e. the input vector is identical to one of the training vectors. One would then ask what would happen if the input vector $\boldsymbol{x}$ is different from the m^{th} training vector $\boldsymbol{s}(m)$ by a small quantity $\Delta\boldsymbol{s}$. As shown in Fig. 4.16, we define

$$\boldsymbol{x} = \boldsymbol{s}(m) + \Delta\boldsymbol{s} \tag{4.109}$$

where

$$\Delta\boldsymbol{s} = b[\boldsymbol{s}(m+1) - \boldsymbol{s}(m)] \tag{4.110}$$

and

$$0 \leq b \leq 1. \tag{4.111}$$

As has been discussed, the interpolative associative memory approximates the inverse mapping

$$F(\boldsymbol{x}) = F(\boldsymbol{s}(m) + \Delta\boldsymbol{s}) = \boldsymbol{\theta}_x \tag{4.112}$$

with an estimate

$$\hat{F}(\boldsymbol{x}) = \hat{F}(\boldsymbol{s}(m) + \Delta\boldsymbol{s}) = \boldsymbol{\theta}(m) + \boldsymbol{\varepsilon}. \tag{4.113}$$

Let us define the approximation error to be

$$\epsilon_a = |F(\boldsymbol{x}) - \hat{F}(\boldsymbol{x})|. \tag{4.114}$$

In the linear associative memory case, it follows from (4.84) and (4.85) that

$$\boldsymbol{\varepsilon} = \left[\boldsymbol{\theta}(1)\boldsymbol{s}^T(1) + \boldsymbol{\theta}(2)\boldsymbol{s}^T(2) + ... + \boldsymbol{\theta}(M)\boldsymbol{s}^T(M)\right]\Delta\boldsymbol{s}. \tag{4.115}$$

There are two observations we may make from (4.115).

1. Substituting (4.110) into (4.115) and using (4.84) and (4.85), we obtain

$$\boldsymbol{\varepsilon} = b[\boldsymbol{\theta}(m+1) - \boldsymbol{\theta}(m)]. \tag{4.116}$$

As expected, the linear associative memory provides a linear interpolation if an input point falls in-between two training points. That is, the linear associative memory provides the first order approximation. The magnitude of ϵ_a, therefore, depends on the degree of nonlinearity of the mapping which is to be approximated. Obviously, the higher degree of nonlinearity the mapping has, the larger the value of ϵ_a is.

2. As b approaches zero, the output of the linear associative memory will produce a perfect recall, i.e.

$$\lim_{\|\Delta \boldsymbol{s}\| \to 0} \|\boldsymbol{\epsilon}\| = 0 \tag{4.117}$$

and

$$\lim_{\|\Delta \boldsymbol{s}\| \to 0} \boldsymbol{\theta}_x = \boldsymbol{\theta}(m). \tag{4.118}$$

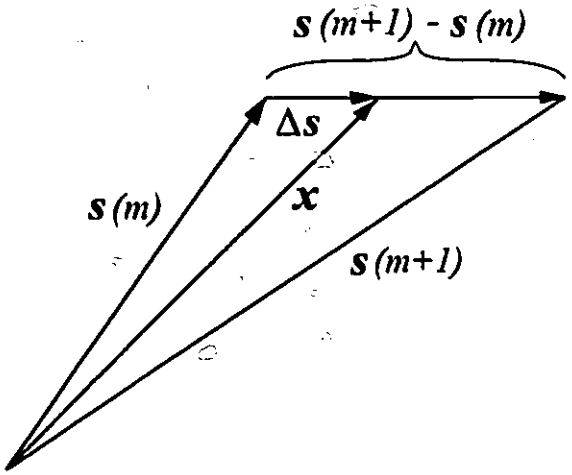

Figure 4.16: Geometrical interpretation of interpolation in the input vector space.

For the RBF network, the evaluation of $\boldsymbol{\epsilon}$ is not as explicit as in the linear case. Fortunately, the properties of the approximation error ϵ_a produced by the RBF network when used to approximate a continuous function have been well studied and documented. The approximation error or the interpolation error given by the RBF

network can be arbitrarily small if the function to be approximated is a continuous function and defined on a compact set [59, 27, 29, 28]. That is , if $F(\boldsymbol{x})$ is a continuous function and defined on a compact set, its approximation $\hat{F}(\boldsymbol{x})$ obtained using the RBF network would be arbitrarily close to the original function.

4.6 Using the Generalised RBF Networks

The RBF formulation under the strict interpolation condition is prohibitively expensive to implement in terms of computations. Its complexity grows polynomially with M, since an $M \times M$ matrix has to be inverted to find the coefficients of the expansion. In the AOA application the value of M can be large, and the inversion of such a large matrix is not only formidable but could also be meaningless since the likelihood of ill-conditioning is higher for a larger matrix. To overcome the computational difficulty, the strict interpolation condition can be relaxed by making the number of centres, N, lower than the number of data points, M. The co-ordinates of the centres are allowed to be chosen arbitrarily [30]. This class of RBF networks are usually referred to as the generalised RBF networks and their structure is shown in Fig. 4.17. To further reduce the intensive computations required for the determination of the value of σ_m for each centres, which involves a high-dimensional nonlinear optimisation, a set of widths with the pre-determined values can be used. This is, of course, at the expense of the optimality. That is, we may only obtain sub-optimal estimation. In this case, denoting with $\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_N$ the co-ordinates of the N centres ($N < M$) the RBF network represents the following approximation,

$$\boldsymbol{\Theta} \simeq \tilde{\boldsymbol{\Phi}} \tilde{\boldsymbol{W}} \tag{4.119}$$

where Θ is defined by (4.99) and $\tilde{W}$ and $\tilde{\Phi}$ are defined as

$$\tilde{\Phi} = \begin{bmatrix} \phi(s(1), c_1, \sigma) & \phi(s(1), c_2, \sigma) & \cdots & \phi(s(1), c_N, \sigma) \\ \phi(s(2), c_1, \sigma) & \phi(s(2), c_2, \sigma) & \cdots & \phi(s(2), c_N, \sigma) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(s(M), c_1, \sigma) & \phi(s(M), c_2, \sigma) & \cdots & \phi(s(M), c_N, \sigma) \end{bmatrix} \tag{4.120}$$

and

$$\tilde{W} = \begin{bmatrix} w_1^1 & w_1^2 & \cdots & w_1^L \\ w_2^1 & w_2^2 & \cdots & w_2^L \\ \vdots & \vdots & \ddots & \vdots \\ w_N^1 & w_N^2 & \cdots & w_N^L \end{bmatrix}. \tag{4.121}$$

The construction of the associative memory becomes finding the linear mapping $\tilde{W}$ that minimises the Euclidean norm of a matrix defined as the difference between the desired response Θ and the actual response $\tilde{\Phi}\tilde{W}$. The mapping $\tilde{W}$ as postulated here is referred to as the least-squares (LS) solution to the pattern association problem. The LS solution for $\tilde{W}$ can be written as

$$\tilde{W} = \tilde{\Phi}^{+}\Theta \tag{4.122}$$

where $\tilde{\Phi}^{+}$ is the pseudo-inverse of the key matrix $\tilde{\Phi}$, given by

$$\tilde{\Phi}^{+} = (\tilde{\Phi}^T\tilde{\Phi})^{-1}\tilde{\Phi}^T. \tag{4.123}$$

The advantage of using a rectangular matrix is that adaptive algorithms such as the Least Mean Squares (LMS) algorithm and the Recursive Least Squares (RLS) algorithm [60] can be used to estimate the weights. This further reduces the computational time and the hardware complexity for using this model in signal processing applications. However, these advantages come at the cost of the approximation capability of the network. For example, the best approximation property of the network

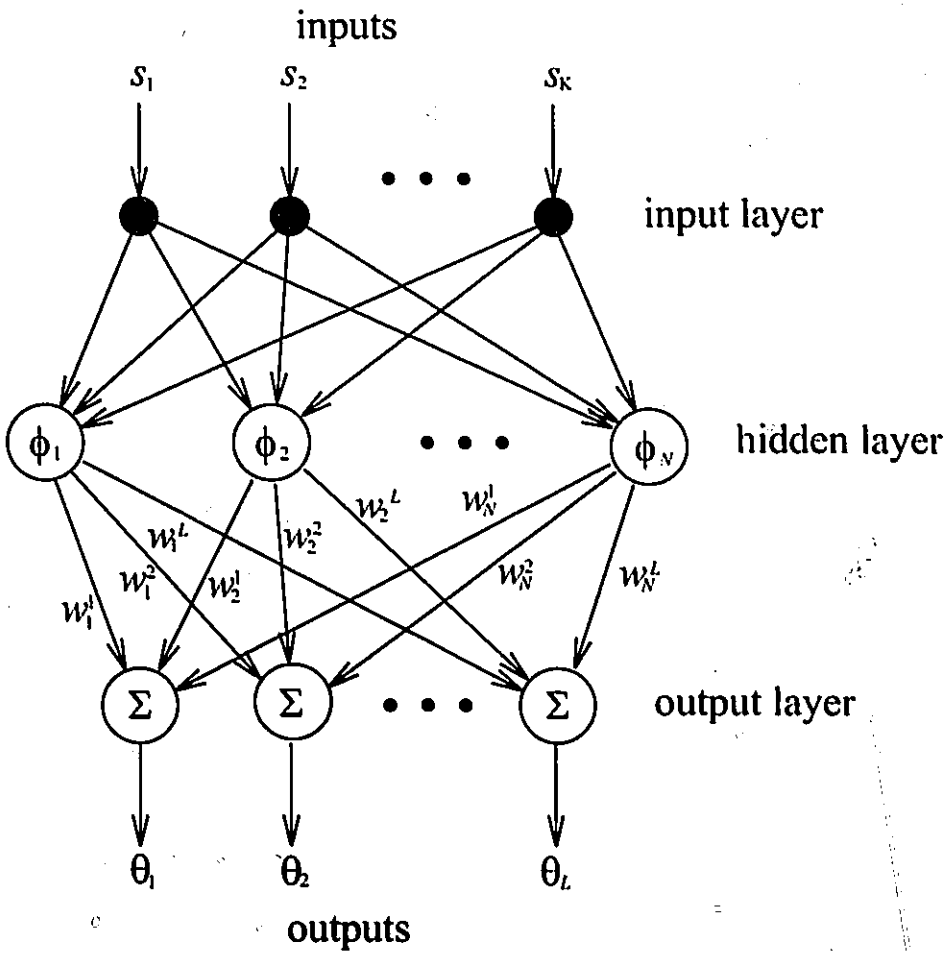

Figure 4.17: Structure of the generalised radial basis-function network.

will be lost [29], a perfect recall will no longer be achievable, and the interpolation accuracy will be degraded. It should be pointed out, however, that the generalized RBF network has almost the best approximation property provided that the centres are fixed to a subset of positions [29]

If the training data are used to test the LS-RBF network, it will produce the estimates

$$\hat{\Theta} = \tilde{\Phi}\tilde{W} = \tilde{\Phi}(\tilde{\Phi}^T\tilde{\Phi})^{-1}\tilde{\Phi}^T\Theta. \tag{4.124}$$

Proposition 4.1 *Let us partition both Φ and W in (4.102) as follows,*

$$\Phi = \left[\tilde{\Phi} \mid \breve{\Phi}\right] \tag{4.125}$$

and

$$W = \left[\begin{array}{c} W_1 \\ \hline W_2 \end{array}\right] \tag{4.126}$$

where $\Phi \in \Re^{M\times M}$, $\tilde{\Phi} \in \Re^{M\times N}$, $\breve{\Phi} \in \Re^{M\times(M-N)}$, $W_1 \in \Re^{N\times L}$, and $W_2 \in \Re^{(M-N)\times L}$. Furthermore, suppose that Φ has M singular values $\{\lambda_1 > \lambda_2 > \cdots > \lambda_M\}$. Then the 2-norm of the difference between Θ and $\hat{\Theta}$ is given by

$$\|\Theta - \hat{\Theta}\|_2 \leq \lambda_{N+1}\|W_2\|_2. \tag{4.127}$$

Proof Using (4.124), we write

$$\Theta - \hat{\Theta} = \Theta - \tilde{\Phi}(\tilde{\Phi}^T\tilde{\Phi})^{-1}\tilde{\Phi}^T\Theta. \tag{4.128}$$

Since $\Theta = \Phi W$,

$$\Theta - \hat{\Theta} = \Phi W - \tilde{\Phi}(\tilde{\Phi}^T\tilde{\Phi})^{-1}\tilde{\Phi}^T\Phi W. \tag{4.129}$$

Expressing $\boldsymbol{\Phi} = \left[\dot{\boldsymbol{\Phi}} \mid \breve{\boldsymbol{\Phi}}\right]$, it follows that

$$\begin{aligned}
\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}} &= \left(\left[\dot{\boldsymbol{\Phi}} \mid \breve{\boldsymbol{\Phi}}\right] - \dot{\boldsymbol{\Phi}}(\dot{\boldsymbol{\Phi}}^T\dot{\boldsymbol{\Phi}})^{-1}\dot{\boldsymbol{\Phi}}^T\left[\dot{\boldsymbol{\Phi}} \mid \breve{\boldsymbol{\Phi}}\right]\right)\boldsymbol{W} \\
&= \left(\left[\dot{\boldsymbol{\Phi}} \mid \breve{\boldsymbol{\Phi}}\right] - \left[\dot{\boldsymbol{\Phi}} \mid \dot{\boldsymbol{\Phi}}(\dot{\boldsymbol{\Phi}}^T\dot{\boldsymbol{\Phi}})^{-1}\dot{\boldsymbol{\Phi}}^T\breve{\boldsymbol{\Phi}}\right]\right)\boldsymbol{W} \\
&= \left[\boldsymbol{O} \mid \left(\breve{\boldsymbol{\Phi}} - \dot{\boldsymbol{\Phi}}(\dot{\boldsymbol{\Phi}}^T\dot{\boldsymbol{\Phi}})^{-1}\dot{\boldsymbol{\Phi}}^T\breve{\boldsymbol{\Phi}}\right)\right]\boldsymbol{W} \\
&= \left[\boldsymbol{O} \mid \left(\boldsymbol{I} - \dot{\boldsymbol{\Phi}}(\dot{\boldsymbol{\Phi}}^T\dot{\boldsymbol{\Phi}})^{-1}\dot{\boldsymbol{\Phi}}^T\right)\breve{\boldsymbol{\Phi}}\right]\boldsymbol{W}
\end{aligned} \tag{4.130}$$

where $\boldsymbol{O}$ is an $M \times N$ null matrix. Taking the norm of the difference and using the relation $\|\boldsymbol{I} - \dot{\boldsymbol{\Phi}}(\dot{\boldsymbol{\Phi}}^T\dot{\boldsymbol{\Phi}})^{-1}\dot{\boldsymbol{\Phi}}^T\| = \min(M-N, 1)$ [61], we have

$$\begin{aligned}
\|\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}}\|_2 &= \left\|\left[\boldsymbol{O} \mid \left(\boldsymbol{I} - \dot{\boldsymbol{\Phi}}(\dot{\boldsymbol{\Phi}}^T\dot{\boldsymbol{\Phi}})^{-1}\dot{\boldsymbol{\Phi}}^T\right)\breve{\boldsymbol{\Phi}}\right]\left[\frac{\boldsymbol{W}_1}{\boldsymbol{W}_2}\right]\right\|_2 \\
&= \min(M-N, 1)\|\breve{\boldsymbol{\Phi}}\boldsymbol{W}_2\|_2 \\
&\leq \min(M-N, 1)\|\breve{\boldsymbol{\Phi}}\|\ \|\boldsymbol{W}_2\|_2.
\end{aligned} \tag{4.131}$$

Since $M - N \geq 1$ in our case and $\|\breve{\boldsymbol{\Phi}}\|_2 = \lambda_{N+1}$, the inequality (4.127) follows. Q.E.D.

According to Proposition 4.1, the norm of the AOA estimation error given by the generalised RBF network is bounded by the quantity that is proportional to the $(N+1)^{th}$ singular value of $\boldsymbol{\Phi}$. That is, if λ_{N+1} is small, the norm of the AOA estimation error will be small.

