Lecture presented at the International School of Geodesy "A. Marussi",
4th Course: "*Applied and Basic Geodesy: Present and Future Trends*"
Ettore Majorana Centre for Scientific Culture, Erice-Sicily, 15-25 June 1987.

From: ***Eratosthenes***, 22, (229-262), 1988.

Geodetic Applications of Interpolation and Prediction

Athanasios Dermanis

Department of Geodesy and Surveying
The Aristotle University of Thessaloniki

1. Introduction

Interpolation in its classical sense familiar from mathematical textbooks, is a method for the "construction" of a function from its discrete values at a finite number of points (elements of its domain of definition). Here interpolation is meant in a somewhat more generalized sense, where the available data are not restricted to be values of the function itself, but they can be any type of real valued quantities which depend on the relevant function (see e.g. Davis, 1975, p. 26).

Geodetic applications related to data analysis aim at the determination of an "unknown" entity from information which becomes available from relevant observations. Since both the unknowns and the data can be of either a discrete or continuous type, geodetic applications can be classified accordingly. With this in mind, interpolation can be identified as the class of geodetic applications where an unknown of continuous type, i.e. one or more functions, is to be determined from discrete data. One can also consider more complicated situations where unknowns and/or data contain both discrete and continuous elements. Interpolation methods can be extended to the case where the unknowns are a mixture of continuous and discrete elements, the data being always of a discrete type only. Continuous data methods seem to be of no practical importance at first sight, since observations can be carried out at discrete points only. Even if a continuous record is obtained by an analog instrument, it has to be discretized before any further computational processing takes place. However, theoretical solutions treating problems with continuous data become of practical importance in two ways. The most obvious way is to obtain first a continuous data record from the interpolation of discrete data. An other possibility is to seek numerical approximations to the rigorous solution which is based on continuous data, in such a way that the approximation utilizes only discrete available information.

The most important case of interpolation is related to the choice of an interpolating function among an infinite number of candidates. Such a choice can only be based on an appropriate optimality criterion which is explicitly or implicitly related to the degree of smoothness of the interpolation function. The situation is similar to that in classical adjustment theory which treats the problem with discrete data and discrete unknowns. The number of unknowns, which include the problem parameters and the unknown observational errors, exceeds the number of the available observations and the relevant equations admit an infinity of solutions. In particular, when the method of condition equations is used the observational errors are the only unknowns. The choice of a single solution can be initially based on a deterministic (weighted) least squares principle which is an optimality criterion reflecting the fact that errors are small in magnitude. However the problem of the proper choice of the weights can not be solved from deterministic considerations alone. Taking into account the stochastic behavior of the observational errors the well known optimality criterion of best (minimum variance) linear unbiased estimation (BLUE) provides an answer to the choice of weights problem: The optimal weight matrix is the inverse of the covariance matrix of the observations.

Returning to the interpolation problem we find a quite similar situation. The smoothness of the interpolating function is expressed by a scalar quantity, the norm, in the same way that the scalar sum of weighted residuals expresses the smallness of the observational errors. The problem of the choice of an appropriate norm can be solved with the introduction of a stochastic model. The unknown function is considered to be a sample function from a corresponding random function (stochastic process) in the same way that the actual errors in the classical adjustment problem are considered to be samples of relevant random variables. The corresponding stochastic optimality criterion is that of best (minimum mean square error) linear prediction. A link is established between the stochastic prediction problem and the deterministic interpolation problem through the identification of the covariance function of the stochastic process with a similar function, related to the norm of the interpolating function, which is called the reproducing kernel. Again, this link is similar to the link between the stochastic estimation problem and the deterministic least squares

problem in classical adjustment theory, which is established through the identification of the error covariance matrix with the inverse of the weight matrix.

We shall try here to explain this connection between interpolation and prediction with particular consideration to the relevant geodetic applications. Since our aim is to emphasize the basic ideas underlying this connection, we shall avoid much of the rigor which is nevertheless necessary for a proper mathematical treatment.

