Computers Math. Applic. Vol. 24, No. 12, pp. 139–148, 1992
Printed in Great Britain.

0097-4943/92 $5.00 + 0.00

KRIGING, COKRIGING, RADIAL BASIS FUNCTIONS AND THE ROLE OF POSITIVE DEFINITENESS

DONALD E. MYERS
Department of Mathematics, University of Arizona
Tucson, Arizona 85721, U.S.A.

Abstract—There are at least three developments for interpolators that lead to the same functional form for the interpolator; the thin plate spline, radial basis functions and the regression method known as kriging. The key to the interrelationship lies in the positive definiteness of the kernel function. Micchelli has shown that a weak form of positive definiteness is sufficient to ensure a unique solution to the system of equations determining the coefficients in the interpolator. Both the positive definiteness and the interpolator can be extended to vector valued functions via the kriging approach which is also independent of the dimension of the underlying space. The kriging approach leads naturally to various methods for simulation as well.

1. INTRODUCTION

Given data locations $x_1, \ldots, x_n$ in k-dimensional Euclidean space and observed data $Z(x_1), \ldots, Z(x_n)$, consider an interpolator of the form

$$Z^*(x) = \sum_{i=1}^{n} b_i \, g(\|x_i - x\|) + \sum_{k=0}^{p} c_k \, f_k(x), \tag{1}$$

where the $f_k(x); k = 0, \ldots, p$ are linearly independent monomials in the position coordinates of x. In the case where the interpolator is exact, i.e., $Z^*(x_i) = Z(x_i)$ for $i = 1, \ldots, n$, Micchelli [1] has shown that conditional positive definiteness of an appropriate order of the function $g(x)$ is a sufficient condition for uniquely determining the weights in the interpolator. This interpolator as well as Micchelli's results will be generalized in several ways.

Alternatively, consider an interpolator cum estimator of the form

$$Z^*(x) = \sum_{i=1}^{n} \lambda_i(x) \, Z(x_i). \tag{1'}$$

These two interpolators represent two different approaches to the same problem and in certain special cases they are equivalent. This has been noted by Kimeldorf and Wahba [2], Matheron [3–5], Watson [6], Myers [7] and Cressie [8]. The form given by (1) can be derived, as in the case of the thin plate and smoothing spline, by imposing a smoothness condition on the interpolating function. Alternatively this form can be assumed as in Micchelli. It can also be obtained indirectly by assuming an interpolator as in (1′) satisfying certain conditions such as unbiasedness and minimal estimation variance then showing its equivalence with that in (1). The form given in (1′) of the interpolator is easily generalized. By exploiting the equivalence between the two forms, generalizations or extensions of one form can be extended to the other form. For the sake of clarity we digress to delineate and emphasize the use of the terms "multidimensional"

Although the research described in this article has been funded wholly or in part by the U.S. Environmental Protection Agency through a Cooperative Research Agreement with the University of Arizona, it has not been subjected to Agency review and therefore does not reflect the views of the Agency and no official endorsement should be inferred.

and "multivariate" since some authors have used these terms interchangeably. In the following, *multidimensional* will refer to the dimension of the *domain* of the interpolated/interpolating function whereas *multivariate* will refer to the dimension of the *range* of these functions. It would seem to be desirable for the definition as well as the derivation of an interpolator to independent of both of these dimensions, this is clearly not the case for either type of spline in their usual formulations.

2. POSITIVE DEFINITENESS

It was noted above that positive definiteness plays a central role with respect to the uniqueness of the coefficients in the interpolators (1) and (1'). In order to generalize the interpolators, it is necessary to consider more general forms of positive definiteness, although these could have been stated for the complex valued case only real valued functions will be considered herein.

DEFINITION 1. *Let $f_0, \ldots, f_p$ be linearly independent real valued functions defined on R^k, and $g(x, y)$ be a function from $R^k \times R^k$ into the ring of $m \times m$ real symmetric matrices. Then, g is said to be positive definite with respect to the $f_0, \ldots, f_p$ if for all sets of points $x_1, \ldots, x_n$ in R^k*

(a) $\operatorname{Tr} \sum_{i=1}^n \sum_{j=1}^n \Gamma_i^T g(x_i, x_j) \Gamma_j \geq 0$ *for all $m \times m$ real valued matrices $\Gamma_1, \ldots, \Gamma_n$ (not all identically zero) such that*

(b) $\sum_{j=1}^n f_i(x_j) \Gamma_j = 0$ *for $i = 0, \ldots, p$.*

LEMMA 1. *Let $f_0, \ldots, f_p$ and $g(x, y)$ be as in Definition 1 above. Then, g is positive definite with respect to the $f_0, \ldots, f_p$ if and only if for all sets of points $x_1, \ldots, x_n$ in R^k*

(a') $\sum_{i=1}^n \sum_{j=1}^n \Gamma_i^T g(x_i, x_j) \Gamma_j \geq 0$ *for all $m \times 1$ real valued vectors $\Gamma_1, \ldots, \Gamma_n$ (not all identically zero) such that*

(b') $\sum_{j=1}^n f_k(x_j) \Gamma_j = 0$ *for $k = 0, \ldots, p$.*

Clearly, if g is positive definite in the sense of Definition 1, then it suffices to consider weight matrices which have all zeros except in one column in which case the value of the trace is the same as the value of (a') when using those columns. Conversely, if (a') and (b') are satisfied, then the trace from Definition 1 can be written as a sum using the columns of the weight matrices as the vectors satisfying (a') and (b').