Chapter 5

APPLICATION OF RBF NETWORK TO AOA ESTIMATION

5.1 Existence of the Inverse Mapping

The AOA mapping H given by (2.31) is a highly nonlinear function, and hence its inverse mapping F is also highly nonlinear if it exists. Since H is a very complicated nonlinear function, it would be very difficult, if not impossible, to prove, in a strictly mathematical sense, that its inverse mapping F exists. However, we are able to infer from the following arguments that F does exist within a valid region.

1. In Fig. 5.18 is shown a simple AOA case where a plane wave impinges upon a two-element array at an angle θ with respect to the array normal. The outputs of the two array elements, when normalised with respect to that of the first (on the left) element, are 1 and $u + jv$, respectively. From Chapter 2, we know the phase of the second element's output is $\frac{2\pi}{\lambda} d \sin\theta$. It follows that the angle-of-

arrival can be expressed as a function of the array outputs, i.e.

$$\theta = \arcsin\left(\frac{\lambda}{2\pi d}\arctan\frac{v}{u}\right) = F(u, v), \tag{5.132}$$

which is, as expected, a highly nonlinear function. In theory, $-\infty \leq u \leq \infty$ and $-\infty \leq v \leq \infty$. If $d \leq \lambda$, then θ is uniquely defined in the range $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$. That is, in this case, F is a continuous function. It would be desirable to extend this explicit derivation of F to a more complicated case. Unfortunately, this is, as we already mentioned, very difficult. Indeed, this is one of the major reasons why we use a neural network to approximate F.

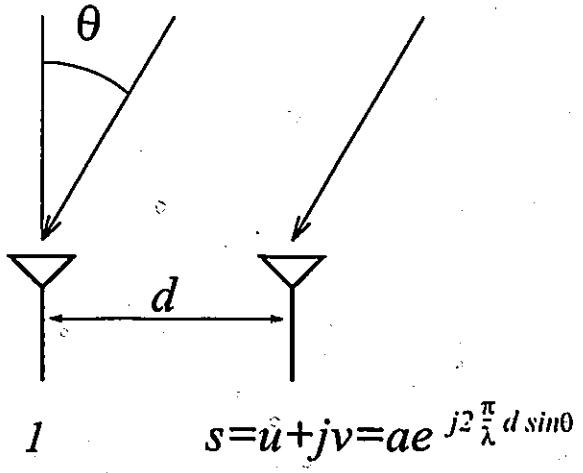

Figure 5.18: A plane wave impinges upon a two-element array.

2. One can also argue the existence of F from a philosophical point of view. Let us assume that F does not exist. Based on this assumption, there would be no technique or method that can provide the AOA solution. As we have seen, however, this is not the case. Using the MLE method, we are able to estimate the AOA, in spite of the computational complexity of the MLE method. In other words, F must exist for the MLE method to provide the AOA solution.

3. One can also argue the existence of F from a physical point of view. The currents at the elements in the receiving antenna array, which are induced by a

the radio source s_o at a particular location θ with respect to the array normal, can be expressed as a function, i.e.

$$\boldsymbol{s} = H(s_o, \theta). \tag{5.133}$$

The reciprocity theorem [42] states that if a voltage is applied to the terminals of an antenna A and the current is measured at the terminals of a separate antenna B, then an equal current will be obtained at the terminals of antenna A if the same voltage source is applied to the terminals of antenna B, provided that the transmission medium is linear, passive, and isotropic. In essence, if the same currents $\boldsymbol{s}$ are applied to the corresponding array elements, an equal current s_o will be induced at the source location, i.e.

$$(s_o, \theta) = F(\boldsymbol{s}) \tag{5.134}$$

where F is the inverse of H. That is, the inverse mapping is guaranteed by the reciprocity theorem.

5.2 Network Configuration

We have presented the principles of using the RBF network for AOA estimation in the previous chapter. In this section, the specific configuration and implementation of the network for AOA estimation in multipath will be described in detail. The RBF network for AOA estimation is shown in Fig. 5.19. Following Section 4.6, the input-output relationship has the following form,

$$\boldsymbol{\Theta} = \boldsymbol{\Phi W} \tag{5.135}$$

where

$$\boldsymbol{\Theta} = \begin{bmatrix} \theta^d(1) & \theta^i(1) \\ \vdots & \vdots \\ \theta^d(M) & \theta^i(M) \end{bmatrix}, \tag{5.136}$$

$$\boldsymbol{\Phi} = \begin{bmatrix} \phi(\|\boldsymbol{z}(1) - \boldsymbol{c}_1\|) & \cdots & \phi(\|\boldsymbol{z}(1) - \boldsymbol{c}_N\|) \\ \vdots & \ddots & \vdots \\ \phi(\|\boldsymbol{z}(M) - \boldsymbol{c}_1\|) & \cdots & \phi(\|\boldsymbol{z}(M) - \boldsymbol{c}_N\|) \end{bmatrix}, \tag{5.137}$$

and

$$\boldsymbol{W} = \begin{bmatrix} w_1^d & w_1^i \\ \vdots & \vdots \\ w_N^d & w_N^i \end{bmatrix}. \tag{5.138}$$

The network consists of three layers: the input layer, the hidden layer, and the output layer. The number of nodes in the input layer is equal to the dimension of the input vector $\boldsymbol{z}$. The hidden layer is composed of nonlinear units that are connected directly to all of the nodes in the input layer. This layer is usually of higher dimension than the input layer, i.e. the number of hidden units is larger than the number of source units. The output layer consists of two output units. It supplies the response (θ^d, θ^i) of the network to the input pattern $\boldsymbol{z}$ that is applied to the input layer.

5.2.1 Input Layer

The signals from the antenna elements can be fed directly to the network as the input patterns. However, we choose to use the second order statistics of the signals. That is, the covariance matrix $\boldsymbol{R}$ of the signal represented by (2.24) is first formed. There are a number of reasons to use the second order statistics. First, the initial phase ξ, which represents no particularly useful information to the AOA estimation problem, will be eliminated. This also reduces the dimensionality of the signal parameter space in which the network learns. Second, the matrix $\boldsymbol{R}$ is a Hermitian matrix, i.e. it is equal to its conjugate transpose or $R_{kl} = R_{lk}^*$. To the associative memory, the lower triangular matrix of $\boldsymbol{R}$ represents the redundant information of the upper one. Therefore, we may use either the upper or the lower triangular matrix of $\boldsymbol{R}$. Finally,

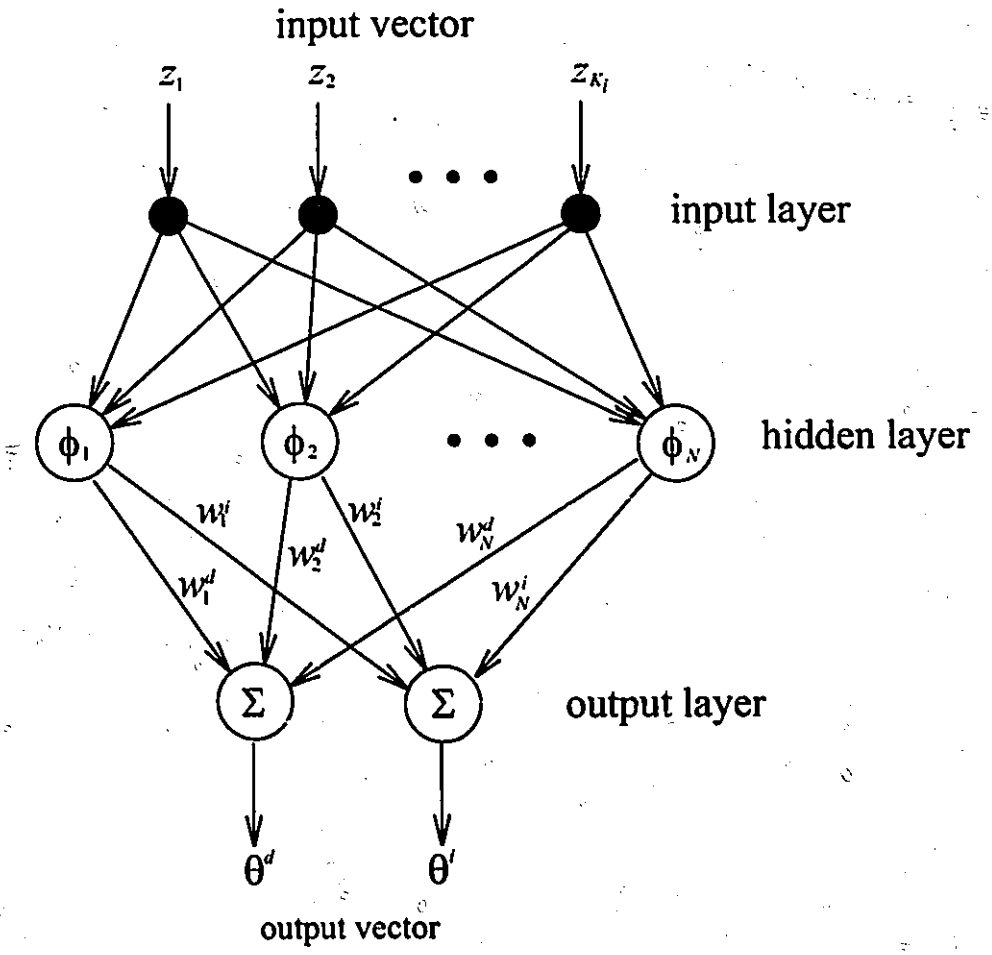

Figure 5.19: Structure of the RBF network used for AOA estimation in multipath environment.

as we have discussed in Chapter 2, the autocorrelation function of the array signal is approximately uniformly sampled when a minimum redundant array is used. When the minimum redundant array is used with its elements located at [0 d 4d 10d 16d 22d 24d 27d 29d], the upper triangular matrix of $\boldsymbol{R}$ consists of 29 distinct spatial correlation terms and seven redundant terms that result from the redundancy in the coarray (see Appendix). The redundant terms are discarded. Rearranging the $K_i = 29$ terms into a new $K_i \times 1$ vector $\boldsymbol{z}$, we form the input vector to the network,

$$\boldsymbol{z} = [z_1\ z_2 \cdots z_{K_i}]^T. \tag{5.139}$$

Therefore, the required number of input units of the network is K_i. The problem now becomes to find the inverse mapping $\boldsymbol{\Theta} = F(\boldsymbol{z})$. Although the covariance terms are used as the inputs to the network, the essence of the problem remains the same.

5.2.2 Hidden Layer

To construct the associative memory, N hidden units are used to evaluate the radial basis functions,

$$\phi(\|\boldsymbol{z} - \boldsymbol{c}_n\|) = e^{-\frac{\|\boldsymbol{z}-\boldsymbol{c}_n\|^2}{\sigma^2}}, \quad n = 1, \cdots, N \tag{5.140}$$

where

$$\|\boldsymbol{z} - \boldsymbol{c}_n\|^2 = [\boldsymbol{z} - \boldsymbol{c}_n]^H[\boldsymbol{z} - \boldsymbol{c}_n]. \tag{5.141}$$

There are a number of approaches for selecting the centres for the radial basis functions [62]:

1. The centres are chosen as a subset of the training data in a suitable manner. Once chosen, the centres are fixed during the rest of the learning process. This is considered to be a sensible approach, provided that the training data are distributed in a representative manner for the problem at hand.

2. The centres are selected in a self-organised manner, i.e., during the learning

process the centres are allowed to move to the regions of the input space where significant data are present.

3. The centres undergo a supervised learning process. In other words, the RBF network takes its most generalised form.

In this study, we use the first approach for the following reasons:

1. Training data are generated based on the model given by (2.24). Therefore, they can be generated so that they are distributed evenly over the hyper-surface associated with F. The N centres $\{\boldsymbol{c}_n,\ n = 1, ..., N\}$ are, in a representative manner, chosen from the training set.

2. Since the training data are generated to be evenly distributed in the input space, the self-organised approach may not be cost-effective in terms of the trade-off between the performance and computational complexity.

3. Nonlinear optimisation of the parameters of the network is beneficial when a minimal network configuration is required. However, it has been found that the same performance of generalisation may also be achieved by using a larger RBF network with fixed centres [63].

4. The fixed-centre approach is the simplest amongst the three in terms of implementation, i.e. it only requires the solution to a linear system.

The choice of the number of hidden units is another consideration. There are various methods for determining the number of hidden units in a RBF network. A few examples of these methods are the minimum description length (MDL) technique [64], the cross-validation method [65], and the canonical subspace analysis [66]. However, these methods were developed for generalisation or prediction where the network is required to generalise optimally for a given set of training examples in a noisy environment. If the number of hidden units is too small, the network cannot sufficiently

express the input-output relationship of the given examples. On the other hand, if the number of hidden units is excessive, the network may overfit the training data according to Occam's razor principle [67]. Our application here, however, requires different strategies. First, the training data are noise-free. Second, it is desirable to fit every point in the training data set to satisfy the strict interpolation condition. Ideally, the number of hidden units should be the same as the number of training data points. However, as mentioned above, this will cause computational difficulties if the number of training data is large. Therefore, in determining the number of hidden units, we take an *ad hoc* approach. That is, we choose the number of hidden units that gives the best training result. In next chapter, we will demonstrate the effects of the number of centres N through numerical examples.

5.2.3 Output Layer

In our application, we only need two output units since there are only two estimates to produce. The outputs of the two output units are respectively given by

$$\theta^d(z) = \sum_{n=1}^{N} w_n^d \phi\left([z - c_n]^H [z - c_n]\right) \tag{5.142}$$

and

$$\theta^i(z) = \sum_{n=1}^{N} w_n^i \phi\left([z - c_n]^H [z - c_n]\right). \tag{5.143}$$

The operation is a simple vector inner product between the outputs from the hidden units and the optimal synaptic weights.

5.3 Implementation

The implementation of the RBF associative memory for AOA estimation consists of two phases: learning and estimation. As shown in Fig. 5.20(a), the learning phase consists of acquiring the array output vectors, forming the input vectors by using the array vectors, and applying the least squares method to obtain a set of optimal

weights. The estimation phase consists of forming the input vector from sensor outputs and producing the AOA estimate by retrieving it from the memory, as shown in Fig. 5.20(b).

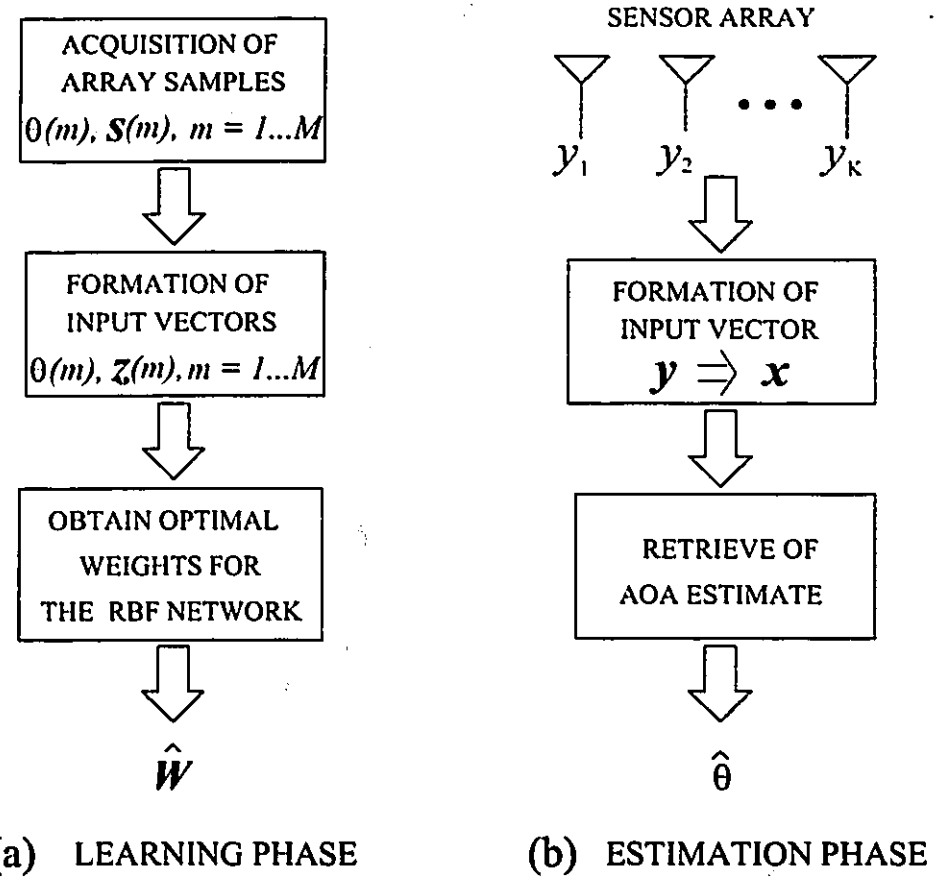

Figure 5.20: Implementation of RBF network for AOA estimation in a multipath environment.

