Appl. Statist. (1999)
48, *Part* 2, *pp.*197–209

A basis function approach to position estimation using microwave arrays

Andrew R. Webb and Philip N. Garner

Defence Evaluation and Research Agency, Malvern, UK

[Received October 1996. Final revision May 1998]

Summary. We consider the problem of estimating the bearing of a remote object given measurements on a particular type of non-scanning radar, namely a focal-plane array. Such a system focuses incoming radiation through a lens onto an array of detectors. The problem is to estimate the angular position of the radiation source given measurements on the array of detectors and knowledge of the properties of the lens. The training data are essentially noiseless, and an estimator is derived for noisy test conditions. An approach based on kernel basis functions is developed. The estimate of the basis function weights is achieved through a *regularization* or *roughness penalty* approach. Choosing the regularization parameter to be proportional to the inverse of the input signal-to-noise ratio leads to a minimum prediction error. Experimental results for a 12-element detector array support the theoretical predictions.

Keywords: Cross-validation; Errors-in-variables model; Focal-plane arrays; Least squares approximation; Nonparametric regression; Radar signal processing; Radial basis functions; Regularization; Roughness penalties

1. Introduction

Radar (radio detection and ranging) is an electromagnetic system for the detection and location of objects that has found widespread application, both civil and military. These applications include surveillance and navigation, control and guidance of weapons, air traffic control, law enforcement (motor car speed monitoring), collision avoidance and satellite remote sensing.

A simple form of radar operates by emitting radiation from a transmitter antenna which is intercepted by a reflecting object which then reradiates the energy in all directions. This reradiated energy is collected by a receiver antenna and passed to a receiver. The distance to an object is determined by measuring the time taken for the transmitted radiation to travel to the reflecting object and back. The direction, or angular position, of an object is usually obtained by scanning a narrow antenna beam over the field of view and recording the antenna position at which the signal is detected. It is this problem of source position estimation which we address in this paper.

An alternative to the single-antenna scanning radar is the *array* radar comprising an array of spatially distributed receivers which is not mechanically scanned. Measurements made on these receivers are combined to perform direction finding. In this paper, we use data from a particular type of radar array, namely a *focal-plane array*, in which electromagnetic detectors are positioned on the focal-plane of a lens. The biological equivalent of this is the system of

Address for correspondence: Andrew R. Webb, Room EX24, Defence Evaluation and Research Agency, St Andrews Road, Malvern, Worcs., WR14 3PS, UK.
E-mail: webb@signal.dra.hmg.gb

receptors in the retina of the eye, namely the rods and cones. The main differences between the focal-plane array and its biological counterpart are the array size and resolution (the ability to resolve closely spaced objects): the human eye has about 10^8 receptors, each with an angular resolution of about 0.005°; our focal-plane array comprises 12 sensors, each with an angular resolution of 30°.

The problem that we are addressing in this paper is a particular type of inverse problem, namely one in image restoration. Here our 'image' comprises the set of measurements on the focal-plane array elements. Given this image, together with knowledge of the focusing system (in particular, properties of the lens), we wish to reconstruct the radiating scene, i.e. we wish to reconstruct the spatial distribution of scatterers that gives rise to the measured image. However, as discussed by Andrews and Hunt (1977), the image restoration problem is ill conditioned or ill posed in that a small change in the image may lead to a large change in the restored object. Therefore, noise in the image means that a solution must be selected from an infinite family of candidate solutions. This requires prior knowledge of the properties of admissible solutions (e.g. bounds on the extent of the object, positivity constraints and smoothness conditions). Here, our constraint is that the scene consists of a single scatterer whose position we wish to estimate.

The problem of estimating the bearing of a source using measurements made on a radar focal-plane array has been treated, both theoretically and experimentally, in previous work (Webb, 1991). A maximum likelihood approach was adopted and error in the estimate of the bearing of the source was evaluated as a function of source bearing. Theoretical predictions were compared with experimental results. In this paper we adopt an alternative approach based on kernel basis functions. This is motivated by a hardware requirement. The array of focal-plane sensors is fabricated on a silicon wafer. Recent research (Collins *et al.*, 1994) has shown that kernel basis functions can also be implemented using the same processing technology. Thus, the sensors and processing can be integrated in hardware, giving cost and size advantages.

The purpose of this paper is to present a development of a basis function approach, driven by the source bearing estimation application, and to apply it to sensor measurements. Section 2 briefly reviews radar focal-plane technology and the problem that has motivated this work. Section 3 presents the results of the basis function approach to regression when there is noise on the predictor variables and considers the application to the microwave imaging problem (with further details in Appendix A). It is related to the roughness penalty approach to nonparametric regression. The main result of this section is that we obtain a ridge regression type of solution for the weights of the basis function approach, with the ridge parameter inversely proportional to a signal-to-noise ratio term. The consequence of this is that we do not need to perform any search procedure to determine the ridge parameter. It can be set by measuring the signal-to-noise ratio of the radar system. Section 4 describes experimental results of a study on position estimation, using a 12-element detector array.