As will be seen later, it is more natural to define positive definiteness as in Definition 1, but it is also convenient to have the equivalent form given in Lemma 1. In conformity with usual definition, we say that g is *strictly* positive definite if the expression in (a) (respectively (a')) is positive for all Γ_j's satisfying (b) (respectively (b')) and not all are zero matrices. In the case where the set of f_k's is empty, we consider condition (b) to be vacuously satisfied, in which case the definition would coincide with the usual definition of positive definiteness albeit for matrix valued functions. It is also easy to see that Definition 1 is a matrix valued generalization of conditional positive definiteness as given in [1,9]. The importance of this definition is seen in the following theorem.

THEOREM 1. *Let $g(x, y)$ be strictly positive definite as in Definition 1 and $x_1, \ldots, x_n$ points in R^k, then the following matrix is invertible:*

$$\begin{bmatrix} g(x_1,x_1) & \cdots & g(x_1,x_n) & f_0(x_1)I & \cdots & f_p(x_1)I \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ g(x_n,x_1) & \cdots & g(x_n,x_n) & f_0(x_n)I & \cdots & f_p(x_n)I \\ f_0(x_1)I & \cdots & f_0(x_n)I & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_p(x_1)I & \cdots & f_p(x_n)I & 0 & \cdots & 0 \end{bmatrix}.$$

PROOF. The proof is completely analogous to the proof of the counterpart theorem given in [1]. For simplicity, consider the above matrix to be written in the form

$$\begin{bmatrix} G & F \\ F^T & 0 \end{bmatrix},$$

and suppose by way of contradiction that the matrix is not invertible. Then there is a vector $[U^T V^T]^T$ not identically zero such that

$$\begin{bmatrix} G & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

or $GU + FV = 0$ and $F^T U = 0$. The latter equation is the same as the set of conditions given in (b) of Definition 1, except where the matrices $\Gamma_1, \ldots, \Gamma_n$ are simply column vectors. Clearly, $F^T U = 0$ implies $U^T F = 0$ and hence, $U^T FV = 0$. This implies that $U^T GU = 0$ which contradicts the strict positive definiteness of G unless U is the zero vector. If U is the zero vector, then $GU = 0$ and hence, $FV = 0$, but the scalar matrix functions are linearly independent and hence, $FV = 0$ implies that $V = 0$. Since both U, V are zero vectors, the original matrix must be invertible.

LEMMA 2. *Let $g(x, y)$ be positive definite in the sense of Definition 1 and A an $m \times 1$ column vector (not the zero vector) then $h(x, y) = A^T g(x, y) A$ is real valued positive definite with respect to $f_0, \ldots, f_p$.*

3. LINEAR ESTIMATORS

It will be seen that initially it is easier to generalize the estimator given in (1′) than it is to generalize the interpolator in (1), although these are essentially equivalent. As a way of modeling or quantifying the uncertainty, associated with the lack of knowledge about the function to be interpolated, the data will be considered to be a finite (non-random) sample of a vector valued random function $\overline{Z}(x)$ with values in R^m defined in R^k such that $\overline{Z}(x) = \overline{Y}(x) + \overline{M}(x)$ where

(i) $E[\overline{Z}(x)] = \overline{M}(x) = [f_0(x), \ldots, f_p(x)]M$. $f_0, \ldots, f_p$ are known real valued linearly independent functions defined in R^k, M is an unknown $(p+1) \times m$ matrix of constants.

(ii) $\bar{\gamma}(h) = 0.5\, E[\overline{Y}(x+h) - \overline{Y}(x)]_T[\overline{Y}(x+h) - \overline{Y}(x)]$ exists and depends only on h. When $m > 1$ $\bar{\gamma}(h)$ is matrix valued, then the diagonal entries (variograms) quantify the spatial correlation of each component of $\overline{Z}(x)$ with itself, the off diagonal entries (cross-variograms) quantify the intercomponent spatial correlation. Condition (ii) could be replaced by the following

(ii′) $C(h) = E\{[\overline{Z}(x+h)]^T[\overline{Z}(x)]\}$ exists and depends only on h.