5.3.1 Learning Phase

Learning samples must be generated for the RBF network to learn. One of the issues regarding the training data generation is how many samples are sufficient so that the inverse mapping can be approximated with a required accuracy. If we have the a priori knowledge about the properties of the inverse mapping F, we may apply the sampling

theory. However, this would be difficult since F is a highly nonlinear function and we do not know its exact closed form. Thus, the problem is equivalent to a class of ill-posed, inverse problems involving the reconstruction of a function from a sparse set of examples [68]. In this case, the option is to assume that F is a reasonably smooth function such that a reasonably large number M of learning samples are sufficient for the approximation [69]. Furthermore, it has also been shown that the Gaussian kernel functions can be used to solve this type of problem [27].

It is highly desireable to use real-life data to train the neural network because its strength is that it is able to provide itself, based on real-life data, a statistical model of the environment, in which it operates, to carry out the estimation. However, it may take a very long time or may be impractical to obtain sufficient quantity of data for training. Therefore, a hybrid approach is used here. That is, the neural network can be trained using synthetic data derived from the mathematical model to carry out the estimation. The training data are generated based on (2.24). Specifically, for a finite set $\{\boldsymbol{\theta}(m), m = 1, ..., M\}$, a corresponding array output vector set $\{\boldsymbol{s}(m), m = 1, ..., M\}$ is generated. Similar to the MLE method, the set $\{\boldsymbol{\theta}(m), m = 1, ..., M\}$ is formed by the combination of two smaller sets $\{\theta^d(m^d), m^d = 1, .., M^d\}$ and $\{\theta^i(m^i), m^i = 1, .., M^i\}$, i.e. $\boldsymbol{\theta}(m) = [\theta^d(m^d)\ \theta^i(m^i)]$. Both $\theta^d(m^d)$ and $\theta^i(m^i)$ are uniformly taken from the specific angular intervals V^d and V^i, respectively, which corresponds, in our case, to the low grazing angle region.

The input vectors $\{\boldsymbol{z}(m), m = 1, ..., M\}$ are formed as described in Section 5.2.1. The absolute power of the signal is usually unknown in practice. In order to avoid training the network to estimate the absolute power of the signal, we will normalise the input vector by its norm in both the training and estimation phases, i.e.,

$$\boldsymbol{z}(m) \rightarrow \frac{\boldsymbol{z}(m)}{\|\boldsymbol{z}(m)\|}, \quad m = 1, ...M. \tag{5.144}$$

This is followed by the formation of the data matrix $\boldsymbol{\Phi}$ in the form given by (5.137). The optimal weights are obtained using the recursive least squares method or the method of singular value decomposition. In Table 5.1, we summarise the learning

procedures.

Table 5.1: Procedure in the learning phase for AOA estimation using RBF network.

1.	Generate array output vectors $\{\boldsymbol{s}(m),\ m = 1,...M\}$ using (2.24).
2.	Evaluate the covariance matrices $\{\boldsymbol{R}(m), m = 1,...,M\}$.
3.	Form the input vectors $\{\boldsymbol{z}(m), m = 1,...M\}$ using (5.139).
4.	Normalise the input vectors $\{\boldsymbol{z}(m) \rightarrow \boldsymbol{z}(m)/\|\boldsymbol{z}(m)\|, m = 1,...M\}$.
5.	Form the data matrix $\boldsymbol{\Phi}$.
6.	Compute the least squares solution for $\hat{\boldsymbol{W}} = \boldsymbol{\Phi}^{+}\boldsymbol{\Theta}$.

5.3.2 Estimation Phase

The estimation phase consists of forming the input vector from sensor outputs and producing the AOA estimate by retrieving it from the memory, as shown in Fig. 5.20(b). If the array output vector is corrupted with noise, i.e. $\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{n}$, the estimate of $\boldsymbol{R}$ is evaluated as

$$\begin{aligned} \hat{\boldsymbol{R}} &= \frac{1}{N_s}\sum_{i=1}^{N_s} \boldsymbol{y}(n)\boldsymbol{y}^{\dagger}(n) \\ &= \frac{1}{N_s}\sum_{i=1}^{N_s} \boldsymbol{s}(n)\boldsymbol{s}^{\dagger}(n) + \boldsymbol{\epsilon} \end{aligned} \tag{5.145}$$

where N_s is the number of snapshots used in the estimation and $\boldsymbol{\epsilon}$ is a $K \times K$ matrix

$$\boldsymbol{\epsilon} = \frac{1}{N_s}\sum_{i=1}^{N_s}\left[\boldsymbol{s}(i)\boldsymbol{n}^{\dagger}(i) + \boldsymbol{n}(i)\boldsymbol{s}^{\dagger}(i) + \boldsymbol{n}(i)\boldsymbol{n}^{\dagger}(i)\right] \tag{5.146}$$

or

$$\epsilon_{kl} = \frac{1}{N_s}\sum_{i=1}^{N_s}\left[s_k(i)n_l^{*}(i) + n_k(i)s_l^{*}(i) + n_k(i)n_l*(i)\right]. \tag{5.147}$$

Because $\boldsymbol{s}$ and $\boldsymbol{n}$ are assumed uncorrelated with each other, the terms ϵ_{kl}, $k \neq l$, approximately follows the normal distribution with a mean value equal to zero, based on the large number theorem if N_s is sufficiently large. The expected value of $E[\epsilon] =$

$\sigma_n^2 \boldsymbol{I}$, where σ_n^2 is the noise power and $\boldsymbol{I}$ is an identity matrix. When the input vector is formed based on (5.139), the terms $\{\epsilon_{kl},\ k \neq l\}$ can be considered as additive noise terms, i.e.

$$\boldsymbol{x} = \boldsymbol{z} + \boldsymbol{\eta} \tag{5.148}$$

where $\boldsymbol{\eta} = [\epsilon_{12} \ldots \epsilon_{1K}\ \epsilon_{23} \ldots \epsilon_{2K} \ldots \epsilon_{K(K-1)/2}]^T$. It is easy to see that $\boldsymbol{x}$ approaches $\boldsymbol{z}$ as N_s goes to infinity. In other words, one may expect the estimation performance of the network to improve as N_s increases. In this sense, the RBF network is an asymptotic estimator.

Table 5.2: Procedure in the estimation phase for AOA estimation using RBF network.

1.	Evaluate the sample covariance matrix $\hat{\boldsymbol{R}}$ using the array outputs $\{\boldsymbol{y}(i), i = 1, \ldots N_s\}$.
2.	Form the input vector $\boldsymbol{x}$ as in (5.139).
3.	Normalise the input vector $\boldsymbol{x} \rightarrow \boldsymbol{x}/\|\boldsymbol{x}\|$.
4.	Compute the vector $\boldsymbol{\phi}$.
5.	Recall the corresponding AOA $\hat{\boldsymbol{\theta}} = \boldsymbol{\phi}\hat{\boldsymbol{W}}$.

5.4 Implementation Considerations

5.4.1 Computational Efficiency

Computational efficiency is one of the implementation considerations. As previously discussed, the MLE method requires M complex inner product operations if the synthetic array vectors are stored in memory. Each complex inner product operation in turn requires $8 \times K$ floating point operations (flops), where K is the number of array elements. Therefore, to produce an estimate $\hat{\boldsymbol{\theta}}$, the MLE method requires $8 \times K \times M$

flops, in addition to the computational requirement of M searching operations and the time required for retrieving the pre-calculated vectors from memory.

In contrast, the RBF network approach requires much less computations to produce an estimate $\hat{\boldsymbol{\theta}}$, if only the operations required to carry out the training are counted. To be more specific, the computations required to form an input vector are $12 \times K_i$ flops, where K_i is the number of input units. The number of flops required for evaluating the N exponential functions is $N(10 \times K_i + 3)$. The multiplication of weights with the outputs of the hidden units and the summation are counted as $2N$ flops. For L output units, the total number of flops required by the RBF network is

$$12K_i + L[N(10 \times K_i + 3) + 2N] = 12K_i + LN(10 \times K_i + 5). \tag{5.149}$$

It should be noted that this number is independent of M, the number of training data points used in the training. For example, if $L = 2$, $N = 120$, and $K_i = 29$, the total number of flops required by the network is 71,148. In Fig. 5.21 is shown the comparison of the numbers of flops required by the RBF network and the MLE method. When $M \simeq 1000$, the number of flops required by the MLE method is of the same order as that required by the RBF network. However, as M increases, the number of flops required by the MLE method increases at a rate proportional to $8 \times K$, whilst the number of flops required by the RBF network remains constant.

5.4.2 Real-Time Implementation

In the approach when an RBF network is used to carry out the estimation, the associative recall process directly provides the AOA estimates. There is no need to search for peaks or maxima in the values of some objective function. As a result, the computational efficiency increases. Furthermore, the parallel implementation of the RBF network makes it feasible to carry out real-time processing. Although a large learning set is required in the learning phase, the learning process itself can be carried out off-line. Thus, it should not diminish the possibility of real-time implementation.

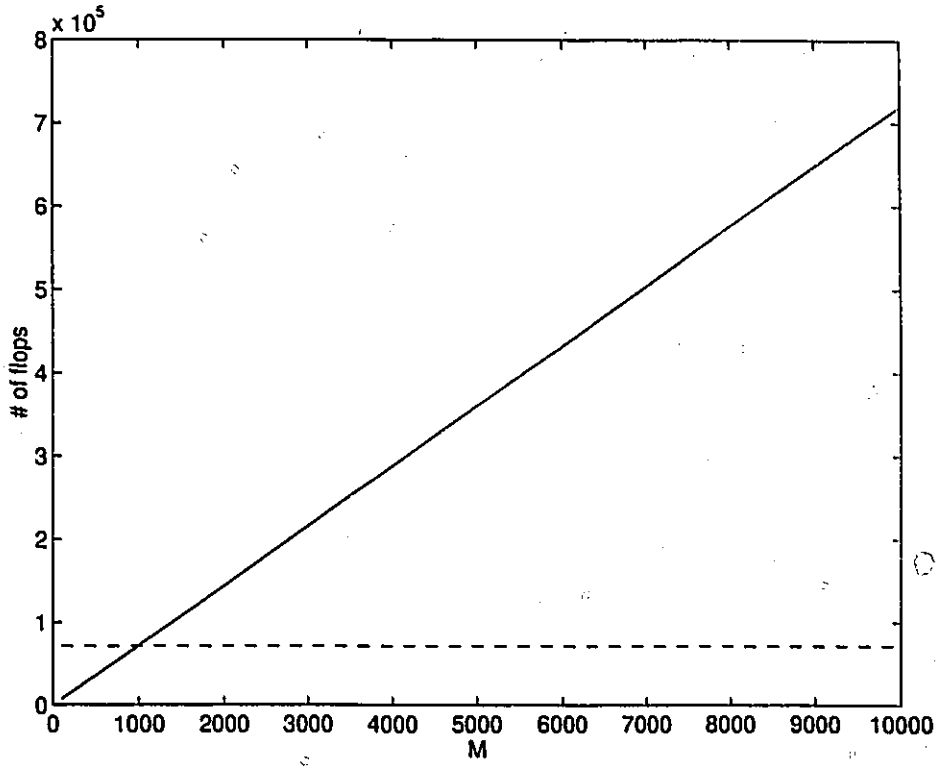

Figure 5.21: Comparison of the numbers of flops required by the RBF network (- -) and the MLE method (—).

The progress in VLSI (Very Large Scale Integration) has lowered implementation costs for array processors, which provide very high throughput rates and CAD (Computer Aided Design) techniques have facilitated speedy prototyping and implementation for customised processors. In recent years, a great deal of attention has been drawn to ASIC (application specific integrated circuit) technology. Emerging as the mainstream of VLSI design, ASIC technology is revolutionising the design, manufacturing, and marketing practice in electronics-related industries. The essence of this revolution is that in pre-ASIC era system designers dealt with microprocessors and standard integrated components, a design process of building systems from chips; whereas ASIC design is inherently a process of integrating systems into chips. The VLSI-ASIC technology permits the implementation of many neural-computing algorithms. A number of analogue VLSI neural processors [70, 71, 72, 73] and digital VLSI neural processors [74, 75, 76] have been reported in the literature. The analogue neural processors provide a computationally efficient capability for mimicking neurobiology. The digital neural processors, on the other hand, provides computational flexibility which permits the implementation of complex algorithms in many applications, in addition to the advantages of being easy to design and manufacture.

For the specific application of the RBF network in the AOA problem discussed in this thesis, we believe that the digital neural processor is more suitable than its analogue counterpart because of the flexibility offered by the digital neural processor. In terms of computation, the RBF network is of relatively low complexity. However, the AOA application may require a certain degree of flexibility. For example, the dimension of the input vector may vary from system to system, depending on the number of antenna elements. The network may have to be trained and re-trained as often as necessary, because of the changing environment.

Chapter 6

SIMULATIONS AND REAL DATA TESTS

In this chapter, computer simulations will be carried out to study the performance of both the MLE and RBF techniques developed in the thesis. Analysis will be conducted to determine the learning performance of the RBF network. Furthermore, Monte-Carlo simulations are performed to evaluate the estimation performance of the MLE method and the RBF network. In addition to simulation study, real data will be used to test and validate the two techniques.

6.1 Simulations

6.1.1 Simulation Parameters

In the simulation, the performance is measured by the normalized overall mean squared error (MSE), i.e.

$$\mathcal{E} = \frac{1}{N_t}\sum_{n=1}^{N_t}\left(\frac{\theta_n^d - \hat{\theta}_n^d}{BW}\right)^2 + \frac{1}{N_t}\sum_{n=1}^{N_t}\left(\frac{\theta_n^i - \hat{\theta}_n^i}{BW}\right)^2 \tag{6.150}$$

where N_t denotes the number of ensembles to be averaged and BW denotes the beamwidth of the array.

The simulations are based on a 9-element minimum redundant linear array with its elements located at [7.89 7.95 8.12 8.46 8.80 9.15 9.26 9.43 9.55] metres. The corresponding antenna aperture is 1.66 metres. Since we assume that the frequency of the signal is 10.2 GHz, the beamwidth of the array is 0.9°. In Table-6.3 are listed the values of the parameters used in the simulations. The values for M, N, and h are normally kept fixed, except in those few cases where they are allowed to vary. At $h = 15\ m$, the true AOA's of the direct and indirect signals are 0.066° and −0.306°, respectively, and their angular separation is 40% BW. The searching intervals V^d and V^i are chosen to correspond to a low grazing angle region.

6.1.2 Learning Performance of the RBF network

To determine the learning performance of the RBF network, the number of learning samples and the number of hidden units in the network are taken into consideration for their effect on the learning process. In the analysis, noiseless cases ($SNR = \infty$) are assumed. The learning performance of the network is measured by the learning MSE, which is defined as the MSE produced by the learned network when noiseless input vectors are applied.