The data analysed in this paper can be accessed at

`http://www.blackwellpublishers.co.uk/rss/`

2. Description of the problem

2.1. Radar focal-plane arrays

The sensor system that we are considering in this paper is a focal-plane array of microwave detectors (Alder *et al.*, 1990). This is illustrated in Fig. 1. The two principal elements are a

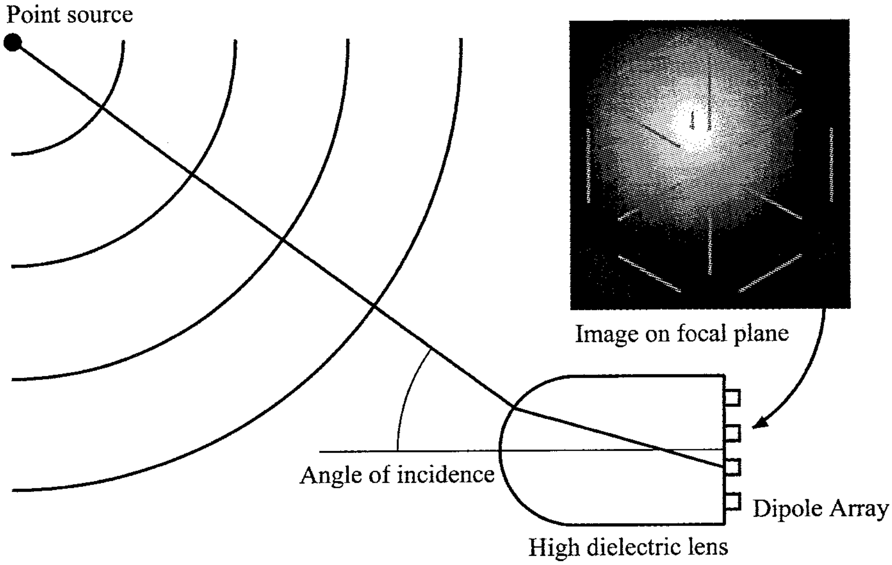

Fig. 1. Focal-plane array concept

lens and an array of detectors. The lens focuses incoming radiation onto the focal-plane. This produces an image which is sampled by an array of detectors (denoted by the small squares on the back of the lens in Fig. 1). An example of an image of a point source is also shown in Fig. 1 (shown as an intensity image). It consists of a bright spot that decays in intensity away from a single peak value.

In practical applications the point source is remote from the array. It may be a moving object, such as an aircraft, reflecting radiation transmitted from another radar. Alternatively, the point source may be transmitted from a satellite and the focal-plane array could be on a moving platform.

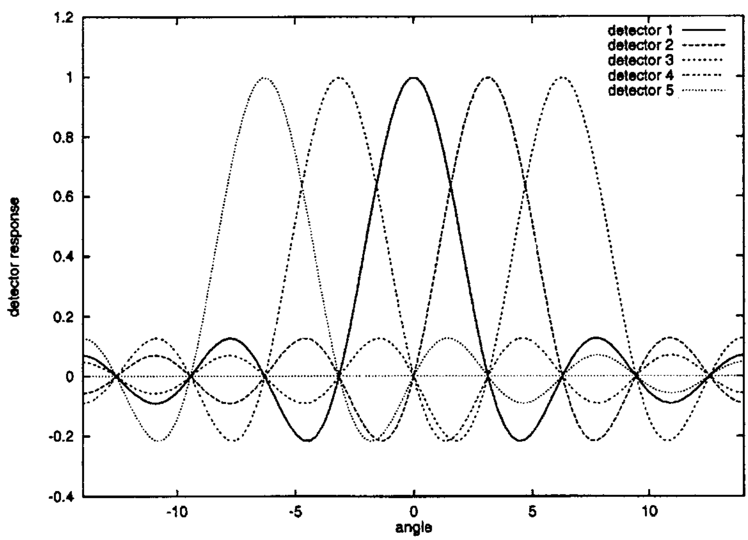

Fig. 2. Array responses for an idealized five-element linear array as a function of angle

As the remote point source moves relative to the array, the image moves across the detectors in the focal-plane (the detectors are denoted by straight lines in the image) and consequently the responses of the detectors change. The response of a detector as a function of the position of the source is called the *point spread function*. Fig. 2 shows idealized point spread functions of an array of five linear array elements as a function of position. Each element has a peak response in a specific direction (termed the *direction of look*), but there is some (residual) response to all directions.

An important factor is that complete arrays can be implemented within a small area of low cost silicon. This provides the main motivation for the current study in which we seek an approach for analysing the detector outputs that can also be integrated with the sensors and implemented in silicon on the focal-plane. This would provide a physically compact 'front end' combining sensors and signal processing. This is very desirable since it reduces the physical size, the weight and the cost of manufacture.

2.2. Description of the problem

In the particular application that we are addressing, the scene consists of a single radiating source of amplitude A at position θ. The measurements on the array at a particular time can be represented by the vector $\mathbf{z}$ given by

$$\mathbf{z} = A\,\mathbf{h}(\theta) + \boldsymbol{\xi}$$

where $\boldsymbol{\xi}$ is a noise vector and $\mathbf{h}$ is the *point spread function* of the lens–array combination: it is the response of the array to a source of unit amplitude at position θ. Thus, we assume that measurements comprise a scaled and corrupted (by additive noise) point spread function. We assume that the point spread function is known (obtained by calibration of the system over the range of values of θ of interest). In practice, calibration is performed by placing the array in an anechoic chamber (an echoless room so that there is no extraneous signal) and recording the array signals and source position as a transmitting source is moved relative to the array. Measurements are made at discrete steps in angle θ at a very high signal-to-noise ratio. The position variable θ may comprise two angular co-ordinates — the azimuth and the elevation.