The use of (ii′) implies an assumption of second order stationarity but does not require that $C(h)$ be symmetric. In contrast, $\bar{\gamma}(h)$ is symmetric by construction but requires only a weaker form of stationarity. In the case that the components of $\overline{Z}(x)$ are second order stationary, the matrix variogram function $\bar{\gamma}(h)$ can be expressed in terms of the matrix covariance function, namely

$$\bar{\gamma}(h) = C(0) - C(h).$$

Although a matrix covariance function should be positive definite, it is the negative of a matrix variogram function that should be positive definite. This distinction has no real effect with respect to Theorems 1 and 2.

THEOREM 2. *Let $\overline{Z}(x)$ be a random function satisfying (i), (ii) above, then given data $\overline{Z}(x_1), \ldots, \overline{Z}(x_n)$, the unbiased, minimum variance linear estimator of $\overline{Z}(x)$, is given by*

$$\overline{Z}^*(x) = \sum_{i=1}^{n} \overline{Z}(x_i)\, \Gamma_i(x), \tag{1″}$$

where

$$\begin{aligned} \sum_{i=1}^{n} \bar{\gamma}(x_i - x_j)\, \Gamma_i(x) + \sum_{k=0}^{n} f_k(x_j)\, \mu_k &= \bar{\gamma}(x - x_j), \qquad j = 1, \ldots, n, \\ \sum_{i=1}^{n} f_k(x_i)\, \Gamma_i(x) &= f_k(x), \qquad k = 0, \ldots, p, \end{aligned} \tag{2}$$

and "variance" means the sum of the respective error variances when $m > 1$.

The case of $m = 1$ is given in [10–12], each with slight variations. The general case of $m > 1$, more commonly known as cokriging, was given in [13]. A limited form of cokriging associated with the "undersampled problem" is given in [14] and was subsequently shown in [15,16] to be a special case of the above general form. A program implementing cokriging and incorporating the adjustment for the undersampled case was given in [17]. The system of equations given in (2) requires that $\bar{\gamma}(h)$ be known. Of course in practice this is usually not the case and the entries in $\bar{\gamma}(h)$ must usually be estimated/modelled from the data. There is also an implied presumption that the random function and the $\bar{\gamma}(h)$ are unique (up to certain equivalence relations). This may not quite be the appropriate viewpoint, since any valid variogram will produce an unbiased, minimal variance linear estimator. However, the minimized estimation variance only pertains to the choice of the weights in the estimator relative to that particular variogram and does not serve as a discriminator between variograms, i.e., between estimators obtained with different valid variogram models. Moreover, once the variogram is known, i.e, modeled, the minimized estimation variance is really a characteristic of the data location pattern more than it is a true quantification of the error. Note also that no distributional assumptions are invoked although only in the case of multivariate Normality is the interpolator the conditional mean.

The system of equations (2) corresponds to a point estimator rather than to an estimator of a spatial average such as the average grade of a block but the extension to spatial averages is relatively easy in the kriging form of the estimator.

A. Radial Basis Functions

For the case of $m = 1$, Powell [18], Micchelli [1] and others have proposed the use of an interpolator of the form

$$Z^*(x) = \sum_{i=1}^{n} b_i \, g(||x_0 - x||) + \sum_{k=0}^{p} c_k \, f_k(x), \tag{1}$$

where the f_k's are linearly independent functions as in the kriging estimator above and g is a kernel function satisfying an appropriate positive definiteness condition. By requiring that $Z^*(x_i) = Z(x_i)$ for $i = 1, \ldots, n$, i.e., the interpolator is exact, a set of linear equations is obtained for the unknown coefficients. Hardy [19] utilized a simpler form of (1), namely without the second summation and chose g to be a biharmonic function since in that case the coefficient matrix was invertible without additional restrictions. Micchelli showed that in the case where the f_k's are monomials in the position coordinates and the interpolator is required to be exact then conditional positive definiteness of g is a sufficient condition for obtaining a unique choice of the coefficients in (3). As noted in [7], isotropy is not necessary and a more general form of positive definiteness may be used that is still sufficient for obtaining unique solutions for the coefficients in (3). In fact, by using the "dual" form of the cokriging estimator as given in [15,16], it is seen that the radial basis function interpolator can easily be extended to the case of $m > 1$. This extension is dependent on the equivalence of several formulations of positive definiteness for matrix valued functions as given in [16].

B. Splines

In the case of $k = 1$ (and of course $m = 1$), the simplest form of a spline, the thin plate spline, is obtained as follows. Given data $Z(x_1), \ldots, Z(x_n)$ which are the values of an unknown function at the points $x_1, \ldots, x_n$ then the spline on the interval [a,b] containing these points is a function $Z^*(x)$ having a continuous second derivative on [a,b] and satisfying two conditions

- $Z^*(x_i) = Z(x_i), \qquad i = 1, \ldots, n,$
- $\int \left\{ \dfrac{d^2 Z^*(x)}{dx^2} \right\}^2 dx$ is minimal.