2. The interpolation problem

In the classical interpolation problem (which we shall call "simple interpolation" from now on) the available data are the discrete values

$$s_i = f(P_i), \qquad i=1,2,\ldots,n, \tag{1}$$

of an unknown function f at n known points P_i. We use the term point for the sake of simplicity. In fact P_i can be any element in the domain of definition of the function f, e.g. an epoch in time, an event (place and epoch) in space-time, etc. The problem is to construct an interpolating function f from the available values s_1, s_2, …, s_n and the coordinates of the points P_1, P_2, …, P_n. In mathematical terms we seek an operator

$$J: R^n \times R^{(nd)} \to F: \big((s_1,\ldots,s_n),(\mathbf{r}_1,\ldots,\mathbf{r}_n)\big) \to f \tag{2}$$

which we could call *interpolating operator* or *interpolator* for short. In the above relation d stands for the dimension of the domain of definition of f and $\mathbf{r}_i$ for the coordinate vector of point P_i with respect to a fixed system of coordinates. F is a set of functions where f is assumed to belong. The solution to the interpolation problem can also be expressed in a point-wise way, in which case, for every point P in the domain of definition of f, a mapping

$$J_P: R^n \times R^{(nd)} \times R \to R: \big((s_1,\ldots,s_n),(\mathbf{r}_1,\ldots,\mathbf{r}_n),\mathbf{r}\big) \to f(P) \tag{3}$$

where $\mathbf{r}$ is the coordinate vector of point P.

Interpolation in a generalized sense is not bound to the case where the available data are the values of the unknown function at discrete points. On the contrary, the discrete value s_i can be any real valued quantity related to the unknown function f by means of

$$s_i = L_i(f) \tag{4}$$

where

$$L_i: F \to R: f \to s_i \tag{5}$$

is a mapping from the function space F, where f belongs, into the reals. Such a real valued mapping is called *functional*. We will consider here only linear functionals, i.e. functionals having the linearity property ($\alpha_1, \alpha_2 \in R$)

$$L(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1 L(f_1) + \alpha_2 L(f_2). \tag{6}$$

Nonlinear functionals which occur in practice are eventually replaced by linear functionals with the help of a linearization process where an original unknown function f^a is replaced by a "small" function $f = f^a - f^0$ with the help of a known approximation (reference function) f^0.

The functionals in the simple interpolation problem are the evaluation functionals

$$E_P(f) = f(P) \tag{7}$$

which assign to each function its value at a specific point. Typical in geodetic applications are functionals which result from the combination of evaluation functionals with (usually linear) operators which map the original function f into some other function. If A is such an operator we obtain a functional $L = E_P \circ A$, defined by

$$L(f) = (E_P \circ A)(f) = E_P(Af) = (Af)(P). \tag{8}$$

It is exactly the presence of the evaluation functional in the above relation which justifies the use of the term *discrete* for the data $s = L(f)$. However it must be understood that the known value s does not depend on the value of the function f at point P only. When A is a differential operator, s depends on the values of f in an infinitesimal neighborhood of the point P. When A is an integral operator, s depends on the values of f on the whole domain of integration.

Seeking the interpolation function f among "all possible functions" is quite a hopeless task. We must therefore confine ourselves to an appropriate set of functions F and try to find an element $f \in F$, which is the best choice according to an optimality criterion that remains to be defined. The elements of F which agree with the available data

$$s_i = L_i(f), \qquad i=1,2,\ldots,n, \tag{9}$$

constitute a subset F_0 of F. When F_0 is the empty set, i.e. when no element of F satisfies the data equations (9), it is not possible to have an interpolation in the real sense, that is an exact interpolation. In such a case only a smoothing interpolation is possible, where the interpolating function does not agree with the available data s_1, s_2, …, s_n. According to the number of elements in F_0 we distinguish between three types of interpolation:

a) Least squares smoothing interpolation

When F_0 is empty, i.e., it has no element, we have in general

$$s_i - L_i(f) = v_i \neq 0, \qquad i=1,2,\ldots,n, \tag{10}$$

and a function $f \in F$ can be chosen which fits the data best, according to the least squares optimality criterion

$$\sum_i v_i^2 = \mathbf{v}^T \mathbf{v} = \min . \quad (11)$$

A somewhat more general optimality criterion is the weighted least squares criterion

$$\mathbf{v}^T \mathbf{P} \mathbf{v} = \min \quad (12)$$

where $\mathbf{P}$ is a positive definite weight matrix.