The data analysis problem is simply this: given a measurement $\mathbf{z}$ and knowledge of the point spread function $\mathbf{h}$, estimate the angle of arrival, θ, of the radiation.

Assuming a multivariate, zero-mean normal distribution for the noise $\boldsymbol{\xi}$, a maximum likelihood approach may be implemented (see Webb (1991) for a description of this approach and an evaluation on real radar data). However, two of the disadvantages of the maximum likelihood approach that limit its practical implementation in the application of interest are the physical storage of the point spread function $\mathbf{h}(\theta)$ and the amount of computation involved, both of which may be excessive.

An approach based on function approximation is proposed, i.e. for a measurement vector $\mathbf{z}$, we seek a function $\mathbf{f}$ such that

$$\hat{\theta} = \mathbf{f}\{\mathbf{z}(\theta)\}$$

is, in some sense, a good estimate of position θ. This is not a standard non-linear regression problem, since the noise arises on the predictor variables, not the response variables, and is similar to errors-in-variables models. Indeed, nonparametric estimators based on kernel basis functions (see Section 3.1) have been considered (e.g. Fan and Truong (1993) and Ioannides and Alevizos (1997)). The difference between that work and the approach taken here is that we have noiseless training data available to construct the estimator, but we wish to apply them to noisy test conditions.

The training data for constructing the predictor $\mathbf{f}$ comprise measurements on the point spread function: $\{(\mathbf{x}_i, \theta_i); i = 1, \ldots, N\}$, where $\mathbf{x}_i$ is a vector of measurements on the independent variables (array calibration measurements in this problem) and θ_i is the response variable (position). These are measured essentially without error. The test data comprise measurements $\mathbf{z}$ (a scaled and corrupted sample from the training distribution).

The fact that the training and test distributions differ deserves some comment. In our practical application, noise is present. This is characterized by the signal-to-noise ratio which can be measured by the sensor and hence is considered to be known. We require a predictor for a range of values of the signal-to-noise ratio. An obvious way of achieving this is to gather data that are representative of several typical signal-to-noise ratio conditions and to design a predictor for each condition. An alternative approach is to design a predictor by using essentially noiseless data, but to include as a parameter in the model the signal-to-noise ratio of the test conditions. This means that only a single set of training data needs to be gathered. In the following section we show how this may be achieved.

This problem is one example from a wider class in discrimination and regression in which the expected operating conditions (the test conditions) differ from the training conditions in a known way. For example, in a discrimination problem, the class priors may differ considerably from the values estimated from the training data. In a least squares approach, we may compensate for this by modifying the sum-of-squares error criterion appropriately. Also, we may make allowance for expected *population drift* (Hand, 1997) by modifying the error criterion. In the problem considered here, the training conditions are considered 'noiseless' and the test conditions differ in that there is noise (of known variance) on the data. Again, this can be taken into account by modifying the sum-of-squares error criterion, as we show in Section 3.

3. A basis function approach

We present the results for a basis function approach to the bearing estimation problem in two stages. Firstly (in Section 3.1), we consider the situation where we have a noiseless training set and we wish to design a predictor for test conditions in which there is noise of known variance on the regressors.

In Section 3.2, we consider the situation that is appropriate to our application, where the test data differ from the training conditions by a scaling as well as the addition of noise. Here, we state the results, reserving the detail for Appendix A.

3.1. Solution for basis functions approach

Let the design set or, in the microwave array example, the calibration set be $\{(\mathbf{x}_i, \boldsymbol{\theta}_i); i = 1, \ldots, N\}$ where $\mathbf{x}_i \in \mathbb{R}^n$ are the predictors and $\boldsymbol{\theta}_i \in \mathbb{R}^m$ are vectors of m (usually 2 or 3) response variables (angular co-ordinates). For noise $\boldsymbol{\xi}$ on the predictor variables characterized by the distribution $p(\boldsymbol{\xi})$, we seek to minimize the expected sum-of-squares error S, given by

$$S = \frac{1}{N} \sum_{i=1}^{N} \int |\boldsymbol{\theta}_i - \mathbf{f}(\mathbf{x}_i + \boldsymbol{\xi})|^2 \, p(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}$$

with respect to the function $\mathbf{f}$. Usually, this will have a fixed number of parameters to optimize.

We assume that the function $\mathbf{f}$ is a smooth function of its argument. This is justified by referring to the noiseless case in which there is a smooth relationship between $\boldsymbol{\theta}$ and $\mathbf{x}$ (imposed by the lens). Thus, we may expand $\mathbf{f}$ as a Taylor series. The sum-of-squares error S simplifies to (Webb, 1994; Bishop, 1995)

$$S = \frac{1}{N} \sum_{i=1}^{N} |\boldsymbol{\theta}_i - \mathbf{f}(\mathbf{x}_i)|^2 + \frac{\sigma^2}{N} \sum_{i=1}^{N} \|\mathbf{J}^i\|^2 \tag{1}$$

where $\mathbf{J}^i$ is an $m \times n$ matrix given by

$$J^i_{jk} = \left.\frac{\partial f_j}{\partial x_k}\right|_{\mathbf{x}_i}.$$

In the above, we have assumed that $E\{\boldsymbol{\xi}\} = \mathbf{0}$ and $E\{\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}\} = \sigma^2\mathbf{I}$ (where T denotes transpose) and that perturbations are small so that higher order terms in the Taylor expansion may be neglected.