The first condition is the exactness and the second condition is of course a smoothness characteristic. The spline can also be characterized in a more abstract way. Let H_1, H_2 be Hilbert spaces

of scalar valued functions defined on R^k and B a bounded linear operator from H_1 to H_2. For the interpolation/estimation of a linear functional of the function $Z(x)$, the spline is the function g in H_1 such that Bg has minimum norm in H_2 and such that given continous linear functionals $L_1, \ldots, L_n$ have prescribed values when applied to g. For example, B could be the second order differential operator and the linear functionals could be point evaluations, i.e., the prescribed values of the functionals are then the known values of $Z(x)$ at data locations. In this case, the usual thin plate spline is obtained although for $k > 1$ there is some ambiguity in the choice of the second order differential operator.

C. The (Near) Equivalence in the Case $m = 1$

To see the equivalence between the three formulations it is essentially sufficient to re-write (1) and (2) in the so-called dual form which is completely equivalent to the original

$$Z^*(x) = \sum_{i=1}^{n} b_i \gamma(x - x_i) + \sum_{k=0}^{p} a_k f_k(x), \tag{1}$$

where

$$\begin{aligned} \sum_{i=1}^{n} \gamma(x_i - x_j) b_i + \sum_{k=0}^{p} f_k(x_j) a_k = Z(x_j), \qquad & j = 1, \ldots, n, \\ \sum_{i=1}^{n} f_k(x_i) b_i = 0, \qquad & k = 0, \ldots, p. \end{aligned} \tag{2'}$$

This form is obtained by solving the system (2), writing the estimator (1′) in matrix form as

$$[Z(x_1), \ldots, Z(x_n), 0, \ldots, 0][\Gamma_1(x), \ldots, \Gamma_n(x), \mu_0, \ldots, \mu_p]^T,$$

substituting the solution into this form and taking the transpose. That is, the interpolator (1″) is of the form

$$\begin{aligned} [\overline{Z} \quad 0][\Gamma^T \quad \mu^T]^T &= [\overline{Z} \quad 0] \begin{bmatrix} G & F \\ F^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} G_0 \\ F_0 \end{bmatrix} \\ &= \left\{ [G_0^T \quad F_0^T]^T \begin{bmatrix} G & F \\ F^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} \overline{Z}^T \\ 0 \end{bmatrix} \right\}^T \\ &= \{[G_0^T \quad F_0^T]^T [B^T \quad A^T]\}^T, \end{aligned}$$

where

$$\begin{bmatrix} G & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} B \\ A \end{bmatrix} = \begin{bmatrix} Z^T \\ 0 \end{bmatrix}.$$

This link between interpolator formulations was called "near" equivalence because the thin plate spline corresponds to a specific generalized covariance, whereas the kriging estimator or the radial basis function interpolator only require the use of a kernel with appropriate positive definiteness properties. This allows adapting the kernel function to a particular data set. This adaptability is discussed in [8,20,21].

When the estimator is written in the form (1), it is easy to see that the effect of the second summation is largely to determine the behavior of the interpolator when it is used as an EXtrapolator, i.e., outside the convex hull of the sample locations. If the variogram is constant beyond a certain distance (called the range) it is easy to show that the sum of the b_i's is zero assuming that one of the f_k's is a constant function and hence the first summation is zero when all the distances $x - x_i$ exceed the range. More generally it can be shown that even if the variogram does not become constant the first summation in (1′) will go to zero asymptotically. In comparing the two forms of the equations (2) and (2′), we see that the first set of equations in (2) corresponds to the requirement that Cov $\{Z^*(x), Z(x_j)\}$ = Cov $\{Z(x), Z(x_j)\}$ for each sample location and the

corresponding equations in (2') are the conditions $Z^*(x_j) = Z(x_j)$. The second set of equations in (2) is sufficient to ensure unbiasedness whereas the equations in (2') together with the positive definiteness of the kernel function are sufficient to ensure that the coefficient matrix is invertible.