First, the effects of the number of hidden units on the learning performance of the network are examined. As was shown in Chapter 4, by taking $N < M$ the weakened interpolation condition will result in a learning error. Since a perfect recall is unachievable for a rectangular key matrix, the output of the network will be different from the true $\boldsymbol{\theta}$. The magnitude of the error should depend on the value of N. To be more precise, the error depends on the rank of the key matrix $\boldsymbol{\Phi}$, as stated in Proposition 4.1. In Fig. 6.23 is plotted the singular value spectrum, which exhibits the down trend as the number of hidden unit, N, increases. Therefore, it is expected that the learning MSE, evaluated based on (6.150), should also decrease as N increases.

Table 6.3: Simulation parameters.

frequency f	$10.2GHz$
wavelength λ	$0.02941m$
array aperture	$1.66\ m$
array beamwidth	$0.9°$
source range	$4610\ m$
refractivity gradient $\frac{dN}{dh}$	-39
dielectric constant of the sea water ϵ_c	$44.9 - 40.96j$
significant wave height σ_H	0.2
radius of earth R	$6375000\ m$
source height h	15
M^d for both RBF and MLE	75
V^d for both RBF and MLE	$-0.25°$ to $0.35°$
M^i for both RBF and MLE	75
V^i for both RBF and MLE	$-1.25°$ to $0.25°$
number of learning samples M	5625
number of hidden units N	120
σ in the hidden units	1
number of ensembles N_t	100

In Fig. 6.22 is shown the learning MSE as a function of number of hidden units for noiseless input vectors. One may observe a trend that the MSE decreases as N increases. However, this trend does not continue beyond $N = 120$. This is because the probability that the matrix $\boldsymbol{\Phi}$ becomes ill-conditioned is high for a large value of N. This is indeed the case if one examines the singular value distribution shown in Fig. 6.23. It can be seen that there is a threshold at about $N = 130$, beyond which the singular values are insignificant in the sense that they are almost equal and relatively small. In theory, the singular vectors associated with these 'insignificant'

singular values are defined in the null space. However, in a practical computation implementation, the values assigned to these vectors become arbitrary. Therefore, acting as noise, they affect the evaluation of the pseudo-inverse of the key matrix. As a result, relatively larger learning errors occur if $N > 130$, as shown in Fig. 6.22.

Next, the effects of the number of learning samples M are examined on the learning performance of the RBF network. As have been pointed out in Section 5.3.1, M is required to be reasonably large so that the RBF network is able to accurately approximate the inverse mapping. The question, however, is how large should M be? Analytically, it would be formidable task to determine an optimal value for M unless F is known. Fortunately, computer simulations can be used as a tool to determine how the value of M may affect the performance of the RBF network. Based on simulations, we may be able to determine the value of M in accordance with the desired performance in a particular application. In Fig. 6.24 is given the learning MSE as a function of M. In this case, $N = 120$ and $h = 15\ m$. Qualitatively, it can be seen that as M increases, the learning MSE decreases. That is, a large value of M results in a small sampling spacing which in turn leads to a small error. Furthermore, the figure reveals the nonlinear relationship between the value of M and the MSE. That is, the rate of decrease in MSE becomes faster and faster as M increases. In contrast, we find that in the case of the MLE method, if the resolution at each search-dimension doubles, the MSE will decrease by half (3 dB). That is, if the same type of curve is plotted for the MLE method, it will appear to be a straight line.

6.1.3 Estimation Performance

In the above analysis, the focus is on the learning capability of the network without noise components in the signal. In the following performance analysis, the main objective is to investigate the effects of noise on the two methods we have developed. To determine the estimation performance of the RBF network, Monte-Carlo simulations are carried out. That is, input vectors with different levels of noise are repeatedly

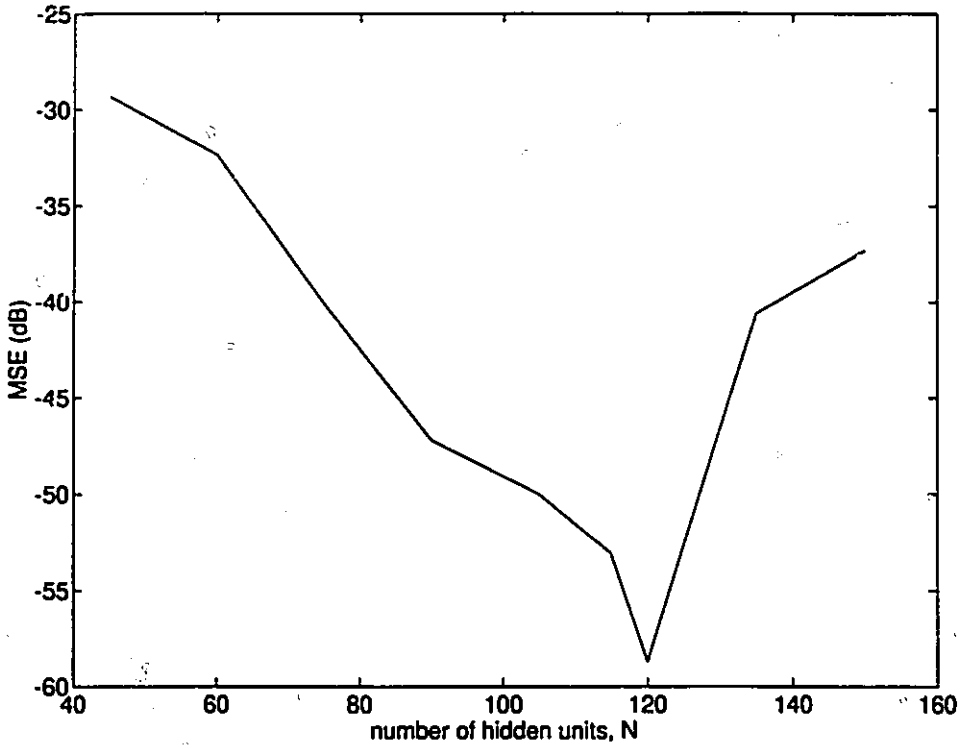

Figure 6.22: Learning MSE as a function of the number of hidden units N.

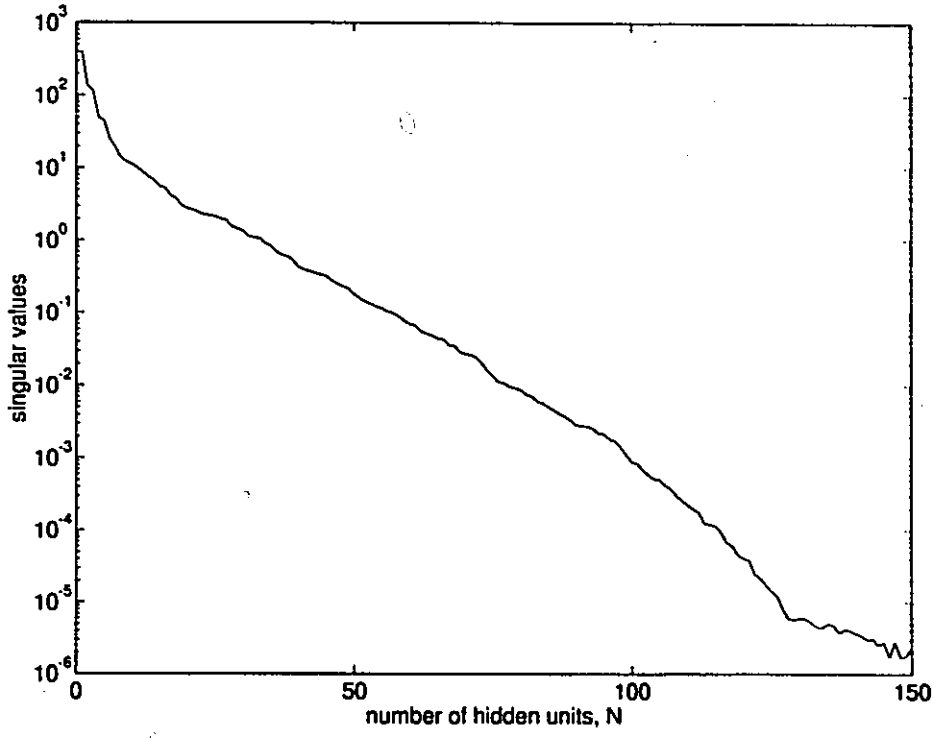

Figure 6.23: Singular value spectrum showing the distribution of the singular values as a function of the number of hidden units N.

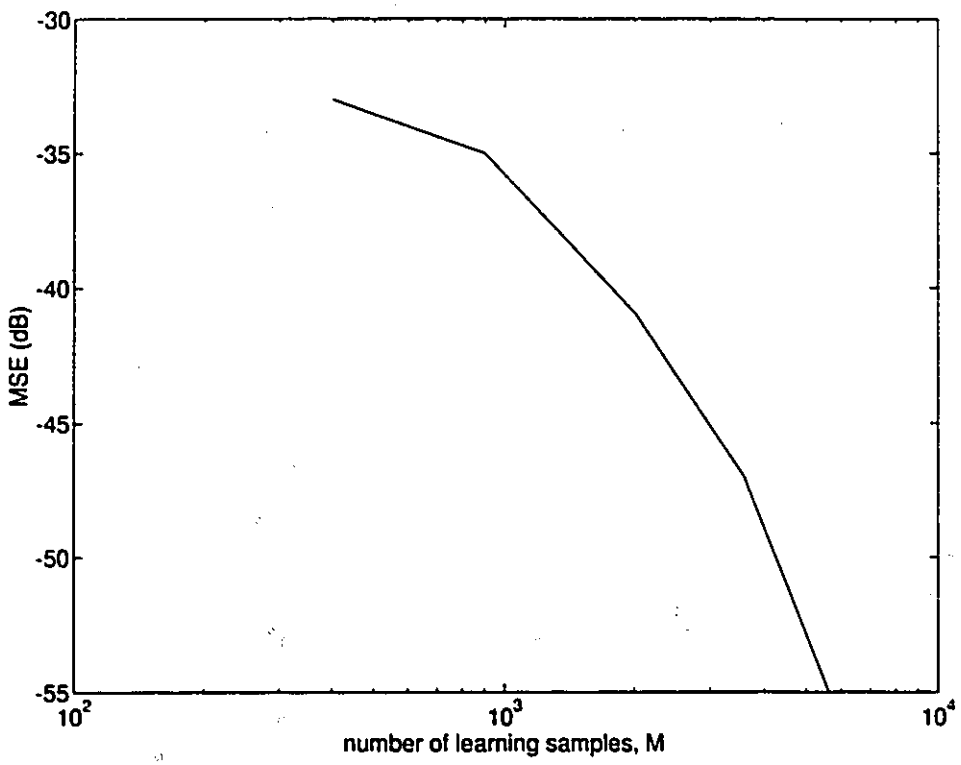

Figure 6.24: Learning MSE as a function of the number of learning samples M.

applied to the trained network and the values of MSE produced by the network are then determined.

In Fig. 6.25 are shown a family of MSE curves, for different values of N, as a function of SNR. At low SNR ($< 15\ dB$), the performance of the RBF network is insensitive to height, as well as to the number of centres, N. That is, at low SNR, the errors due to noise is such a dominant factor that the performance difference due to different values of N is not observable. Only when the noise power is reduced, is one able to observe the effect of different values of N, namely the fact that the larger the N, the smaller the value of MSE. This is completely consistent with what Proposition 4.1 predicts. In other words, as the value of N increases, λ_{N+1} becomes smaller, resulting in a smaller MSE.

Similar MSE curves are plotted in Fig. 6.26, except that they are given for different numbers of training samples M. At low SNR ($< 15\ dB$), the estimation performance of the RBF network is statistically the same for different values of M. That is, at low SNR, the performance of the network is affected more by noise than by the number of training samples. Only when the SNR is increased to a certain level, is there a need to consider the effect of M on the estimation performance.

In the following analysis, the performance of both the RBF network and the MLE method will be evaluated. In order to compare the performance with the Cramer-Rao (CR) lower bound derived specifically for the multipath case as described in Chapter 3, we set the value of M relatively large (75×75) for the MLE method so that the errors due to the finite value of M is insignificant. In Fig. 6.27, the performance MSE curves for both the MLE method and the RBF network are plotted as a function of number of snapshots N_s. In this case, the SNR is set to 20 dB. As is expected, the MSE decreases as N_s increases for both the RBF network and the MLE method. That is, estimates obtained using either technique exhibit asymptotic behaviour. It can also be observed that the MSE produced by the RBF network is larger by a few dB than that by the MLE method. In other words, by using the MLE method one is

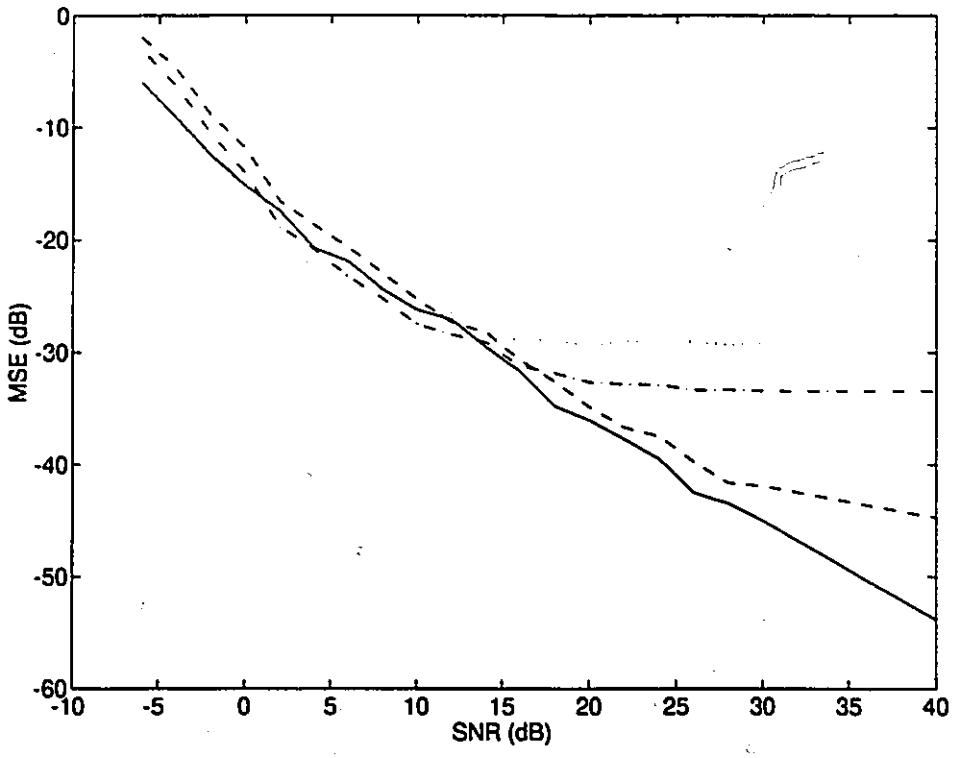

Figure 6.25: Estimation MSE as a function of SNR for $h = 15m$; $\cdots$ $N = 45$, $-\cdot-$ $N = 60$, - - $N = 90$, — $N = 120$.