Thus, under the assumption of low noise, minimizing the expected square error over the test data is equivalent to minimizing a modified error criterion over the training set, i.e. we may write the sum-of-squares error (1) as the sum of two terms: the average square error on the training samples and a regularization term, with regularization parameter proportional to σ^2. It is a penalized sum-of-squares criterion, identical in form with that minimized in a roughness penalty approach to smoothing (Green and Silverman, 1994).

We shall model the function $\mathbf{f}$ as a sum of l basis functions of the form

$$f_j(\mathbf{x}; \mathbf{W}) = \sum_{i=1}^{l} w_{ij}\, \phi_i(\mathbf{x}) \qquad j = 1, \ldots, m$$

where ϕ_i, $i = 1, \ldots, l$, are non-linear functions of $\mathbf{x}$ and the weights w_{ij}, $i = 1, \ldots, l$, $j = 1, \ldots, m$, are a set of parameters to be determined, i.e. we assume that each component of the transformation is a different linear combination of the same set of basis functions. We may write

$$\mathbf{f}(\mathbf{x}) = \mathbf{W}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}) \tag{2}$$

where $\mathbf{W}$ is the $l \times m$ matrix $(\mathbf{W})_{ij} = w_{ij}$ and $\boldsymbol{\phi}$ is the vector with jth component $\phi_j(\mathbf{x})$. The form of the functions ϕ_j and the number of basis functions must be specified or determined from the data.

By writing $\mathbf{f}$ in the form (2), the problem of minimizing S, given by equation (1), with respect to $\mathbf{f}$ reduces to a problem of estimating the matrix $\mathbf{W}$. Denoting by $\boldsymbol{\Phi}$ the $l \times N$ matrix with kth column $\boldsymbol{\phi}(\mathbf{x}_k)$ and $\mathbf{G}$ the $m \times N$ matrix with kth column $\boldsymbol{\theta}_k$, then the sum-of-squares error (1) may be written for model (2) for $\mathbf{f}$ as

$$S = \frac{1}{N} \operatorname{tr}\{(\mathbf{G}^{\mathrm{T}} - \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{W})(\mathbf{G}^{\mathrm{T}} - \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{W})^{\mathrm{T}}\} + \frac{\sigma^2}{N} \operatorname{tr}(\mathbf{W}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{W}) \tag{3}$$

where tr denotes the matrix trace operation and the $l \times l$ matrix $\boldsymbol{\Lambda}$ has (k, j) component

$$\Lambda_{k,j} = \sum_{i=1}^{N} \left(\frac{\partial \phi_k}{\partial \mathbf{x}}\right)^{\mathrm{T}}_{\mathbf{x}_i} \left(\frac{\partial \phi_j}{\partial \mathbf{x}}\right)_{\mathbf{x}_i}$$

where

$$\left(\frac{\partial \phi_j}{\partial \mathbf{x}}\right)_{\mathbf{x}_i} = \left(\frac{\partial \phi_j}{\partial x_1}, \ldots, \frac{\partial \phi_j}{\partial x_n}\right)^{\mathrm{T}}\Bigg|_{\mathbf{x}_i}$$

is the vector of derivatives of the jth basis function with respect to the input variables, evaluated at $\mathbf{x}_i$.

Equation (3) shows that S is quadratic in the parameters of the transformation $\mathbf{f}$, i.e. the matrix $\mathbf{W}$. Minimizing S given by equation (3) with respect to $\mathbf{W}$ leads to the solution for $\mathbf{W}$

$$\mathbf{W} = (\mathbf{\Phi\Phi}^{\mathrm{T}} + \sigma^2 \mathbf{\Lambda})^{-1} \mathbf{\Phi G}^{\mathrm{T}}, \tag{4}$$

assuming that the inverse exists.

In equation (4), the matrices $\mathbf{\Phi}$, $\mathbf{G}$ and $\mathbf{\Lambda}$ are calculated from the training data; σ^2 is representative of the test conditions. The solution for $\mathbf{W}$ above has the form of a ridge regression estimate. In regularized multiple linear regression (Brown, 1993) in which we have a $p \times N$ (p variables and N samples) data matrix $\mathbf{X}$ and a vector of response variable measurements $\mathbf{y}$, the solution for the ridge regression estimate for the coefficients β is

$$\beta = (\mathbf{XX}^{\mathrm{T}} + \alpha \mathbf{I}_p)^{-1} \mathbf{Xy}^{\mathrm{T}}$$

where α is the ridge parameter and $\mathbf{I}_p$ is the $p \times p$ identity matrix. In our example, the ridge term is not simply proportional to the identity matrix but has some structure described by the matrix $\mathbf{\Lambda}$. For the special case of a linear transformation (so that $l = n$ and $\phi_i(\mathbf{x}) = x_i$, $i = 1, \ldots, n$) $\mathbf{\Lambda}$ reduces to N times the identity matrix (with the solution for $\mathbf{W}$ becoming the standard ridge regression solution).

3.2. Application to microwave imaging

In Section 3.1, a predictor was developed for test conditions that differed from the training conditions by the addition of a noise term. We now turn to the microwave array problem. Here, the test data do not differ from the training data simply by the addition of a noise term, but they are scaled (by an amplitude A) and noise is added. Thus, we wish to make a prediction based on a measurement

$$\mathbf{z} = A\mathbf{x} + \boldsymbol{\xi}$$

for signal amplitude A and noise $\boldsymbol{\xi}$.