Conditionally positive definite functions provide the link between the three formulations, these functions are either induced by a particular scalar product in an appropriate Hilbert space or induce that product. In the formulation of the interpolator given by (1), positive definiteness of an appropriate form is required to ensure that the variance of the error of estimation is non-negative and hence has a non-negative minimum. In contrast, the positive definiteness in the reformulation (1') and (2') is imposed solely to ensure the existence and uniqueness of the solution of the system of equations. Matheron linked weaker forms of positive definiteness to the representation of the variance of certain linear combinations that behave as generalized increments and hence would filter out polynomials of a given order. The radial basis formulation, as is clearly shown in [1], defines weaker forms of positive definiteness in terms of the non-negativity of certain quadratic forms which are related to the existence and uniqueness of solutions of linear systems. By bringing these together, a slightly more general form of positive definiteness can be defined and such kernel functions are still appropriate for representing the interpolator. Matheron [9] gives a generalization of the Bochner Theorem to characterize conditionally positive definite functions of a given order. There is a subset that are polynomials, Matheron then obtains the conditions on the coefficients in these polynomials. Alternatively, one may use the complete monotonicity characterization given in [1]. Because of the connection between the weak form of positive definiteness and the variances of generalized increments, Matheron refers to these kernels as generalized covariances. Conditionally positive definite functions appear in a number of contexts and are closely related to (strongly) positive definite functions. For example, conditionally positive definite functions are related to the exchangeability of random variables and as shown in [22] that they are essentially the logarithms of positive definite functions.

The "weight" matrices $\Gamma_1, \ldots, \Gamma_n$ in Definition 1 are taken to be $m \times m$ because that is the natural interpretation when deriving the cokriging estimator, i.e., one minimizes the sum of the estimation variances for the components of $\overline{Z}(x)$ and this sum arises as the trace of a quadratic form like (a).

4. THE GENERALIZATION

Consider now a vector valued estimator/interpolator of the form

$$\overline{Z}^*(x) = \sum_{i=1}^{n} B_i\, g(x_i, x) + \sum_{k=0}^{p} A_k\, F_k(x), \tag{1'''}$$

where $B_1, \ldots, B_n$ and $A_0, \ldots, A_p$ are $m \times m$ matrices, $g(x, y)$ is an appropriate kernel and the $F_k(x)$'s are of the form $f_j(x)I$. Given data $\overline{Z}(x_1), \ldots, \overline{Z}(x_n)$ and the requirement that the interpolator is exact, i.e., $\overline{Z}^*(x_i) = \overline{Z}(x_i)$ for $i = 1, \ldots, n$, a linear system of matrix equations is obtained. Unless the kernel function is positive definite in the strong sense, then the coefficient matrix for this system may not be invertible. However, by imposing additional conditions on the $B_1, \ldots, B_n$, a unique solution is obtained. As is the case of $m = 1$, the term, $\sum_{k=0}^{p} A_k\, F_k(x)$, is seen to determine the behavior of the estimator outside of the convex hull of the sample/data locations. The form of the estimator given by (1''') is also seen to be a slight generalization of the dual form of the cokriging estimator as given in [15].

THEOREM 3. *Let $g(x, y)$ be positive definite with respect to $f_0, \ldots, f_p$ be as in the definition preceding; $B_1, \ldots, B_n$ and $A_0, \ldots, A_p$ as in (1'''), then the following system has a unique solution*

and (1‴) will be an exact interpolator

$$\begin{bmatrix} g(x_1,x_1) & \cdots & g(x_1,x_n) & F_0(x_1) & \cdots & F_p(x_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ g(x_n,x_1) & \cdots & g(x_n,x_n) & F_0(x_n) & \cdots & F_p(x_n) \\ F_0(x_1) & \cdots & F_0(x_n) & & & \\ \vdots & \ddots & \vdots & & 0 & \\ F_p(x_1) & \cdots & F_p(x_n) & & & \end{bmatrix} \begin{bmatrix} B_1^T \\ \vdots \\ B_n^T \\ A_0^T \\ \vdots \\ A_p^T \end{bmatrix} = \begin{bmatrix} Z(x_1)^T \\ \vdots \\ Z(x_n)^T \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

The uniqueness of the solution of this system of equations is a direct consequence of Theorem 1 and the first n (matrix) equations are simply the exactness conditions.

If as was assumed above, the identity matrix is one of the linearly independent matrix functions then the sum of the B's is the zero matrix (i.e., this is one of the equations in (b)); hence, if $g(x,y)$ is a constant matrix for the distances between x and y sufficiently large, then $\overline{Z}^*(x)$ is determined only by the second summation in (1‴) when the minimum of the distances $x - x_i$ is large enough. The upper portion of the system in Theorem 2 is obtained by imposing the exactness condition and the lower part corresponds to the unbiasedness conditions if formulated in the context of cokriging. However, the estimator given by (1‴) can be obtained without the stochastic formulation and provides a natural generalization of a spline. From the geostatistical/cokriging perspective g is assumed to be "uniquely" determined by the data and the principal problem is one of adequately estimating/modeling the kernel function as well as determining the appropriate order of the polynomial functions whereas from the perspective of splines or radial basis functions the choice of the linearly independent functions and the kernel function is more arbitrary and determined by external conditions such as the smoothness of the interpolating surface(s).