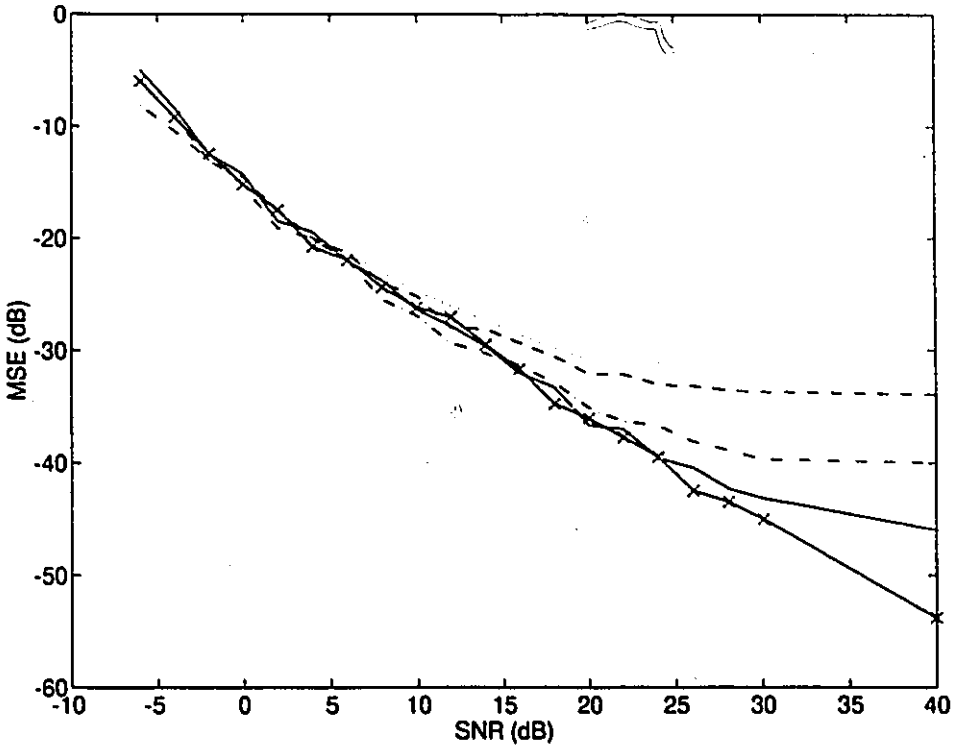

Figure 6.26: Estimation MSE as a function of SNR for $h = 15m$; $\cdots$ $M = 20^2$, - - $M = 30^2$, $-\cdot-$ $M = 45^2$, — $M = 60^2$, -x- $M = 75^2$.

able to obtain the estimation performance consistently closer to the CR bound than using the RBF network.

Similar observations can also be made from a comparison of the estimation performance and the CR bound for different levels of SNR, as shown in Fig. 6.28. The performance is obtained with $N_s = 32$. The results show that the performance of the MLE method is a few dB closer to the CR bound than that of the RBF network. Therefore, based on these simulation results, the MLE method can be said to be a more efficient estimator than the RBF network, provided that the value of M is sufficiently large. In the results, some error is due to the finite value of M for both the RBF network and MLE method. Although these errors are relatively small, they are observable when the SNR is high. This is manifested by the fact that the MSE curves for both the RBF network and MLE method deviate from the CR bound when SNR>30 dB.

The above performance comparison is carried out without considering the computational complexity. On the surface, it seems that there was no advantage to using the RBF network. It is only when we include the computational complexity in the comparison that the true advantage of using the RBF network can be revealed. In Fig. 6.29 is shown the estimation MSE for the MLE method as a function of number of flops. In the same figure, the asterisk represents the performance of RBF network. In the case shown in the figure, SNR$=20$ dB. Along the vertical dotted line, which indicates the same computational complexity, it can be observed that the MSE produced by the RBF network is at least 6 dB less than that produced by the MLE method. On the other hand, along the horizontal dotted line, which indicates the same MSE, one can see that for both the approaches to have the same performance, the computational complexity required by MLE method must be of the order of 3.2×10^5 flops. That is, the computational requirement for the MLE method is $\frac{3.2\times10^5}{71,148} = 4.5$ times as that for the RBF network. The advantage of the RBF network over the MLE method is attributed to its excellent ability to interpolate and approximate.

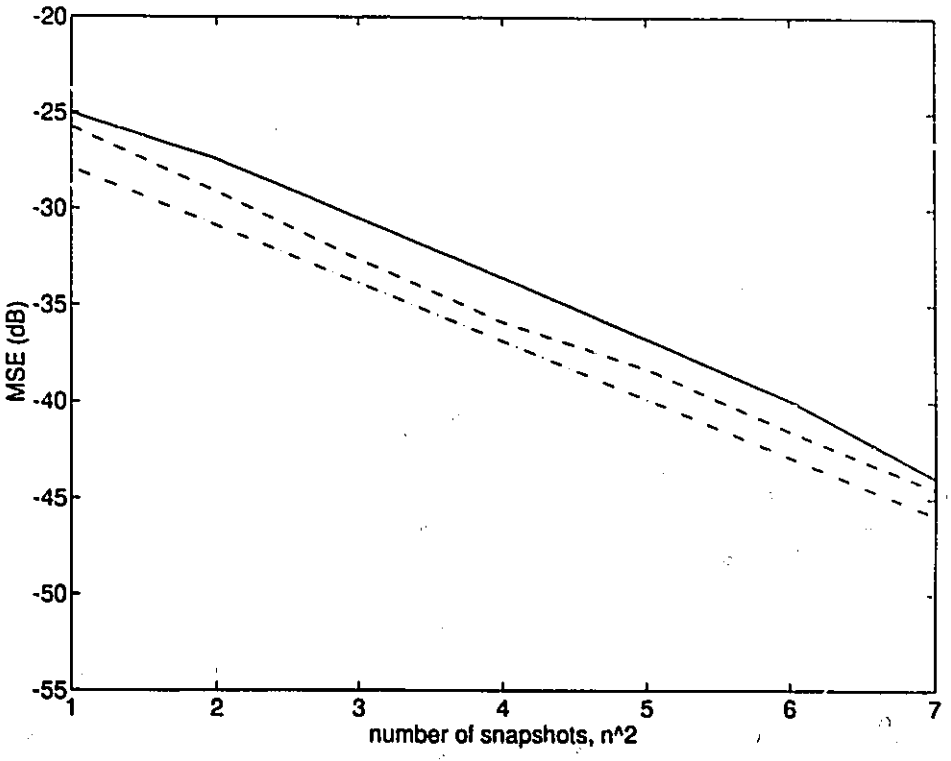

Figure 6.27: Estimation MSE as a function of number of snapshots N_s; — · — CR bound, - - MLE, — RBF.

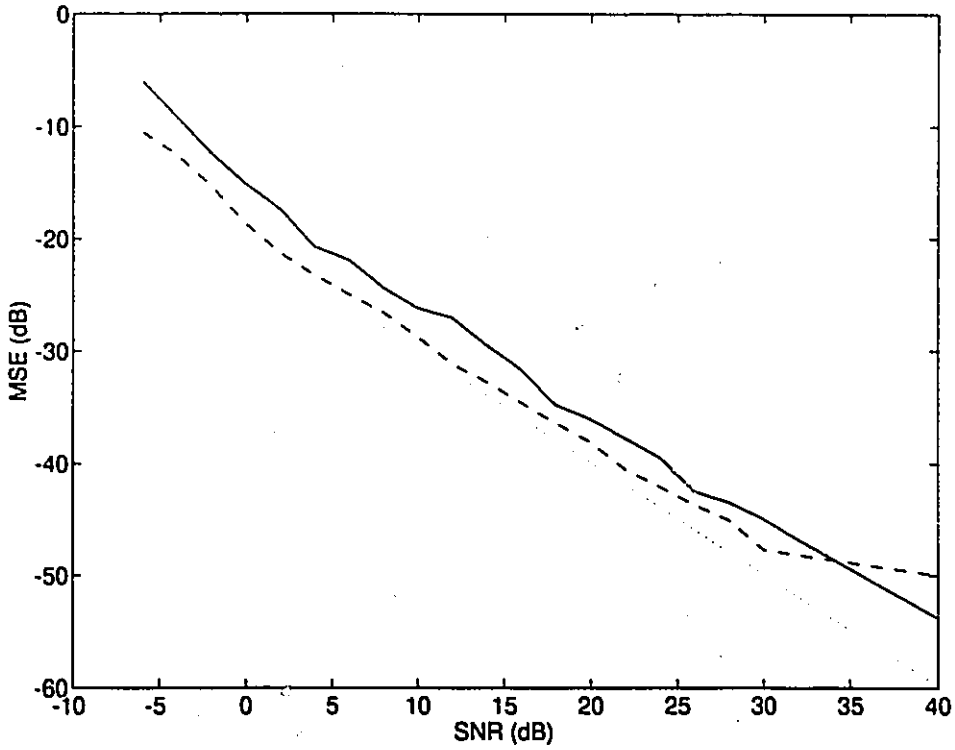

Figure 6.28: Estimation MSE as a function of SNR; ··· CR bound, - - MLE, — RBF.

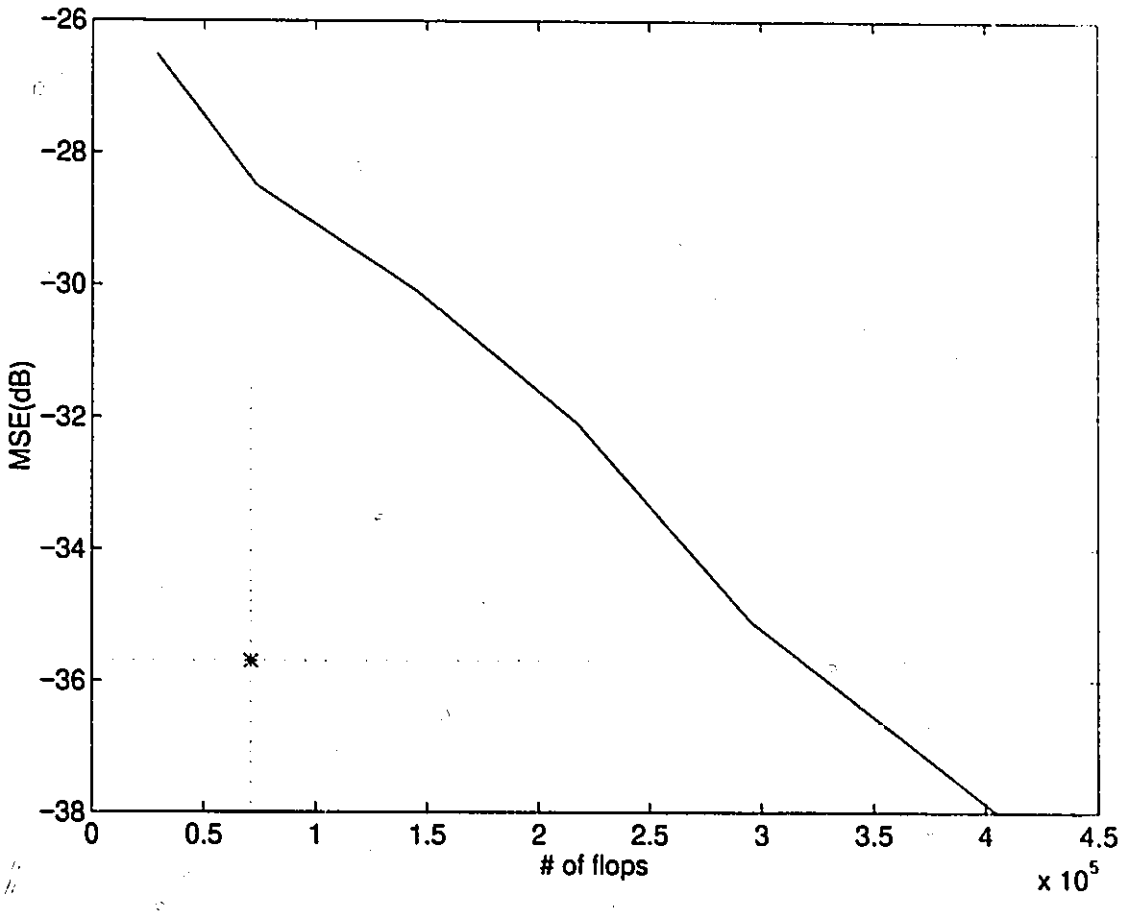

Figure 6.29: Estimation MSE as a function of number of flops; — MLE, * RBF.

6.2 Real Data Tests

6.2.1 Real Database

In late October 1987, a series of experiments were performed on the west coast of the Bruce Peninsula, Ontario, Canada, overlooking Lake Huron. This particular location was chosen because of the high sea states caused by the combination of the westerly winds, the shallow water offshore, and the long fetch across Lake Huron. One of the objectives of these experiments was to acquire data suitable for evaluating the performance of high-resolution estimation techniques. A 32-element sampled aperture antenna, developed at the Communications Research Laboratory, McMaster University, Hamilton, Ontario, Canada, was used to collect multipath data under a variety of meteorological and water surface conditions. The experimental system consisted of a transmitter, a sampled aperture antenna, and a data acquisition system. The transmitter was a 100-mW CW beacon, sited at a distance of 4610 metres from the receiving antenna. Fig. 6.30 gives a vertical-plane view of the geometry for the measurement set-up. The receiving antenna array consists of 32 10-dB H-polarised horns. All elements in the array were precisely aligned and uniformly distributed to within 0.1 mm to minimise errors due to misalignment and uneven spacing between elements. The data were recorded with 12-bit precision at a sampling rate of 62.5 *samples/second*. The system operated at X-band. A more detailed description of the measurement system, the experiment and the calibration is given in [77]. To emulate the minimum redundant array, only the outputs from Elements 1, 2, 5, 11, 17, 23, 25, 28, and 30 are used. The emulated array has a 1.66-metre aperture. The height of the transmitting horn could be varied through the interval from 3.5 to 18.5 metres above the water level. It follows from the geometry that the angular separation of the direct and reflected signals varied from 9% to 50% of a BW. The height of the first element in the array is at 7.8 metres above the water level. Each data set consists of 127 snapshots which were taken within a period of 2 seconds.

6.2.2 Test Results

As a demonstration of the quality of the data collected with the sampled aperture antenna, comparisons are made between synthetic and measured data. Fig. 6.31 gives a comparison of a typical distribution of signal amplitude over the array face. In Fig. 6.32 is given the corresponding phase distribution. The synthetic data are derived using (2.24). In the case of the amplitude distribution, one observes an interference pattern due to the superposition of the direct and indirect signals at the individual antenna elements. In Fig. 6.33 is shown the comparison of DFT beamforming results for both the synthetic data and measured data. It should be noted that the highest sidelobe in the beampatterns is above $-7\ dB$ because the array is a minimum redundant array. From Fig. 6.33, it is clear that the beampatterns are similar to each other. The mainlobes of the two patterns are almost the same. However, the sidelobe level corresponding to the measured data is higher than that for the synthetic data. This discrepancy is a manifestation of array errors, which are due to [77] : (1) imperfect array calibration, (2) misalignment of array elements, (3) quantization errors, (4) diffuse component in the reflected signal, and (5) other random errors. Furthermore, this beamforming result clearly shows that the DFT technique failed to resolve the direct and indirect signals which are separated by less than a BW.

A number of sets of data are used to test both the MLE and the RBF network. The RF frequency of the signal used for these data sets was 10.2 GHz. The water surface was characterised as having a significant wave height (SWH) between 1.5 and 2.0 metres or SWH deviation between 0.375 and 0.5 metres. The roughness of the surface is well within the Rayleigh roughness criterion (3.50) which is 0.75 metres for the maximum grazing angle realised in the experimental set-up.

In each of the cases to be presented, eight snapshots ($N_s = 8$) are used to obtain one AOA estimate for both the MLE and the RBF network. Thus, the scaling used for the time axis in the figures to be shown, one unit is equivalent to 0.128s. The results

are shown in Figs. 6.34 through 6.39. In Table 6.4 is given the comparison of the mean values of AOA estimates shown in each of these figures. whereas presented in Table 6.5 is a comparison of the angular separations of the direct and indirect signals based on the experimental set-up geometry ($\Delta\theta_g$), and the mean angular separations based on the estimation ($\Delta\theta_e$). It should be noted that we do not compare the actual AOA values with those based on the experimental set-up geometry because it would require the knowledge of the tilt angle of the array antenna with respect to the vertical when the data were recorded. Unfortunately, the tilt measurement was not available. Therefore, there is no appropriate way to compare the estimates with the actual AOA.