The modifications to the previous analysis (detailed in Appendix A) are that the sum-of-squares error term, equation (1), becomes

$$S = \frac{1}{N} \sum_{i=1}^{N} |\theta_i - \mathbf{f}(\mathbf{x}_i)|^2 + \frac{\sigma^2}{NA^2} \sum_{i=1}^{N} \|\mathbf{J}^i \mathbf{D}^i\|^2 \tag{5}$$

where $\mathbf{D}^i = \mathbf{D}(\mathbf{x}_i)$ and the $n \times n$ matrix $\mathbf{D}(\mathbf{x})$ is given by

$$\mathbf{D} = \frac{1}{|\mathbf{x}|}(\mathbf{I} - \hat{\mathbf{x}}\hat{\mathbf{x}}^{\mathrm{T}}).$$

Equation (5) is again a penalized sum-of-squares error which is quadratic in the matrix $\mathbf{W}$. The solution for $\mathbf{W}$ minimizing S in equation (5) is now given by

$$\mathbf{W} = \left(\mathbf{\Phi\Phi}^{\mathrm{T}} + \frac{\sigma^2}{A^2} \tilde{\mathbf{\Lambda}}\right)^{-1} \mathbf{\Phi G}^{\mathrm{T}} \tag{6}$$

where $\tilde{\Lambda}$ is given by

$$\tilde{\Lambda}_{k,j} = \sum_{i=1}^{N} \left(\frac{\partial \phi_k}{\partial \mathbf{x}}\right)^{\mathrm{T}}_{\hat{\mathbf{x}}_i} \mathbf{D}^i (\mathbf{D}^i)^{\mathrm{T}} \left(\frac{\partial \phi_j}{\partial \mathbf{x}}\right)_{\hat{\mathbf{x}}_i}$$

and the derivatives are evaluated for normalized input pattern $\hat{\mathbf{x}}_i$. Thus, again we have a ridge regression type of solution for $\mathbf{W}$ with ridge parameter inversely proportional to the 'signal-to-noise ratio' (A^2/σ^2). (Strictly, the signal-to-noise ratio involves the magnitude of the point spread function which is incorporated in the definition of the matrix $\mathbf{D}$.)

This is the main result of this section. It states that the ridge parameter may be set provided that we know the signal-to-noise ratio. In practice, this may be measured by the array. Thus, rather than search for an optimum ridge parameter (e.g. by calculating a cross-validation error as a function of the ridge parameter (Brown, 1993)), we may set its value directly from physical measurements.

In the above, the matrix $\mathbf{\Phi}$ is the matrix of non-linear values for *normalized* input data, $\hat{\mathbf{x}}_i$ ($\mathbf{\Phi} = (\phi(\hat{\mathbf{x}}_1), \ldots, \phi(\hat{\mathbf{x}}_N))$). Also, the prediction of position for measurement $\mathbf{z}$ is

$$\tilde{\theta} = \mathbf{W}^{\mathrm{T}} \phi(\hat{\mathbf{z}}),$$

i.e. the measurement vector must be normalized to unit magnitude.

4. Experimental study

The purpose of the experiments described in this section is to evaluate a basis function approach (providing a comparison with a maximum likelihood approach) and to assess whether a good choice for the ridge parameter can be made on the basis of the signal-to-noise ratio.

4.1. Implementational details

The basis functions that we choose to model the non-linear transformation are a set of radially symmetric kernel functions of the form

$$\phi_k(\mathbf{x}) = \phi(r)$$

where $\phi(r)$ is a kernel function, with $r = |\mathbf{x} - \mathbf{c}_k|$ and $\mathbf{c}_k$, $k = 1, \ldots, l$, a 'centre' or knot. Such radially symmetric functions are termed *radial basis functions.*

Many forms for the kernel function have been considered. The two most popular are the normal kernel and the thin plate spline. Given that we are motivated in part by a silicon realization of the technique (Collins *et al.*, 1994), we shall take the functions $\phi(r)$ to be Gaussian kernels of the form

$$\phi(r) = \exp(-r^2/2h^2)$$

where h is the kernel width.

As in kernel density estimation, it is important to make an appropriate choice for the smoothing parameter h. In many cases, a subjective choice may be satisfactory. However, we have chosen to adopt an automatic approach based on least squares cross-validation. The smoothing parameter was varied over a range of values and the value chosen for which the cross-validation estimate of the sum-of-squares error was a minimum.

The centres were chosen by using a k-means procedure and several models with different numbers of centres were assessed.

4.2. Data

The training data comprised the detector outputs of an array of 12 detectors, positioned on a triangular grid in the focal-plane of a lens (Fig. 3). The operating frequency was 35 GHz (wavelength 0.08 m) and the lens aperture was 20 mm, giving a beam width of about 30°. Measurements were taken at 1°-spacing in azimuth and elevation over ±30° at a signal-to-noise ratio of about 45 dB. Thus, there are 3721 training samples. All of these were used to train the model. For the measured data, the training data are essentially 'noiseless' (recorded at a signal-to-noise ratio that was much higher than the expected operating levels). Note also that the data, although lying in a 12-dimensional data space, are constrained to lie on a two-dimensional manifold within that space, characterized by the two angular co-ordinates.