b) Interpolation in the classical sense

When F_0 has only one element, this particular element is automatically the interpolation function. This is the case of simple interpolation studied in mathematical texts, in particular those of numerical analysis. The whole effort is directed towards the choice of F in such a way that only one of its elements satisfies the data. Furthermore the functionals of equation (9) are restricted to be evaluation functionals. A typical example of this type of interpolation is the determination of a polynomial of degree n from its values at $n+1$ points. This type of interpolation has little or no interest for geodetic applications.

c) Minimum norm interpolation

When F_0 has more than one element, (in fact an infinite number of elements,) an optimality criterion must be introduced for the selection of the "best" function among all those satisfying the data equations (9) (exact interpolation). Such an appropriate optimality criterion is the minimization of the norm $\| f \|$ of the function f

$$\| f \| = \min . \quad (13)$$

The norm of a function is a positive scalar quantity, having certain properties (see e.g., Davis, 1975, p. 129), which plays a role similar to that of length for vectors and can be related to the smoothness of the function. Minimum norm interpolation is the most important type of interpolation for geodetic applications, its main characteristic being that the interpolating function is chosen from a rich collection of candidates (the set F) with a criterion related to smoothness. The problem is now to select a function space F equipped with a norm in such a way that the minimization principle of equation (13) can be applied in a computationally tractable way.

c) Hybrid interpolation

A final interpolation method, which can be used irrespectively of whether F_0 has none, one or more elements, is the hybrid interpolation where a tradeoff is made between faithfulness to the available data ($\mathbf{v}^T \mathbf{P} \mathbf{v}$) and smoothness ($\| f \|^2$) with the use of the optimality criterion

$$\mathbf{v}^T \mathbf{P} \mathbf{v} + \| f \|^2 = \min . \quad (14)$$

3. Interpolation with base functions

The simplest way for the construction of the function space F is to consider it as the set of all linear combinations

$$f = a_1 f_1 + a_2 f_2 + \ldots + a_m f_m , \quad (15)$$

with real coefficients a_i, of m known functions f_i, which are called *base functions*. With this definition F becomes a linear space, i.e. a set of elements satisfying certain axioms (see e.g. Davis, 1973, p. 3) the most important one being the linearity property: any linear combination of elements from F is also an element of F.

The choice of the base functions, apart from some obvious characteristics like having the same domain of definition as the unknown function f, is based on considerations related to the physical situation in which the function f arises. If for example f is known to be a solution of some differential equation, the base functions should be solutions of the same equation. The base functions are typically ordered with decreasing degree of smoothness. Examples of base functions are polynomials in one or more independent variables, trigonometric functions, spherical harmonics when f is the harmonic disturbing potential of the gravity field of the earth, etc.

It is further assumed that the base functions are linearly independent, which means that

$$a_1 f_1 + a_2 f_2 + \ldots + a_m f_m = 0 \;\Leftrightarrow\; a_1 = a_2 = \ldots = a_m = 0 . \quad (16)$$

Every function f is in an one-to-one correspondence with the coefficient vector

$$\mathbf{a} = [a_1 \; a_2 \cdots a_m]^T \quad (17)$$

in the expansion of equation (15). Thus the problem of determining the interpolating function f reduces to the one of finding the corresponding coefficient vector $\mathbf{a}$. We introduce also the vector notations

$$\mathbf{s} = [s_1 \; s_2 \cdots s_n]^T = [f(P_1) \; f(P_2) \cdots f(P_n)]^T , \quad (18)$$

$$\mathbf{f} = [f_1 \; f_2 \cdots f_m]^T \quad (19)$$

$$\mathbf{f}(P) = [f_1(P) \; f_2(P) \cdots f_m(P)]^T . \quad (20)$$

When $m<n$ it is not possible to have an exact interpolation. For the determination of the least squares smoothing interpolation we note that for linear functionals

$$L_i(f) = L_i(a_1 f_1 + a_2 f_2 + \ldots + a_m f_m) = \\ = a_1 L_i(f_1) + a_2 L_i(f_2) + \ldots + a_m L_i(f_m) \quad (21)$$

Since $f = \mathbf{a}^T \mathbf{f}$, equation (21) can also be written in the matrix form

$$L_i(f) = L_i(\mathbf{a}^T \mathbf{f}) = \mathbf{a}^T L_i(\mathbf{f}) . \quad (22)$$