Table 6.4: Comparison of mean AOA.

case #	source height (m)	θ^d_{MLE} (deg.)	θ^d_{RBF} (deg.)	θ^i_{MLE} (deg.)	θ^i_{RBF} (deg.)
1	15.53	-0.090	-0.11	-0.63	-0.65
2	15.53	-0.054	-0.084	-0.36	-0.48
3	13.06	-0.10	-0.093	-0.40	-0.38
4	9.53	-0.18	-0.20	-0.49	-0.41
5	4.31	-0.16	-0.19	-0.46	-0.41
6	3.53	-0.16	-0.20	-0.45	-0.41

From the test results in different test cases, the estimates derived using both the RBF network and the MLE method agree with each other, to a large extent, not only in terms of mean AOA values and angular separations, but also in terms of instant AOA values, as shown in Figs. 6.34 through 6.39. Since the data were recorded in a real environment which was continuously changing with time, it is not surprising that the AOA estimates exhibit some variation with time. Small variations can be seen within each figure and a large difference can be observed when we compare the

Table 6.5: Comparison of angular separations.

case #	source height (m)	$\Delta\theta_g$ (BW)	$\Delta\theta_e$ (BW).RBF	$\Delta\theta_e$ (BW).MLE
1	15.53	0.42	0.60	0.59
2	15.53	0.42	0.43	0.34
3	13.06	0.35	0.34	0.33
4	9.53	0.26	0.23	0.35
5	4.31	0.12	0.25	0.32
6	3.53	0.10	0.22	0.32

estimates in Cases 1 and 2 where the source height was the same but the data were recorded at different time periods. Analysis and investigation of the characteristics of the forward-scattered signals have already been carried out [78, 79, 80, 81]. The results of those analyses, which may shed some light on the variation of the AOA estimates, will not be presented here since it is outside of the scope of this thesis. Another observation is that the estimated angular separations are consistent with the change of heights of the radio source. That is, for smaller values of source height, the estimated angular separations $\Delta\theta_e$ are smaller. However, there are some discrepancies between $\Delta\theta_g$ and $\Delta\theta_e$, especially for small values of source heights. These discrepancies may be due to (1) imperfect calibration of the array, (2) system noise, or/and (3) inaccurate measurement of the experimental set-up geometry. The effects of these factors appear to be stronger at lower source height.

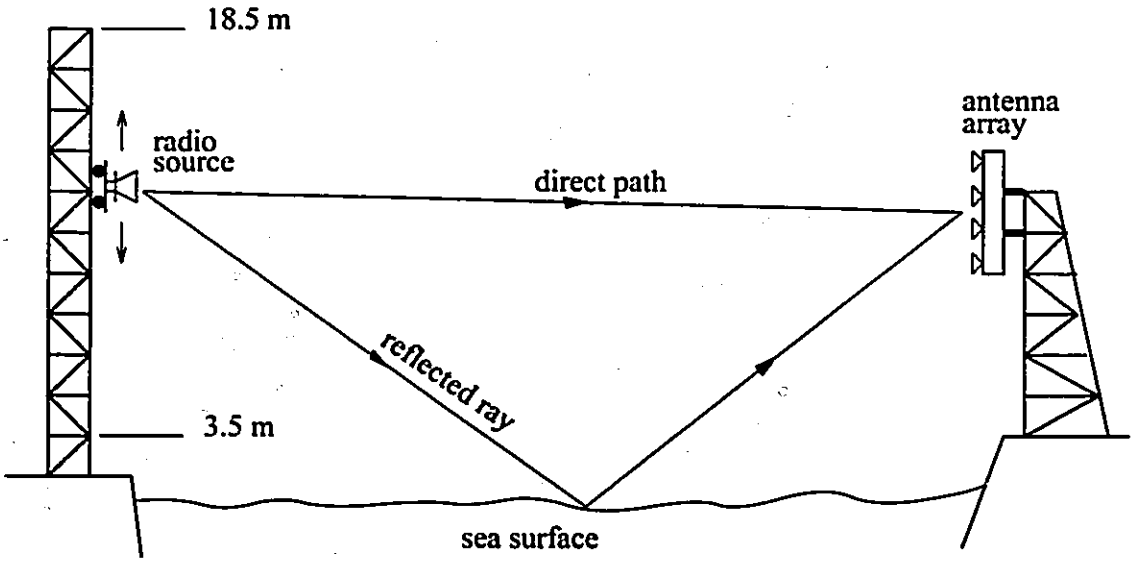

Figure 6.30: Vertical-plane view of the geometry for the measurement set-up.

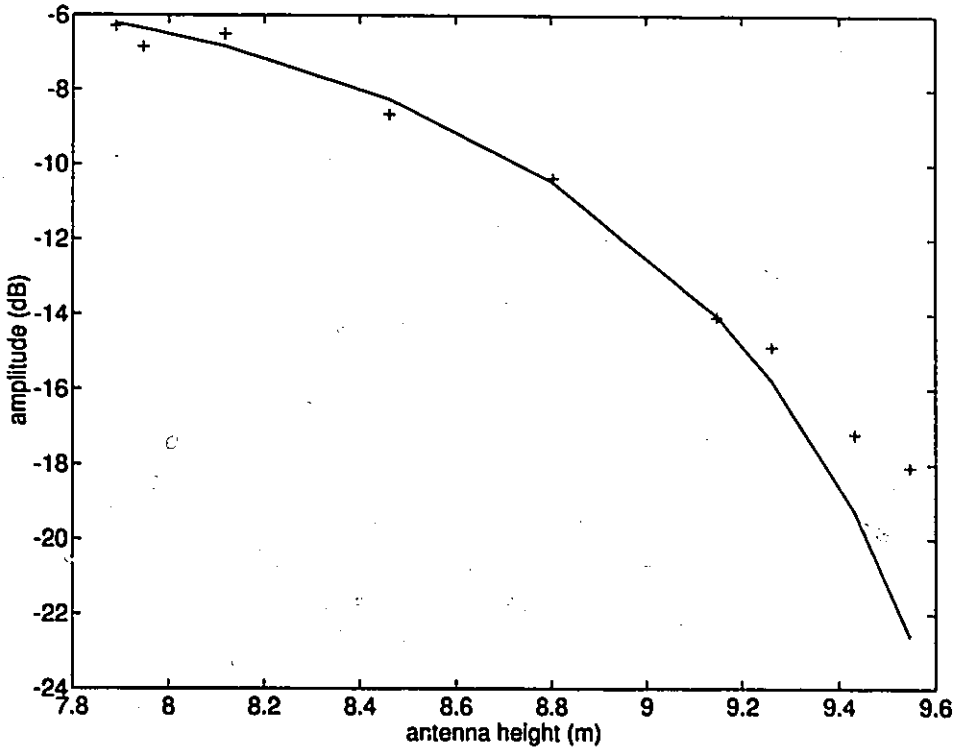

Figure 6.31: Comparison of synthetic (—) and measured (+) signal amplitude distribution over the array face.

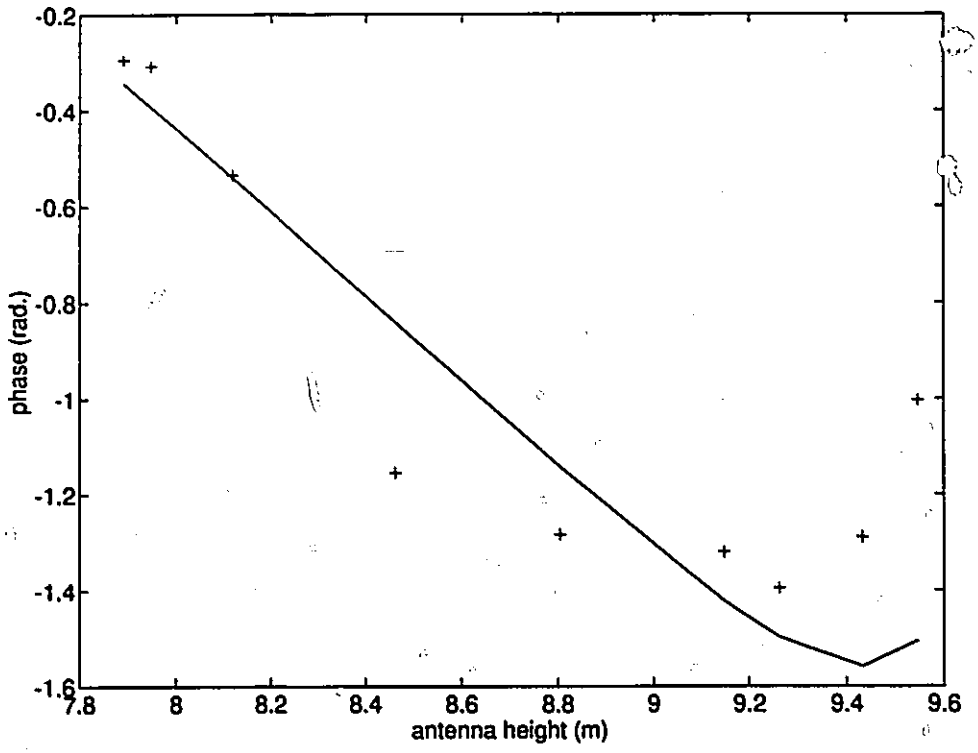

Figure 6.32: Comparison of synthetic (—) and measured (+) signal phase distribution over the array face.

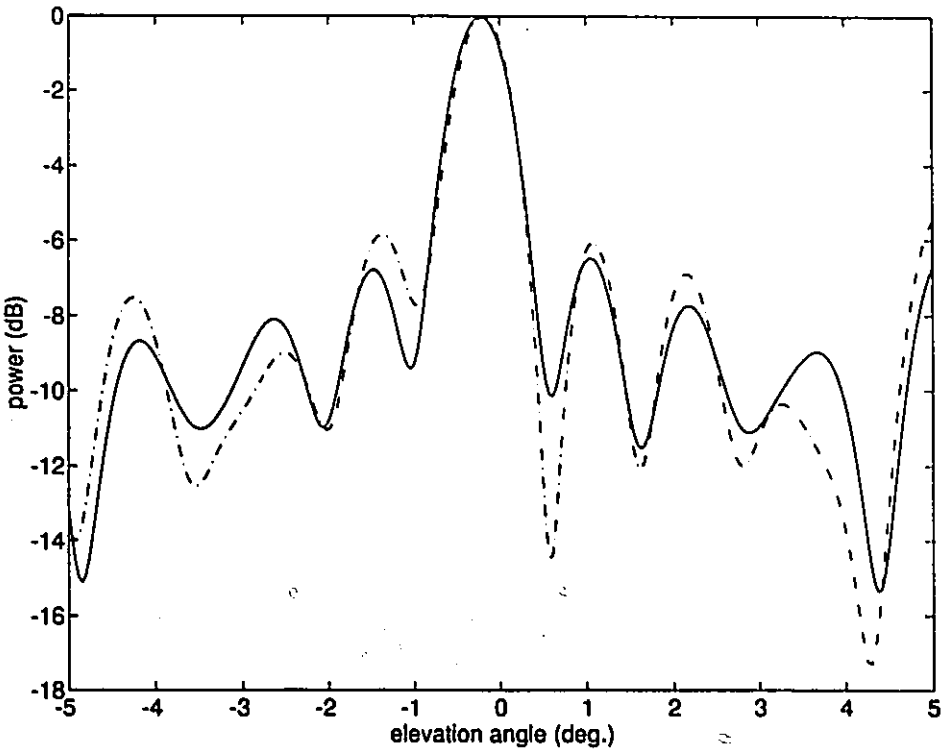

Figure 6.33: Comparison of DFT beamforming results obtained by using synthetic (—) and measured data (– · –) shown in the previous two figures.

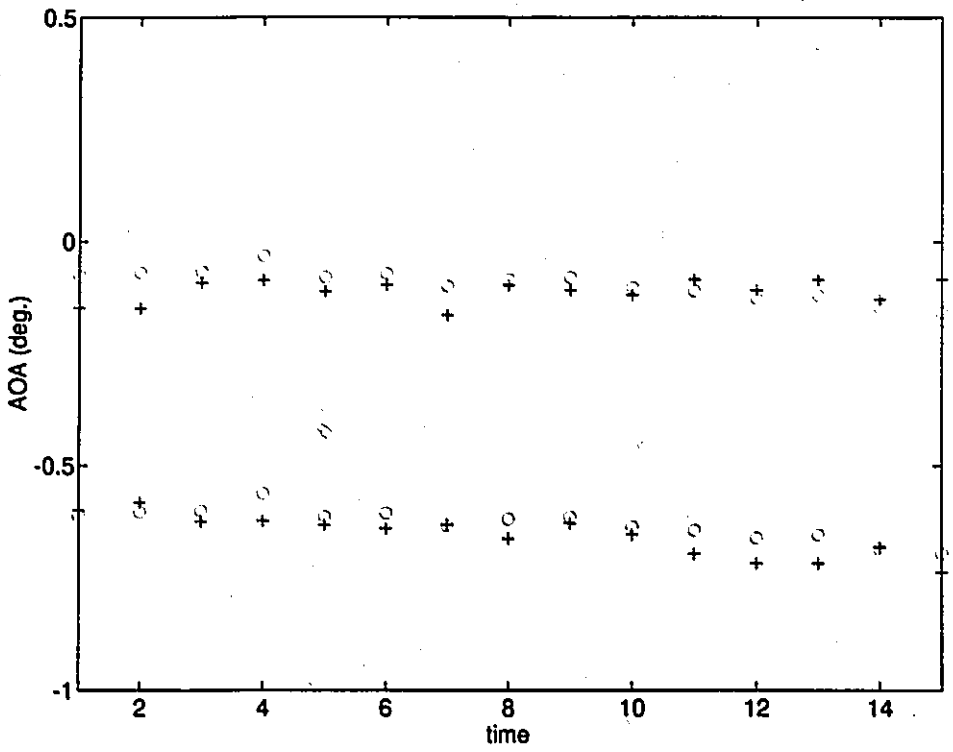

Figure 6.34: AOA estimates derived for Case 1. The results (+) obtained using the RBF network are compared with those using the MLE method (o). Source height = 15.53 m.

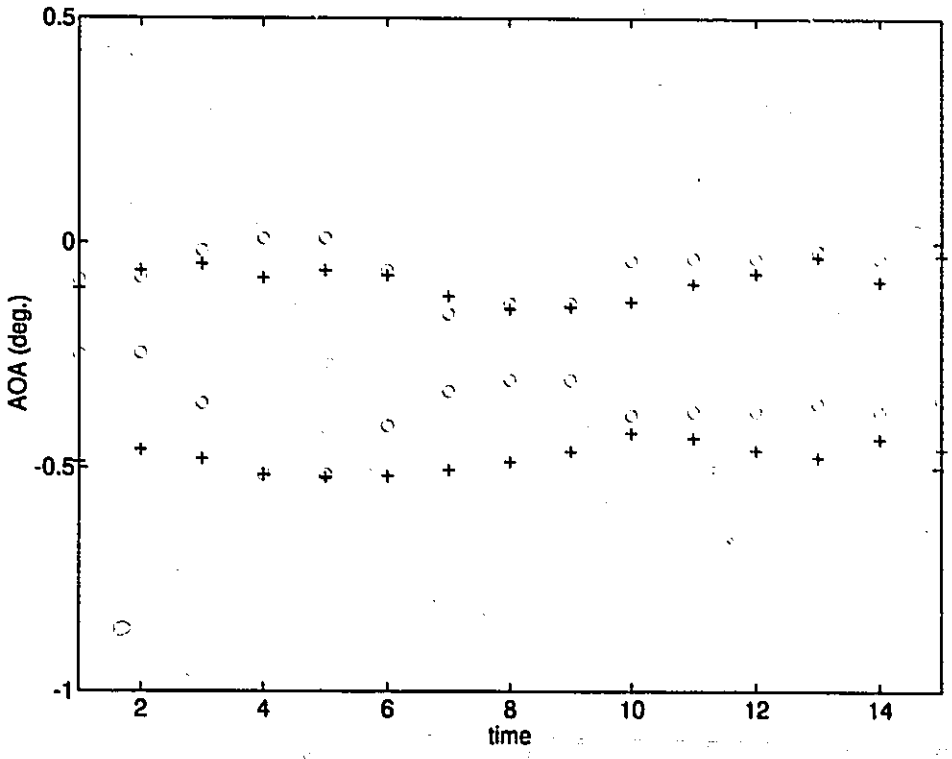

Figure 6.35: AOA estimates for Case 2. The results (+) obtained using the RBF network are compared with those using the MLE method (o). Source height = 15.53 m.