For the test data, measurements were made at five signal-to-noise ratios (representative of expected operating conditions) at 21 points spaced at 3° separation in azimuth and at 0° elevation (a line scan across the centre of the array). 600 samples were taken at each position. For each of the test data sets, the value of σ^2/A^2 (the ratio of the variance to the square of the amplitude scaling of that test set relative to the calibration set) was calculated. This should be the optimum value of the regularization parameter. The five test sets used in the experiments had values of $\log_{10}(\sigma^2/A^2)$ of 1.48, 2.11, 2.49, 2.96 and 3.27 with corresponding signal-to-noise ratios of 37.3 dB, 31.0 dB, 27.2 dB, 22.5 dB and 19.4 dB, with the lowest value of signal-to-noise ratio being the most important from an applications point of view.

4.3. Results

For a given model complexity (number of basis functions and associated cross-validation value of the smoothing parameter), the weight matrix **W** was calculated for several values of the 'ridge parameter' k, i.e. we evaluate

$$\mathbf{W} = (\mathbf{\Phi\Phi}^{\mathrm{T}} + k\tilde{\mathbf{\Lambda}})^{-1}\mathbf{\Phi G}^{\mathrm{T}}$$

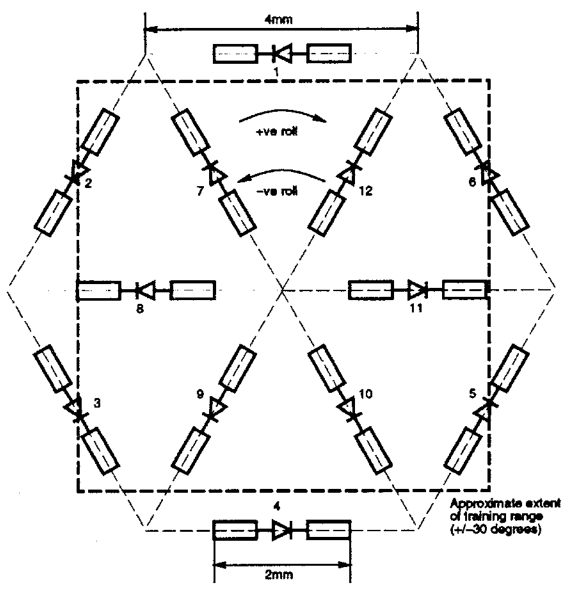

Fig. 3. Array configuration on the focal-plane of the lens, consisting of 12 detectors (▭┤◁├▭) whose centres are on a triangular grid

and calculate the mean-square error on the test set as a function of k. These errors are plotted in Fig. 4 for the five test sets and for a model with 144 basis functions. The minimum error is achieved with a value of k approximately equal to that of the operating conditions, although for the high test set value the minimum is shallow. For example, the results for the test set at a signal-to-noise ratio of 19.4 dB should have an optimum value of $\log_{10}(k)$ of 3.27. The curve indeed has a minimum (the top curve in Fig. 4) in that region, although it is fairly shallow. However, choosing k too small leads to a degradation in performance. The test set at 37.3 dB has an optimum value of $\log_{10}(k)$ of 1.48. The results (bottom curve) show a very broad minimum in this region. Thus, we conclude that basing the choice of the regularization parameter on the signal-to-noise ratio in the test conditions leads to close to optimal

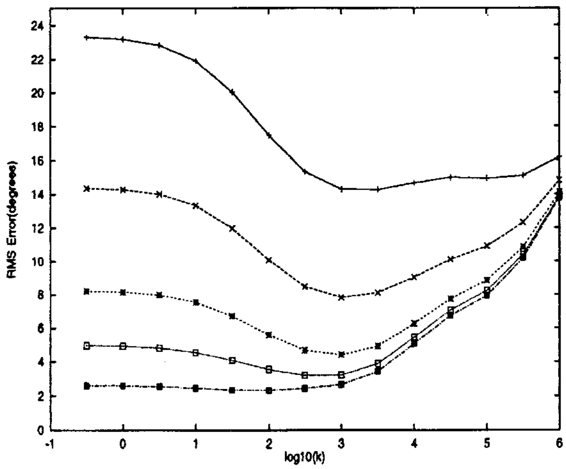

Fig. 4. Root-mean-square error in position estimates as a function of ridge parameter k for the five test sets (number of basis functions $l = 144$), labelled according to values of $\log_{10}(\sigma^2/A^2)$: +, $\log_{10}(\sigma^2/A^2) = 3.27$ (signal-to-noise ratio SNR = 19.4 dB); ×, $\log_{10}(\sigma^2/A^2) = 2.96$ (SNR = 22.5 dB); ⋇, $\log_{10}(\sigma^2/A^2) = 2.49$ (SNR = 27.2 dB); □, $\log_{10}(\sigma^2/A^2) = 2.11$ (SNR = 31.0 dB); ■, $\log_{10}(\sigma^2/A^2) = 1.48$ (SNR = 37.3 dB)