5. APPLICATIONS

Even in the case of $m > 1$, there are many examples of the use of linear interpolators in the earth sciences beginning with the problem of ore reserve estimation in mining. For the case of $m = 1$ the kriging estimator is seen to be a form of a generalization of well-known techniques such as nearest neighbor, inverse distance weighting which even yet are used by some practitioners. The case of $m > 1$ arises rather naturally in mining as well as in hydrology, soil physics and environmental assessment. In most metal mining operations there is a primary metal and one or more secondary metals, that is, the mine is established to extract a primary metal but the extraction process also produces by-products. For example, in the case of copper, secondary metals may include molybdenum, zinc, gold, silver. The "value" of a block of ore then is a function of the respective grades and the market prices of all the metals or minerals extracted, the value is a linear combination of the grades and in some cases such an "equivalent" grade is used in lieu of actual grades. However, these grades are not only spatially correlated but they are also intercorrelated. As shown in [23] estimation of linear combinations is suboptimal in general as compared with joint estimation. Both in the case of mining and in other applications one variable may serve as a proxy for another, one being easier or cheaper to sample. Most of the earliest examples of cokriging were for this "undersampled" problem where in the second variable was only of interest to enhance the estimation of the primary variable. As was shown in [24] the system of equations corresponding to the undersampled case is a special case of the full-sampled case and as shown in [16] the set of equations is the same whether all variables are estimated or only a primary variable, the positive definiteness condition is the same in both cases. Since kriging and cokriging are smoothing operations, they are potentially applicable to image analysis problems as discussed in [25–27].

6. SIMULATION

In the geostatistical formulation of the problem and the interpolator, the data is viewed as a non-random sample from one realization of a random function for which some form of second moment function is known. In this context, it is natural to attempt to generate additional realizations as well as to interpolate across the one realization. Simulation is a useful tool for many purposes such as mine planning, evaluation of potential waste repositories, etc. While it would in general not be possible generate an entire realization, generation of the values of a realization at grid points is quite reasonable. Simulation could mean different things but in geostatistics is usually interpreted to mean that the first and second moments are preserved as is the marginal distribution of $Z(x)$. Essentially all interpolation methods result in some degree of smoothing. For some purposes however, it may be desirable to preserve or enhance the variability, hence the need for simulation. It may still be desirable to generate a realization that conforms to the given data, i.e., to condition the simulation to the given data. Let $Z^*(x)$ denote the interpolated value at the point x and let $Zs(x)$ denote the simulated value. If the interpolator is exact and we write $Zs(x) = Z^*(x) + \{Z(x) - Z^*(x)\}$, then it would be sufficient to simulate the mean zero differences $\{Z(x) - Z^*(x)\}$.

At least two different simulation methods have been used for simulating $Z(x)$, x in R^K. One method reduces the problem of simulation in higher dimensional space to the problem of $k = 1$. Several methods are well-known in the times series context. For example, the Box-Jenkins moving average can be used. Matheron [9] showed that simulations in higher dimensional space could be generated by linear combinations of independent simulations in 1-space. Variations on the method utilize a spectral density function. Fast Fourier transforms can also be used. Alternatively, $Z(x)$ can be simulated at a finite number of points by the use of an $L - U$ decomposition of the covariance matrix as shown by Davis [28,29]. This method has the advantage that it does not depend on k but it requires manipulation of very large matrices. Carr and Myers [30] extended the Turning Bands method to the case where $Z(x)$ is vector valued and Myers [31] showed that the covariance matrix decomposition method extends to the vector valued case as well.

As seen above, the link between the interpolators (1) and (1′) is in the kernel function which in the case of (1′) is interpreted as a generalized covariance. Since this same function appears in (1), there should be a corresponding form of simulation for the radial basis function representation. Certainly one way to do it is to utilize the correspondence, that is, transform the representation from the radial basis function form to the linear combination of data form, simulate and then re-transform. The difference in the approaches is fundamental and easily shows why simulation is not as naturally related to (1). If the number of data points is finite and the number of points to be interpolated is finite then there are two choices for interpreting the random function $Z(x)$. First, it can be considered to be analogous to a random variable but one whose values are functions rather than numbers. Second, for the finite collection of points of interest, one can consider only the collection of jointly distributed random variables. The latter is sufficient for the interpolator (1′) but for (1). One possible approach is to consider the coefficients in (1) as jointly distributed random variables and hence one could simulate $Z(x)$ by simulating those random variables. However, examination of the correspondence between (1) and (1′) shows that the coefficients in (1) depend on the kernel function. It would not be sufficient to jointly simulate the coefficients in order to preserve properties of $Z(x)$. In the case of the smoothing spline, there is a natural form of simulation for (1), namely to add a random noise term to each interpolated value. If however the noise terms were assumed correlated, then the most natural form is that given by (1′).