Equation (10) has now the form

$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} L_1(f_1) & L_1(f_2) & \cdots & L_1(f_m) \\ L_2(f_1) & L_2(f_2) & \cdots & L_2(f_m) \\ \vdots & \vdots & \ddots & \vdots \\ L_n(f_1) & L_n(f_2) & \cdots & L_n(f_m) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \qquad (23)$$

or in combination with equation (12)

$$\mathbf{s}=\mathbf{F}\mathbf{a}+\mathbf{v}\,, \qquad \mathbf{v}^T\mathbf{P}\mathbf{v}=\min\,, \qquad (24)$$

where $\mathbf{F}$ is a matrix with $L_i(\mathbf{f})^T$ as its i$^{\text{th}}$ row and elements

$$F_{ij}=L_i(f_j)\,. \qquad (25)$$

In the simple interpolation case where $L_i(f)=E_{P_i}(f)=$ $=f(P_i)$ the matrix $\mathbf{F}$ has elements $F_{ij}=f_j(P_i)$.

Equations (24) resemble the case of adjustment with observation equations and the well known solution to the relevant minimization problem is

$$\mathbf{a}=(\mathbf{F}^T\mathbf{P}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{P}\mathbf{s}\,. \qquad (26)$$

4. Minimum norm Interpolation

When the number of base functions exceeds the number of available data ($m>n$), (exact) minimum norm interpolation is possible provided that a norm has been defined in the space F of all linear combinations of the base functions. For reasons of computational tractability this norm should result from the definition of an inner product in the same function space. An inner product is a scalar quantity (f,g) defined for any two elements f and g of F, having certain properties (see e.g. Davis, 1973, p. 158) and playing a role similar to that of scalar product for vectors. With the definition of an inner product the linear space F becomes an inner product space. Examples of inner products are

$$(f,g)=\int fg\,dP\,, \qquad (27)$$

where dP is the volume element in the domain of definition of the relevant functions, and

$$(f,g)=\int(fg+\nabla f\bullet\nabla g)\,dP\,, \qquad (28)$$

where ∇f, ∇g are the gradient vectors of the functions f and g, respectively.

If f has the expansion given by equation (15) and g by the corresponding expansion

$$g=b_1f_1+b_2f_2+\ldots+b_mf_m=\mathbf{b}^T\mathbf{f}\,, \qquad (29)$$

we have, due to the fact that the inner product is a bilinear function,

$$(f,g)=\left(\sum_i a_i f_i,\sum_j b_j f_j\right)=\sum_i\sum_j a_ib_j(f_i,f_j)=$$

$$=\sum_{ij}R_{ij}a_ib_j=\mathbf{a}^T\mathbf{R}\mathbf{b}\,, \qquad (30)$$

where we have set

$$R_{ij}=(f_i,f_j)\,. \qquad (31)$$

If we consider the linear space $E^n_{\mathbf{R}}$ of *n*-vectors with inner product

$$(\mathbf{a},\mathbf{b})=\mathbf{a}^T\mathbf{R}\mathbf{b}\,, \qquad (32)$$

there exists an *isomorphism* between the spaces F and $E^n_{\mathbf{R}}$, i.e. an one-to-one correspondence

$$f\leftrightarrow\mathbf{a}\,, \qquad g\leftrightarrow\mathbf{b}\,. \qquad (33)$$

Since this isomorphism preserves the inner product

$$(f,g)=\mathbf{a}^T\mathbf{R}\mathbf{b}=(\mathbf{a},\mathbf{b}) \qquad (34)$$

it is an *isometric isomorphism* or *congruence*, as it is usually called.