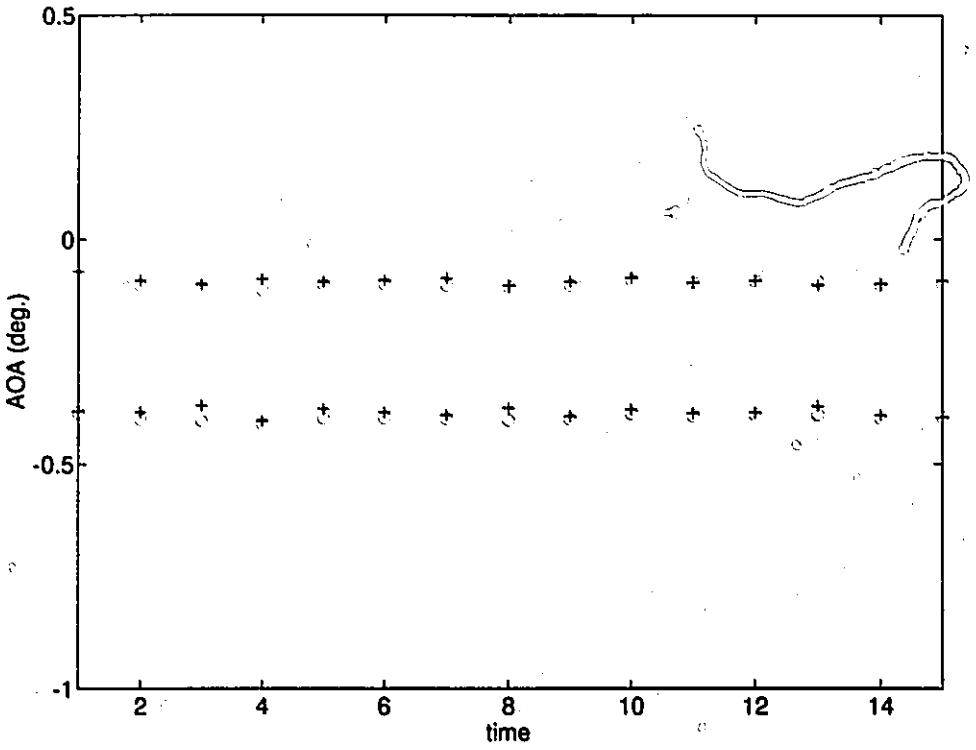

Figure 6.36: AOA estimates for Case 3. The results (+) obtained using the RBF network are compared with those using the MLE method (o). Source height = 13.06 m.

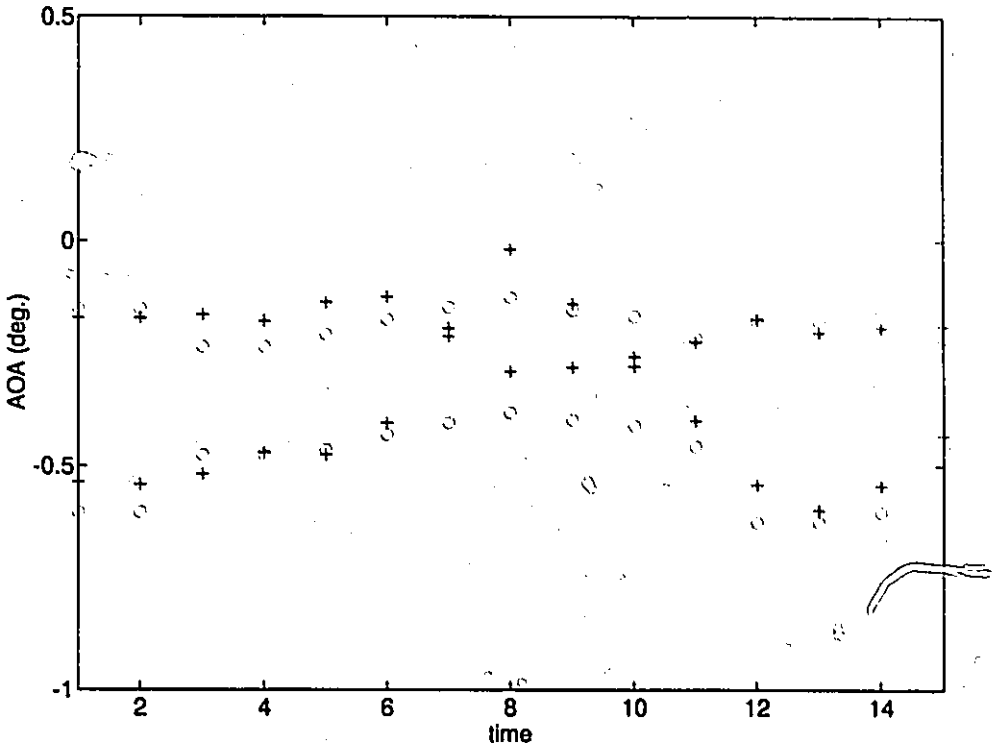

Figure 6.37: AOA estimates for Case 4. The results (+) obtained using the RBF network are compared with those using the MLE method (o). Source height = 9.53 m.

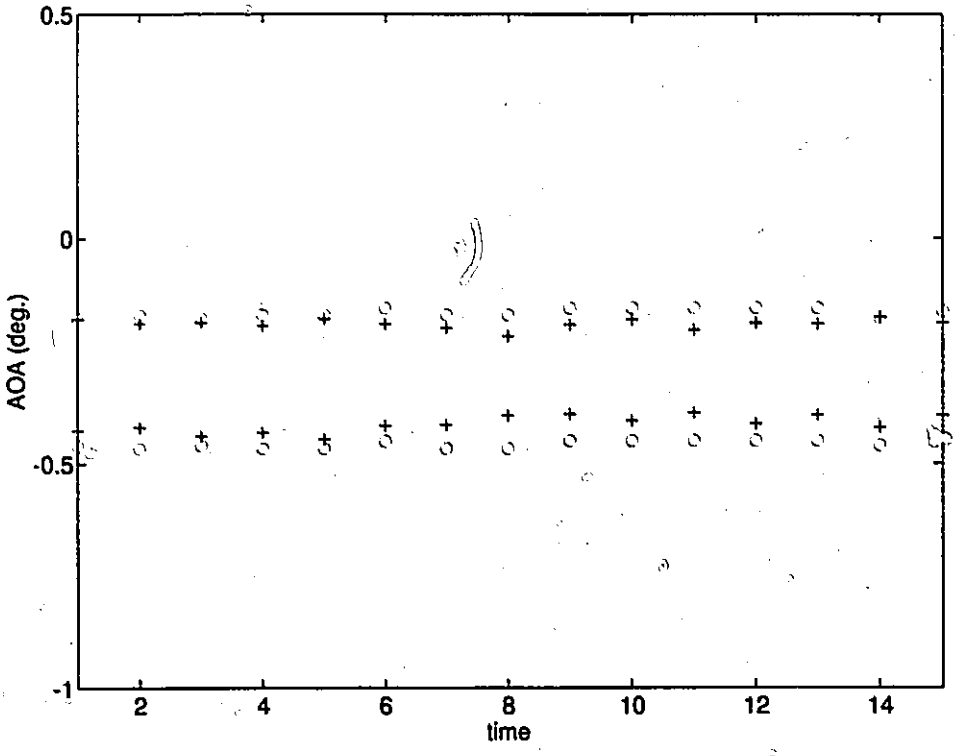

Figure 6.38: AOA estimates for Case 5. The results (+) obtained using the RBF network are compared with those using the MLE method (o). Source height = 4.31 m.

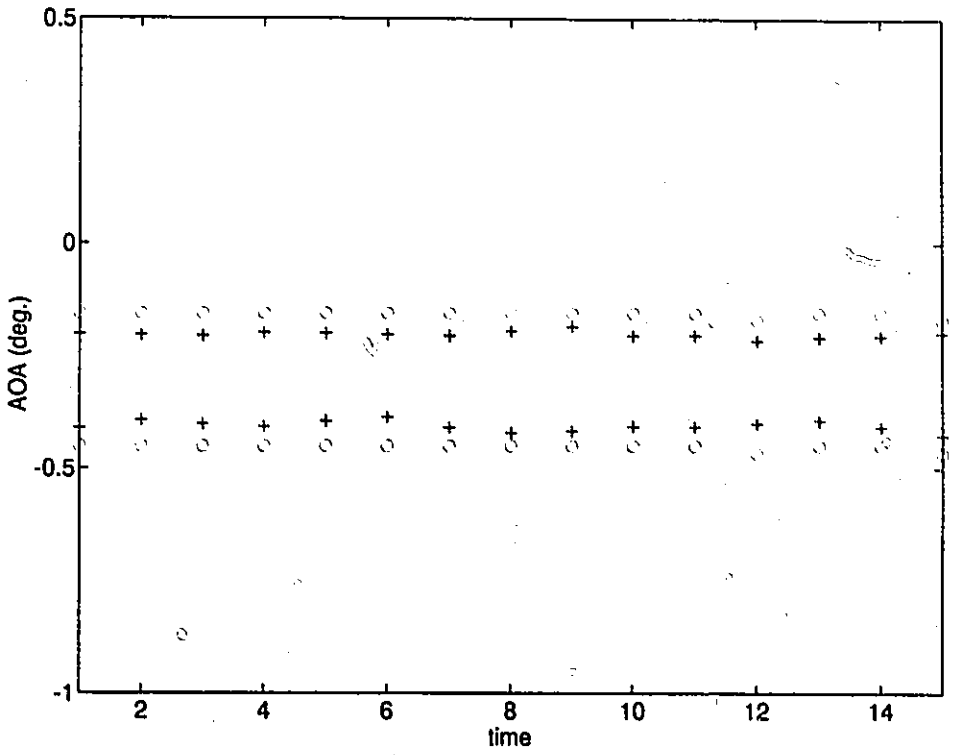

Figure 6.39: AOA estimates for Case 6. The results (+) obtained using the RBF network are compared with those using the MLE method (o). Source height = 3.53 m.

Chapter 7

CONCLUSIONS AND DISCUSSIONS

In this thesis, we have developed two new methods for the estimation of angles-of-arrival in multipath environments. The first technique is based on combining the maximum likelihood estimation (MLE) technique with a deterministic multipath model. The algorithm is unique in that a highly deterministic multipath signal model is used when formulating the likelihood function, which is then maximised with respect to the angles-of-arrival. The deterministic multipath signal model that has been developed to describe the physics underlying the propagation of signals from a signal radio source to a receiver is much more complete than the general angle-of-arrival model commonly used in other maximum likelihood formulations. This model makes use of the geometrical information and a priori knowledge of a number of physical parameters. These parameters include (1) the refractivity gradient, (2) the reflection coefficient, (3) the specular and diffuse scattering coefficient, and (4) the divergence factor. By using the deterministic multipath signal model with the MLE estimator, one is essentially making more information available for the estimation process, with the net result that the estimator's performance can be greatly enhanced. As well, the Cramer-Rao bounds that applies specifically to this model have been derived to

provide a performance measure for the mean-squared errors (MSE) in the estimated angles-of-arrival.

It is an undisputed fact that the MLE method has an optimum performance in the statistical sense and is insensitive to the coherence of the incoming signals. However, it is common knowledge that the computational load of the nonlinear optimisation procedure inherently required by the MLE method is too heavy for real-time processing. Accordingly, an alternative technique has been developed to deal with this shortcoming. In light of using a bank of parallel matched filters to implement the MLE method, we proposed a novel approach to the AOA estimation problem, which is based on the use of an associative memory. The motivation behind the parallelism of the matched-filter implementation provides some insight into the approach of using an associative memory. The connection between the matched-filter implementation and neural networks is that the matched-filter structure resembles the associative memory. The functionality of an associative memory is identical to that of the inverse mapping network. This provides a more comprehensive explanation for the rationale of exploiting the inverse mapping concept in the AOA estimation problem. In particular, the AOA problem is considered as a mapping from the space of AOA to the space of the sensor output. A nonlinear associative memory is used to form the inverse mapping from the space of sensor output to the space of AOA. This memory is realised using the radial basis function (RBF) neural network. Analysis has been carried out to show the network's capability of interpolation and approximation. We have shown that in principle, a perfect recall is possible using the RBF neural network. However, in avoiding ill-conditioning of the key matrix in the case where a large number of training samples are required, we adopted the generalised RBF associative memory. Accordingly, we have carried out the analysis using the generalised RBF network to implement the nonlinear associative memory. It was found that the norm of the AOA estimation error produced by the generalised RBF network is bounded by the quantity that is proportional to the $(N+1)^{th}$ singular value of the key matrix.

In the actual implementation of the RBF network for AOA estimation, the second order statistics of the signals are used as the input vector of the network. This can eliminate the initial phase of the signal and are suitable for the application of the minimum redundant array. Finally, we have shown that the RBF network is much more efficient in terms of computation than the MLE algorithm. This makes the RBF network attractive for real-time implementation.

Simulations have been carried out to understand the efficiency of the RBF neural network approach. In analysing the learning and estimation performance of the network for AOA estimation, we have considered the effects of the number of learning samples and the number of hidden units. The results support the predictions made in the development of the principles of the RBF network for AOA estimation. That is, the learning and estimation MSE is inversely proportional to the number of learning samples and the number of hidden units. It has also been observed that at relatively low SNR, further increasing the either the number of learning samples or the number of hidden units would not improve the estimation performance of the RBF network. Furthermore, we have investigated the estimation performance of both the MLE technique and the RBF network as functions of the number of snapshots and SNR and have compared the performance with the CR bound that we have derived. It has been observed that the performance of the MLE algorithm is consistent with the Cramer-Rao bounds. It has also been observed that the MLE method was more efficient in terms of estimation than the RBF network, if the searching resolution used in the MLE method is sufficiently high (i.e. a large value of M). When the computational complexity is considered, the RBF network has been shown to have much better performance than the MLE method. In particular, with the same computational complexity, the MSE produced by the RBF network is much less than that produced by the MLE method. On the other hand, for the same performance, the computational complexity required by MLE method is much higher than that required by the RBF network. In applications where real-time implementation is re-

quired, the RBF network is much more attractive. On the other hand, if accuracy is the only concern, the MLE method should be used. Finally, we have validated both the MLE technique and the RBF network using real data, which were collected using a 32-element sampled aperture antenna. The results obtained using the RBF network are very similar to those obtained using the MLE method.

In general, when a relatively new concept or technique is conceived and introduced, its development may be incomplete, just as a new-born baby needs time and nursing to grow up. The RBF approach to the AOA problem is not an exception. Although we have demonstrated the viability of this approach in theory and with validations using both simulated and real data, many relevant issues remain to be addressed. For example, the high degree of nonlinearity of the RBF network is not only a blessing but also a curse. On one hand, the high degree of nonlinearity makes the RBF network an excellent functional approximator. On the other, the high degree of nonlinearity also makes it difficult to analytically determine the characteristics of the network. In many cases, we had to resort to computer simulations. However, the general applicability of the characteristics evaluated through simulations remains to be determined. Technically, the performance of the RBF network can be improved by optimising other parameters in the network, such as σ_m in (4.97). As we have mentioned in Chapter 5, this will require a large-dimensional nonlinear optimisation process. Therefore, the trade-off between the computational intensity and performance needs to be investigated. Finally, we have shown that in terms of functionality, the RBF network is able to implement the correlation operation. Therefore, it is possible to use the RBF network in other types of applications that require the use of a correlation function, such as for correlation receivers that are used in a spread spectrum communications system.