7. VARIOGRAM MODELING

It is perhaps more evident in the alternative form (1′) than in the original form (1) that the variogram must be estimated/modeled as a function and not just at a few points. Because the variogram must satisfy an appropriate positive definiteness property, the practical approach is to begin with known valid models (with perhaps unknown parameters) and form positive linear combinations; these are known as nested models. In practice, only a few model types are used although others can be generated by averaging with respect to a parameter. There is extensive

geostatistical literature on the problem of estimating and modeling variograms; for a review see [32]. Briefly, the data is used to generate a sample variogram using distance classes and angle windows, the plot of the sample variogram is compared with the plots of known valid models. The parameters to be determined correspond to characteristics of the plots. The problem is more difficult in the case of cross-variograms since a function is not a valid cross variogram except in the context of an associated pair of variograms. While it is possible to compute and plot a sample cross-variogram it is not as simple to know what the plot of a cross-variogram must look like. As shown in [15], this problem can be resolved by the introduction of new "variables," the sum and the difference for each pair. By modeling the variograms of these the cross variogram can be determined and ensure the positive definiteness property.

As a practical matter, the biggest problem in modeling cross variograms is insufficient data. In order to form the sum and difference of two variables, it is necessary to have data for both variables at the same locations. In one of the common applications, the undersampled problem, one variable is significantly undersampled compared to the other. If only data locations are used where both are sampled, there may be too few data points to provide for adequate estimation and modeling of the cross variogram(s). Clark, Basinger and Harper [33] proposed a pseudo-cross variogram as follows

$$g_{ij}(h) = 0.5\ E\{[Z_i(x+h) - Z_j(x)]^2\}.$$

They also showed that it was possible to derive the equations for a linear estimator using this pseudo-cross variogram although they only considered the case of the estimation of one component. Second order stationarity is sufficient for this function to exist and be dependent only on h but weaker conditions are also adequate. This pseudo-cross variogram is not a variogram nor is it in general a true cross variogram. For example, it is non-negative valued whereas a cross-variogram could be negative, it's value at zero need not be zero whereas a cross-variogram can have a jump discontinuity but is zero at zero. A cross variogram is symmetric with respect to $h = 0$, but the pseudo-cross variogram need not be. Myers [34] has shown the relationship to the usual cross variograms and ways to model them as well as characterizing the positive definiteness property. Only slight modifications in software are necessary in order to use pseudo-cross variograms in the set of cokriging equations.

8. OPEN PROBLEMS

As is pointed out above, any kernel function, i.e., radial basis function or generalized covariance, that satisfies the appropriate positive definiteness condition will result in an exact interpolator. It is known of course that only one of these corresponds to the smoothness condition on the interpolating function. Since smoothness may not always be the relevant condition to impose in all cases, it would be useful to establish a link between the choice of this kernel function and the properties of the interpolating function. In turn, there should be a link between the choice of a bounded linear operator between appropriate Hilbert spaces and the choice of this kernel function.

It is easy to show empirically that in some cases it makes little difference in the interpolated values when the parameters in the kernel function are changed. In particular, this is true for certain sample location patterns (points in space where the values of the function are known). This suggests the defining a topology on the space of valid kernel functions in such a way that the map to the interpolating function is continuous. As pointed out in [34], there are at least three natural ways to define neighborhoods for the kernel functions but no one choice seems best.

In the development of the cokriging estimator, it is easy to use intervariable correlation as the characterization of the interdependence of the components of the vector valued random function. It seems less obvious how to invoke a form of interdependence of the components of vector valued function when using either the spline or radial basis function approach although there is one implied via the duality or equivalence between the different forms of the interpolator.