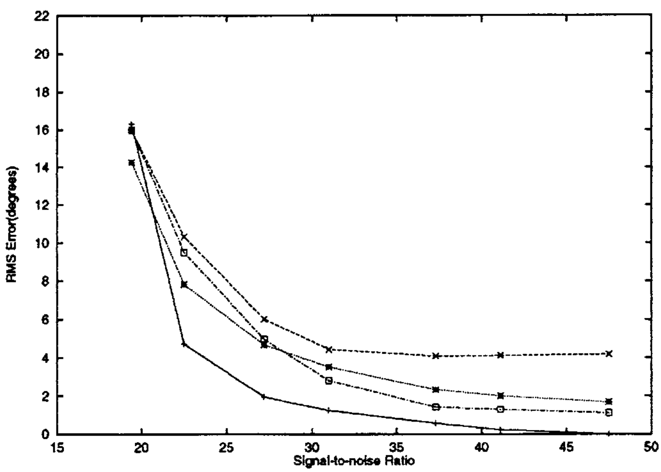

Fig. 5. Root-mean-square in position estimate as a function of signal-to-noise ratio for radial basis function (×, 25 centres; ⋇, 144 centres; □, 225 centres) and maximum likelihood (+) models

performance for this predictor. Experiments with different numbers of centres also showed the same behaviour.

Fig. 5 plots the root-mean-square error in prediction as a function of signal-to-noise ratio for the radial basis function model (with 25, 144 and 225 centres) and for the maximum likelihood approach. For the radial basis function model, the root-mean-square error in prediction generally decreases with the number of basis functions (up to the maximum of 225 considered) but is greater than the maximum likelihood estimator. However, at the lower end of the range (the most important for the application under consideration), all methods give similar performance (approximately half a beam width error). However, we note that the number of data samples stored for the maximum likelihood method (3721) is much greater than that for the largest radial basis function model (225). Thus, the degradation at high signal-to-noise ratio is due to some extent to the number of data samples used to characterize the point spread function.

5. Discussion and conclusions

The purpose of this paper has been to present a kernel-based approach to prediction in which the test conditions differ from the training conditions in a known way. We have applied the methodology to the problem of position estimation of a radiating source by using the outputs of a focal-plane array of detectors. In this situation, the test conditions differ from the training data by an unknown scaling and the addition of noise. For this particular application, the kernel-based approach is attractive in that it has the potential for implementation in circuitry which can be integrated in the focal-plane of the sensor.

The weights in the basis function approach are obtained via a ridge regression procedure in which the ridge or regularization parameter is determined from the test conditions (it is inversely proportional to the signal-to-noise ratio). Experimental data from a microwave detector array have been used to verify these results and the main conclusion is that it is possible to compensate for noisy test conditions through the implementation of a ridge regression type of solution for the model parameters.

This study has formed part of a larger programme on the development of radar focal-plane array technology and techniques for processing the outputs of such arrays. There are several theoretical and experimental areas for further work. These include assessing the performance of source position estimation methods on arrays of different sizes and configurations, investigating the multisource location problem (nonparametric regression techniques can be used as part of a multisource location system) and developing techniques for operation at lower signal-to-noise ratios.

Acknowledgements

We thank Mike Dixon and Chris Alder for providing experimental data.

Appendix A: Basis function solution for non-homogeneous noise

The results of Section 3.1 apply when we have a design set $\{(\mathbf{x}_i, \theta_i)\}$ and we seek a fitting function for the situation when the predictor variables are corrupted by noise ξ. In the radar problem, we wish to make a prediction based on a measurement

$$\mathbf{z} = A\mathbf{x} + \xi$$

for signal amplitude A and noise $\boldsymbol{\xi}$. Thus, there is some unknown scaling (we do not know the amplitude of the source) and corruption by noise.

Making the assumptions that $E\{\boldsymbol{\xi}\} = \mathbf{0}$ and $E\{\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}\} = \sigma^2\mathbf{I}$ and that the signal-to-noise ratio is high, $A^2/(|\mathbf{x}|\sigma)^2 \gg 1$, then

$$\hat{\mathbf{z}} \triangleq \mathbf{z}/|\mathbf{z}|$$

can be written

$$\hat{\mathbf{z}} = \hat{\mathbf{x}} + \beta$$

where $\hat{\mathbf{x}} = \mathbf{x}/|\mathbf{x}|$ and β is a noise term

$$\beta = \frac{1}{A|\mathbf{x}|}(\mathbf{I} - \hat{\mathbf{x}}\hat{\mathbf{x}}^{\mathrm{T}})\boldsymbol{\xi} + \mathcal{O}(\xi^2)$$

with the properties that

$$E\{\eta\} = \mathcal{O}(\sigma^2),$$

$$E\{\eta\eta^{\mathrm{T}}\} = \frac{\sigma^2}{A^2|\mathbf{x}|^2}\{\mathbf{I} - \hat{\mathbf{x}}\hat{\mathbf{x}}^{\mathrm{T}}\}.$$