Appendix A

Spatial Correlation Function

The covariance matrix is formed by taking the expected value of the outer product of the array output vector $\boldsymbol{s}$, i.e.

$$\boldsymbol{R} = E\left[\boldsymbol{s}^{*}\boldsymbol{s}^{T}\right]. \tag{A.151}$$

For example, in the four-element optimum non-redundant array shown in Fig. 2.6, the element positions are [0, 1, 4, 6]. Therefore, the array output vector is given as

$$\boldsymbol{s} = [s_1\ s_2\ s_5\ s_7]^T \tag{A.152}$$

Forming the covariance matrix using the array output vector, we have

$$\boldsymbol{R} = \begin{bmatrix} R(1,1) & R(1,2) & R(1,5) & R(1,7) \\ R(2,1) & R(2,2) & R(2,5) & R(2,7) \\ R(5,1) & R(5,2) & R(5,5) & R(5,7) \\ R(7,1) & R(2,1) & R(7,5) & R(7,7) \end{bmatrix}. \tag{A.153}$$

where $R(k,l) = E[s_k^* s_l]$ is the spatial correlation function with the relative spatial lag $l-k$. Rewriting $R(k,l)$ as R_{l-k}, we express the covariance matrix as

$$\boldsymbol{R} = \begin{bmatrix} R_0 & R_1 & R_4 & R_6 \\ R_{-1} & R_0 & R_3 & R_5 \\ R_{-4} & R_{-3} & R_0 & R_2 \\ R_{-6} & R_{-5} & R_{-2} & R_0 \end{bmatrix} \tag{A.154}$$

The upper triangular matrix of $\boldsymbol{R}$ consists of the spatial correlation function $\{R_k,\ k = 1, 2, \cdots, 6\}$.

Similarly, for the nine-element minimum redundant array with the elements at [0, 1, 4, 10, 16, 22, 24, 27, 29], the upper triangular matrix of the covariance matrix is given as

$$\boldsymbol{\Delta}_u = \begin{bmatrix} R_1 & R_4 & R_{10} & R_{16} & R_{22} & R_{24} & R_{27} & R_{29} \\ & R_3 & R_9 & R_{15} & R_{21} & R_{23} & R_{26} & R_{28} \\ & & R_6 & R_{12} & R_{18} & R_{20} & \boxed{R_{23}} & R_{25} \\ & & & \boxed{R_6} & \boxed{R_{12}} & R_{14} & R_{17} & R_{19} \\ & & & & \boxed{R_6} & R_8 & R_{11} & R_{13} \\ & & & & & R_2 & R_5 & R_7 \\ & & & & & & \boxed{R_3} & \boxed{R_6} \\ & & & & & & & \boxed{R_2} \end{bmatrix} \tag{A.155}$$

where terms in a box are the redundant terms. When these redundant terms are discarded, the spatial correlation function is formed as $\{R_k,\ k = 1, 2, \cdots, 29\}$.

Bibliography

[1] J. E. Evans et al., "MLS multipath studies", Technical Report FAA-RD-79-21, Lincoln Laboratory, MIT, April 1979.

[2] J. E. Evans, J. R. Johnson, and D. F. Sun, "Application of avanced signal processing techniques to angle of arrival estimation in atc navigation and surveillance systems", Technical Report FAA-RD-82-42, Lincoln Laboratory, MIT, June 1982.

[3] D. K. Barton, "Low-angle radar tracking", *Proc. IEEE*, vol. 62, pp. 687–704, June 1974.

[4] R. Schmidt, "Multiple emitter location and signal parameter estimation", *in Proc. of the RADC Spectrum Estimation Workshop, RADC-TR-79-63*, Rome Air Development Centre,Rome, N.Y., 1979.

[5] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood", *Proc. IEEE*, vol. 70, pp. 975–989, Sept 1982.

[6] W. F. Gabriel, "A high-resolution target-tracking concept using spectral estimation techniques", Technical Report 8797, NRL, May 1984.

[7] W. F. Gabriel, "Spectral analysis and adaptive array superresolution techniques", *Proc. IEEE*, vol. 68, pp. 654–666, June 1980.

[8] J. E. Evans, J. R. Johnson, and D. F. Sun, "High resolution angular spectrum estimation techniques for terrain scattering analysis and angle of arrival estimation", *in Proceeding of the First ASSP Workshop on Spectral Estimation*, pp. 5.3.1–5.3.10., McMaster University, Hamilton, Canada, Aug. 1981.

[9] W. D. White, "Low-angle radar tracking in the presence of multipath", *IEEE Trans. Aerosp. Electron. Syst*, vol. AES-10, pp. 835–852, Nov. 1974.

[10] S. Haykin and J. P. Reilly, "Maximum-likelihood receiver for low-angle tracking radar, part 1: The symmetric case", *IEE Proc.-F*, vol. 129, pp. 261–272, Aug. 1982.

[11] A. J. Weiss, A. S. Wilsky, and B. C. Levy, "Maximum likelihood array processing for the estimation of superimposed signals", *Proc. IEEE*, vol. 76, pp. 203–205, Feb. 1988.

[12] E. Bosse, R. Turner, M. Lecours, "Tracking swirling fluctuating target at low altitude over the sea", *IEEE Trans. Aerosp. Electron. Syst.*, vol. 27, pp. 806–822, Sept. 1991.

[13] T. Lo and J. Litva, "Use of a highly deterministic multipath signal model in low-angle tracking", *IEE Proc.-F*, vol. 138, pp. 163–171, April 1991.

[14] T. Lo and J. Litva, "Low-angle tracking using multifrequency sampled aperture radar", *IEEE Trans. Aerosp. Electron. Syst.*, vol. 27, pp. 797–805, Sept. 1991.

[15] K. Sharman, "Maximum likelihood parameter estimation by simulated annealing", *in Proc. Int. Conf. Acoust. Speech Signal Processing*, 1988.

[16] S. Jha and T. S. Durrani, "Direction of arrival estimation using artificial neural networks", *IEEE Trans. Systems, Man and Cybernetics*, vol. 21, pp. 1192–1201, Sept./Oct. 1991.

[17] L. Fa-Long and B. Zheng, "Real-time neural computation of the maximum likelihood criterion for bearing estimation problems", *Neural Networks*, vol. 5, pp. 765-769, 1992.

[18] G. Martinelli and R. Perfetti, "Neural network approach to spectral estimation of harmonic processes", *IEE Proceedings, Part G*, vol. 140, pp. 95-100, 1993.

[19] T. Lo, H. Leung and J. Litva, "Radial basis function neural network for direction-of-arrivals estimation", *IEEE Signal Processing Letters*, vol. 1, pp. 1-3, Feb. 1994.

[20] T. Lo, H. Leung and J. Litva, "Low-angle radar tracking in a naval environment using for-ward-backward nonlinear prediction method", *IEEE J. Oceanic Eng.*, vol. accepted, 1994.

[21] R. P. Lippman, "An introduction to neural net", *IEEE ASSP Magazine*, vol. 4, pp. 4-22, April 1987.

[22] A. S. Lapedes and R. Farber, "Nonlinear signal processing using neural networks: Prediction and system modeling", Technical Report Technical Report LA-UR-87, Los Alamos National Laboratory, Los Alamos, CA, 1987.

[23] J. Hertz, A. Krogh and R. G. Palmer, *Introduction to The Theory of Neural Computation*, Addison-Wesley, Redwood City, 1991.

[24] Y. Pao, *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Redwood City, 1989.

[25] J. J. Hopfield and D. W. Tank, "Neural computation of decisions of in optimization", *Biological Cybernetics*, vol. 52, pp. 141-152, 1985.

[26] D. Goryn and M. Keveh, "Neural networks for narrowband and wide-band direction finding", *in Proc. Int. Conf. Acoust. Speech Signal Processing*, 1988.

[27] T.Poggio and F. Girosi, "Networks for approximation and learning", *Proc. IEEE*, vol. 78, pp. 1481–1496, Sept 1990.

[28] J. Park and I. W. Sandberg, "Universal approximation using radial basis function networks", *Neural Computation*, vol. 3, pp. 246–257, 1991.

[29] F. Girosi and T. Poggio, "Networks and the best approximation property", *Biological Cybernetics*, vol. 63, 1990.

[30] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks", *Complex System*, vol. 2, pp. 321–355, 1988.

[31] J. E. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units", *Neural Computation*, vol. 1, pp. 281–294, 1989.

[32] S. Renals, "Radial basis function network for speech pattern classification", *Electronics Letters*, vol. 25, pp. 437–439, 1989.

[33] T.Poggio and S. Edelman, "A network that learns to recognize three-dimensional objects", *Nature*, vol. 343, pp. 899–910, 1990.

[34] K. Ng and R. P. Lippmann, "Practical characteristics of neural networks and conventional pattern classifiers", in R. P. Lippmann, J. E. Moody, and D. S. Touretzky, editor, *Advances in Neural Information Processing System 3*. Morgan Kaufmann, San Mateo, CA, 1991.

[35] M. Niranjan and F. Fallside, "Neural networks and radial basis functions in classifying static speech patterns", *Computer Speech and Language*, vol. 4, pp. 27–289, 1990.

[36] S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive bayesian feedback equalizer based on a radial basis function network", *in IEEE International Conference on Communications*, pp. 1267–1271, Chicago, IL, 1992.

[37] J. Cid-Sueiro and A. R. Figueiras-Vidal, "Improving conventional equalizers with neural networks", in J. Alspector, R. Goodman, and T. X. Brown, editor, *Applications of Neural networks to Telecommunications*. Lawrence Erlbaum, Hillsdale, NJ, 1993.

[38] T. Lo, H. Leung, and J. Litva, "Non-linear beamforming", *Electronics Letters*, vol. 27, Feb. 1991.

[39] T. Lo, T. Wong, and J. Litva, "A new technique for low-angle radar tracking", *Electronics Letters*, vol. 27, March 1991.

[40] T. Wong, T. Lo, H. Leung, J. Litva, E. Bosse, "Low-angle radar tracking using radial basis function neural network", *IEE Proc.-F*, vol. 140, pp. 323–328, Oct. 1993.

[41] J. Litva, "A new low-angle tracking technique", Technical Report Report 1335, Communications Research Centre, Ottawa, Canada, May 1980.

[42] A. D. Olver, "Basic properties of antennas", in A. W. Rudge et al, editor, *The Handbook of Antenna Design*. Peter Peregrinus, 1986.

[43] A. T. Moffet, "Minimum-redundancy linear arrays", *IEEE Trans. Antennas and Propagation*, vol. AP-16, March 1968.

[44] J. E. Freehafer, W. T. Fishback, W. H. Furry, and D. E. Kerr, "Theory of propagation in a horizontal stratified atmosphere", in D. K. Kerr, editor, *Propagation of Short Radio Wave*. McGraw-Hill, New Yory, 1951.

[45] B. J. Rook and J. Litva, "An improved cha algorithm for tracking low-angle target data", Technical Report 1356, Communications Research Centre, Ottawa, Canada, April 1982.

[46] L. P. Seidman, "Performance limitations and error calculations for parameter estimation", *Proc. IEEE*, vol. 58, pp. 644–652, May 1970.

[47] D. C. Rife, R. R. Boorstyn, "Multiple tone parameter estimation from discrete-time observations", *Bell Syst. Tech. Journ.*, vol. 55, pp. 1389-1410, Nov. 1976.

[48] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley and Sons, New York, 1968.

[49] J. A. Freeman and D. M. Skapura, *Neural Networks: Algorithms, Applications and Programming Techniques*, Addison-Wesley, 1992.

[50] R. J. Marks, et al., "Performance analysis of associative memories with nonlinearities in the correlation domain", *Applied Optics*, vol. 27, pp. 2900-2904, July 1988.

[51] D. R. Selviah, et al., "Correlating matched-filter model for analysis and optimisation of neural networks", *IEE Proceedings, Part F*, vol. 136, pp. 143-148, June 1989.

[52] T. Kohonen, *Self-Organization and Associative Memory, Third Edition*, Springer-Verlag, 1989.

[53] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition", *IEEE Electronic Computers*, vol. EC-14, pp. 326-334, 1988.

[54] M. A. Aizerman, E. M. Braverman, and L.I. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning", *Automatica i Telemekhanika*, vol. 25, pp. 917-936, 1964.

[55] E. Parzen, "On the estimation of a probability density function and mode", *Annals of Mathematical Statistics*, vol. 33, pp. 1065-1076, 1962.

[56] K. Funahashi, "On the approximate realization of continuous mappings by neural networks", *Neural Networks*, vol. 2, pp. 183-192, 1989.

[57] K. Hornit, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", *Neural Networks*, vol. 2, pp. 359–366, 1989.

[58] W. A. Light, "Some aspects of radial basis function approximation", in S. P. Singh, editor, *Approximation Theory, Spline Functions and Applications*. Kluwer Academic Publishers, Boston, MA, 1992.

[59] E. J. Hartman, J. D. Keeler, and J. M. Kowalski, "Layered neural networks with gaussian hidden units as universal approximation", *Neural Computation*, vol. 2, pp. 210–215, 1990.

[60] S. Haykin, *Adaptive Filter Theory, 2nd Ed.*, Prentice-Hall, Englewood Cliffs, NJ, 1991.

[61] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, 1983.

[62] S. Haykin, *Neural Networks — A comprehensive Foundation*, Macmillan College Publishing Company, New York, 1994.

[63] D. Lowe, "Adaptive radial basis function nonlinearities, and the problem of generalization", *in 1st IEE International Conference on Artificial Neural Networks*, pp. 171–175, London, UK, 1989.

[64] J. Rissanen, "Modelling by shortest data description", *Automatica*, vol. 14, pp. 465–471, 1978.

[65] M. Stone, "Cross-validatory choice and assessment of statistical predictions", *Royal Stat. Soc.*, vol. B36, pp. 111–133, 1974.

[66] T. Lo, J. Litva, H. Leung, "Determination of optimal rbf network structure by canonical subspace analysis", in D. W. Ruck, editor, *Science of Artificial Neural Networks*. The International Society for Optical Engineering, Orlendo, Florida, 1993.

[67] A. Blumer, A. Ehrenfeucht, D. Haussler, M. K. Warmuth. "Occam's razor", *Information Processing Letters*, vol. 24, pp. 377-380, 1987.

[68] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-posed Problems*, W. H. Winston, Washington, D. C., 1977.

[69] C. J. Stone, "Optimal global rates of convergence for non-parametric regression", *Ann. Stat.*, vol. 10, pp. 1040-1053, 1982.

[70] R. E. Howard et al., "An associative memory based on an electronic neural network architecture", *IEEE Trans. on Electron. Devices*, vol. 34, pp. 1553 1556, July 1989.

[71] H. P. Gaft and P. deVegvar, "A COMS implementation of a neural networks model", *in Proc. of the 1987 Stanford conference*, pp. 351-362, 1987.

[72] J. P. Sage et al., "An artificial neural network integrated circuit based on MNOS/CCD principles", *in Proc. Conf. Neural Networks for Computing*, 1986.

[73] M. A. Holler et al., "An electronically trainable artificial neural network (ETANN) with 10240 'floating gates' synapses", *in International Joint Conf. on Neural Networks*, pp. 191-196, San Diego, CA, 1989.

[74] D. Hammerstrom, "A VLSI architecture for high-performance, low-cost, or-chip learning", *in International Joint Conf. on Neural Networks*, pp. 537-544, San Diego, CA, 1989.

[75] U. Ramamcher, J. Beichter, and N. Bluls, "Architecture of a general-purpose neural signal processor", *in International Joint Conf. on Neural Networks*, pp. 443-446, Seattle, WA, 1991.

[76] T. Watanabe, K. Kimura, M. Aoki,TSakata, and K. Ito, "A single 1.5 v digital chip for a 10^6 synapses neural network", *IEEE Trans. on Neural Networks*, vol. 4, pp. 387-393, 1993.

[77] T. Lo, "Adaptive beam-space nulling of multipath signals", Master's thesis, McMaster University, Hamilton, Ontario, Canada, 1989.

[78] Titus Lo, Larry Lai, and John Litva, "Multipath propagation effects on low-angle radar tracking: An experimental evaluation", *in IEEE AP-S International Symposium*, 1990.

[79] Titus Lo and John Litva, "Characteristics of diffuse forward-scattering signals", *in Proc. IEEE IGARSS'91*, Espoo, Finland, 1991.

[80] H. Leung, T. Lo, and J. Litva, "Angle-of-arrival estimation in multipath environment using chaos theory", *Signal Processing*, vol. 31, 1993.

[81] Titus Lo, Henry Leung, John Litva, and Simon Haykin, "Fractal characterization of sea scattered signals", *IEE Proceedings Part F*, vol. 140, Aug. 1993.