REFERENCES

1. C. Micchelli, Interpolation of scattered data: Distance matrices and conditionally positive definite functions, *Constructive Approximation* **2**, 11–22 (1986).
2. G.S. Kimeldorf and G. Wahba, A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, *Annals Math. Stat.* **41** (2), 495–502 (1970).
3. G. Matheron, Splines et krigeage: Leur equivalence formelle, Internal Report N-667, 26 p., Centre de Geostatistique, Fontainebleau, (1980).
4. G. Matheron, Remarques sur le krigeage et son dual, Internal Report N-695, 36 p., Centre de Geostatistique, Fontainebleau, (1981).
5. G. Matheron, Splines et krigeage: Le cas fini, Internal Report N-698, 23 p., Centre de Geostatistique, Fontainebleau, (1981).
6. G.S. Watson, Smoothing and interpolation and splines, *Math. Geology* **16** (8), 601–615 (1984).
7. D.E. Myers, Interpolation with positive definite functions, *Sciences de la Terre* **28**, 252–265 (1988).
8. N. Cressie, Geostatistics, *The American Statistician* **43** (4), 197–202 (1989).
9. G. Matheron, The intrinsic random functions and their applications, *Adv. Applied Prob.* **5**, 439–468 (1973).
10. G. Matheron, *Les Variables Regionalisées et Leur Estimation*, Masson et Cie, Paris, (1965).
11. B. Matern, Spatial variation, In *Lecture Notes in Statistics*, 2nd *ed.*, Springer Verlag, London, (1986).
12. A.S. Goldberger, Best linear unbiased prediction in the generalised linear regression model, *J. Am. Stat. Assn.* **57** (298), 369–375 (1962).
13. D.E. Myers, Cokriging: The matrix form, *Math. Geology*, **14** (3), 249–257 (1982).
14. G. Matheron, *The Theory of Regionalized Variables and its Application*, pp. 211, Les Cahiers du Centre de Geostatisque, Fountainebleau, France, (1971).
15. D.E. Myers, Some new aspects of multivariate analysis, In *Statistical Treatment for Estimation of Mineral and Energy Resources*, (Edited by C.-J. Chung), pp. 669–687, D. Reidel Pub., Dordrecht, (1988).
16. D.E. Myers, Multivariate geostatistics for environmental monitoring, *Sciences de la Terre* **27**, 411–428 (1988).
17. J. Carr, D.E. Myers and C. Glass, Co-kriging: A computer program, *Computers and Geosciences* **11** (2), 111–127 (1985).
18. M.J.D. Powell, Radial basis functions for multivariable interpolation: A Review, In *Algorithms for the Approximation of Functions and Data*, (Edited by M.G. Cox and J.C. Mason), Oxford Univ. Press, (1985).
19. R.H. Hardy, Multiquadratic equations of topography and other irregular surfaces, *J. Geophysical Research* **76**, 1905–1915 (1971).
20. N. Cressie, Reply to Letter from G. Wahba, *The American Statistician* **44** (3), 256–258 (1990).
21. G. Wahba, Letter to the Editor (N. Cressie), *The American Statistician* **44** (3), 255–256 (1990).
22. S. Johansen, An application of extreme point methods to the representation of infinitely divisible distributions, *Z. Wahrscheinlichkeitstheorie verw. Beb.* **5**, 304–316 (1960).
23. D.E. Myers, Estimation of linear combinations and cokriging, *Math. Geology* **15** (5), 633–637 (1983).
24. D.E. Myers, Cokriging: New developments, In *Geostatistics for Characterization of Natural Resources*, (Edited by G. Verly *et al.)*, pp. 295–305, D. Reidel Pub., Dordrecht, The Netherlands, (1984).
25. J. Carr and D.E. Myers, Application of the theory of regionalized variables to the spatial analysis of LANDSAT data, In *The Proceedings of the Ninth Pecora Symposium on Remote Sensing, 2-4 Oct.*, IEEE Computer Society Press, Silver Springs, MD, (1984).
26. J. Carr, C. Glass, H. Yang and D.E. Myers, Application of spatial statistics to analysing multiple remote sensing data sets, In *Geotechnical Applications of Remote Sensing and Remote Data Transmission*, ASTM STP 967, (Edited by A.I. Johnson and C.B. Peterson), American Society Testing and Materials, Philadelphia, PA, pp. 138–150, (1987).
27. D.E. Myers, Multivariate, multidimensional smoothing, *Spatial Statistics and Imaging, Lecture Notes—Monograph Series*, (Edited by A. Possolo), pp. 275–285, Institute of Math. Statistics, Hayward, CA, (1991).
28. M. Davis, Production of conditional simulations via the LU triangular decomposition of the covariance matrix, *Math. Geology* **19** (2), 91–98 (1987).
29. M. Davis, Generating large stochastic simulations—The matrix polynomial approximation method, *Math. Geology* **19** (2), 99–107 (1987).
30. J. Carr and D.E. Myers, COSIM: A FORTRAN IV program for co-conditional simulation, *Computers and Geosciences* **11** (6), 675–705 (1985).
31. D.E. Myers, Vector conditional simulation, In *Geostatistics, Proceedings of the Third International Geostatistical Congress, Avignon, 5-9 Sept. 1988*, (Edited by M. Armstrong), pp. 283–293, Kluwer Academic Publishers, Dordrecht, The Netherlands, (1988).
32. D E. Myers, On variogram estimation, In *The Frontiers of Statistical Scientific Theory and Industrial Applications*, (Vol. II, Proceedings of the ICOSCO-I, Cesme, Izmir, Turkey, April 1987), pp. 261–281, American Sciences Press, Syracuse, NY, (1991).
33. I. Clark, K.L. Basinger and W.V. Harper, Muck-a-novel approach to co-kriging, In *Proceedings of the Conference on Geostatistical Sensitivity and Uncertainity Methods for Ground-Water Flow and Radionuclide Transport Modeling*, (Edited by B.E. Buxton), pp. 473–494, Batelle Press, Columbus, OH, (1989).
34. D.E. Myers, Pseudocross-variograms, positive definiteness and cokriging, *Math. Geology* **23** (6), 805–816 (1991).