Thus, provided that we take as our training data $\{\hat{\mathbf{x}}_i, \boldsymbol{\theta}_i\}$ and we normalize our measurement $\mathbf{z}$ before making a prediction, the basis function approach described in Section 3.1 applies, with the important exception that the noise is no longer homogeneous, i.e. the noise depends on the measurement vector $\mathbf{x}$, or equivalently the angle of arrival. In this case, we seek to minimize the error

$$S = \frac{1}{N}\sum_{i=1}^{N}\int |\boldsymbol{\theta}_i - \mathbf{f}(\hat{\mathbf{x}}_i + \eta_i)|^2 p(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}$$

where η_i is the noise term at $\mathbf{x}_i$. Expanding using Taylor's series (assuming a high signal-to-noise ratio so that terms above second order may be neglected) gives

$$S = \frac{1}{N}\sum_{i=1}^{N}|\boldsymbol{\theta}_i - \mathbf{f}(\hat{\mathbf{x}}_i)|^2 - \frac{2}{N}\sum_{i=1}^{N}(\boldsymbol{\theta}_i - \mathbf{f}(\hat{\mathbf{x}}_i))^{\mathrm{T}}\mathbf{J}^i E\{\eta_i\} + \frac{\sigma^2}{NA^2}\sum_{i=1}^{N}\{\|\mathbf{J}^i\mathbf{D}^i\|^2 - \mathrm{tr}(\mathbf{D}^i\mathbf{Q}^i\mathbf{D}^i)\} \tag{7}$$

where $\mathbf{D}^i = (\mathbf{I} - \hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^{\mathrm{T}})/|\mathbf{x}_i|$, $\mathbf{J}^i$ is an $m \times n$ matrix and $\mathbf{Q}^i$ is an $n \times n$ matrix, given by

$$J^i_{jk} = \left.\frac{\partial f_j}{\partial x_k}\right|_{\hat{\mathbf{x}}_i},$$

$$Q^i_{jk} = (\boldsymbol{\theta}_i - \mathbf{f}(\hat{\mathbf{x}}_i))^{\mathrm{T}}\left.\frac{\partial^2\mathbf{f}}{\partial x_j \partial x_k}\right|_{\hat{\mathbf{x}}_i}.$$

Assuming that in the region of the minimum $\mathbf{f}(\hat{\mathbf{x}}) = \boldsymbol{\theta}_i + \mathcal{O}(\sigma)$ and $E\{\eta_i\} \approx \mathcal{O}(\sigma^2)$, then the error simplifies to

$$S = \frac{1}{N}\sum_{i=1}^{N}|\boldsymbol{\theta}_i - \mathbf{f}(\hat{\mathbf{x}}_i)|^2 + \frac{\sigma^2}{NA^2}\sum_{i=1}^{N}\|\mathbf{J}^i\mathbf{D}^i\|^2. \tag{8}$$

For a transformation $\mathbf{f}$ given by equation (2), we have

$$S = \frac{1}{N}\,\mathrm{tr}\{(\mathbf{G}^{\mathrm{T}} - \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{W})(\mathbf{G}^{\mathrm{T}} - \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{W})^{\mathrm{T}}\} + \frac{\sigma^2}{NA^2}\,\mathrm{tr}(\mathbf{W}^{\mathrm{T}}\tilde{\boldsymbol{\Lambda}}\mathbf{W})$$

where $\tilde{\boldsymbol{\Lambda}}$ has (k, j) component

$$\tilde{\boldsymbol{\Lambda}}_{k,j} = \sum_{i=1}^{N}\left(\frac{\partial\phi_k}{\partial\mathbf{x}}\right)^{\mathrm{T}}_{\hat{\mathbf{x}}_i}\mathbf{D}^i(\mathbf{D}^i)^{\mathrm{T}}\left(\frac{\partial\phi_j}{\partial\mathbf{x}}\right)_{\hat{\mathbf{x}}_i}$$

with solution for $\mathbf{W}$ that minimizes V given by

$$\mathbf{W} = \left(\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}} + \frac{\sigma^2}{A^2}\tilde{\mathbf{\Lambda}}\right)^{-1} \mathbf{\Phi}\mathbf{G}^{\mathrm{T}}.$$

References

Alder, C. J., Brewitt-Taylor, C. R., Dixon, M., Hodges, R. D., Irving, L. D. and Rees, H. D. (1990) Lens-fed microwave and millimetre-wave receivers with integral antennas. In *Proc. 20th Eur. Microwave Conf., Budapest*. Tunbridge Wells: Kent Microwave Exhibitions and Publishers.

Andrews, H. C. and Hunt, B. R. (1977) *Digital Image Restoration*. Englewood Cliffs: Prentice Hall.

Bishop, C. M. (1995) Training with noise is equivalent to Tikhonov regularisation. *Neur. Computn*, **7**, 108–116.

Brown, P. J. (1993) *Measurement, Regression, and Calibration*. Oxford: Clarendon.

Collins, S., Marshall, G. F. and Brown, D. R. (1994) An analogue radial basis function circuit using a compact euclidean distance calculator. In *Proc. ISCAS '94: IEEE Int. Symp. Circuits and Systems, London*, vol. 6, pp. 233–236. New York: Institute of Electrical and Electronic Engineers.

Fan, J. and Truong, Y. (1993) Nonparametric regression with errors in variables. *Ann. Statist.*, **21**, 1900–1925.

Green, P. J. and Silverman, B. W. (1994) *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*. London: Chapman and Hall.

Hand, D. J. (1997) *Construction and Assessment of Classification Rules*. Chichester: Wiley.

Ioaniddes, D. A. and Alevizos, P. D. (1997) Nonparametric regression with errors in variables and applications. *Statist. Probab. Lett.*, **32**, 35–43.

Webb, A. R. (1991) Point-source location using a millimetre wave focal-plane array radar. *IEE Proc.* F, **138**, 581–591.

———(1994) Functional approximation in feed-forward networks: a least-squares approach to generalisation. *IEEE Trans. Neur. Netwks*, **5**, 363–371.