When an inner product is defined for F, a corresponding inner product is defined in $E^n_{\mathbf{R}}$. From these inner products corresponding norms are defined by means of

$$\|f\|^2=\sqrt{(f,f)} \qquad (35)$$

and

$$\|\mathbf{a}\|^2=\sqrt{\mathbf{a}^T\mathbf{R}\mathbf{a}}\,. \qquad (36)$$

Due to the above congruence, the norm is also preserved

$$f=\mathbf{f}^T\mathbf{a} \quad\Rightarrow\quad \|f\|=\|\mathbf{a}\|\,. \qquad (37)$$

The problem of minimum norm interpolation becomes now a problem defined by

$$\mathbf{F}\mathbf{a}=\mathbf{s}\,, \qquad \mathbf{a}^T\mathbf{R}\mathbf{a}=\min\,. \qquad (38)$$

Here we have written the data equations (9) in matrix form, taking equations (21) and (25) into account, and we have replaced $\|f\|$ in the optimality criterion (13) with the equivalent vector norm $\|\mathbf{a}\|$. Equations (38) are of the same form as those appearing in adjustment with the method of condition equations and have the well known solution

$$\mathbf{a}=\mathbf{R}^{-1}\mathbf{F}^T(\mathbf{F}\mathbf{R}^{-1}\mathbf{F}^T)^{-1}\mathbf{s}\,. \qquad (39)$$

The interpolating function is given by

$$f(P)=\mathbf{f}(P)^T\mathbf{a}=\mathbf{f}(P)^T\mathbf{R}^{-1}\mathbf{F}^T(\mathbf{F}\mathbf{R}^{-1}\mathbf{F}^T)^{-1}\mathbf{s}\,. \qquad (40)$$

Introducing the matrices

$$\mathbf{K}=\mathbf{F}\mathbf{R}^{-1}\mathbf{F}^{T} \tag{41}$$

and

$$\mathbf{k}(P)=\mathbf{F}\mathbf{R}^{-1}\mathbf{f}(P)\,, \tag{42}$$

equation (40) takes the simpler form

$$f(P)=\mathbf{k}(P)^{T}\,\mathbf{K}^{-1}\mathbf{s}\,. \tag{43}$$

5. The reproducing kernel

We shall next take a closer look at equation (43) which solves the interpolation problem, once the norm in F has been defined and the weight matrix $\mathbf{R}$ has been computed from equation (31). The elements of the matrices $\mathbf{K}$ and $\mathbf{k}(P)$ are

$$K_{ij}=\sum_{pq}(\mathbf{R}^{-1})_{pq}\,F_{ip}F_{jq}=\sum_{pq}(\mathbf{R}^{-1})_{pq}\,L_i(f_p)L_j(f_q) \tag{44}$$

and

$$k_i(P)=\sum_{pq}(\mathbf{R}^{-1})_{pq}\,F_{ip}\,f_q(P)=\sum_{pq}(\mathbf{R}^{-1})_{pq}\,L_i(f_p)\,f_q(P)\,. \tag{45}$$

If the functionals L_i, L_j are associated with the points P_i, P_j, respectively, and taking into account the linearity of the functionals, we have

$$K_{ij}=L_iL_j\sum_{pq}(\mathbf{R}^{-1})_{pq}\,f_p(P_i)\,f_q(P_j) \tag{46}$$

and

$$k_i(P)=L_i\sum_{pq}(\mathbf{R}^{-1})_{pq}\,f_p(P_i)\,f_q(P)\,. \tag{47}$$

The form of the above equations suggests the introduction of a two-point function

$$k(P,Q)=\sum_{pq}(\mathbf{R}^{-1})_{pq}\,f_p(P)\,f_q(Q) \tag{48}$$

so that equations (46) and (47) become

$$K_{ij}=L_iL_j\,k(P_i,P_j) \tag{49}$$

and

$$k_i(P)=L_i\,k(P,P_i)\,. \tag{50}$$

In the case of the simple interpolation described by equation (1), where the linear functionals are simply the evaluation functionals, we have $K_{ij}=k(P_i,P_j)$ and $k_i(P)=k(P,P_i)$.

The function $k(P,Q)$ is a symmetric function since

$$k(P,Q)=k(Q,P) \tag{51}$$

and belongs to the inner product space F both as a function of P and Q, as it can be easily seen from equation (48). Writing equation (48) in the form

$$k(P,Q)=\mathbf{f}(P)^{T}\,\mathbf{R}^{-1}\mathbf{f}(Q) \tag{52}$$

and recalling that $\mathbf{R}^{-1}$ is a positive definite matrix, we have that $k(P,P)>0$. An other property of $k(P,Q)$ is that it is a positive definite function, which means that for any set of points P_i, $i=1,2,\ldots,n$, the matrix with elements $k(P_i,P_j)$ is a positive definite matrix.

The function $k(P,Q)$ is called the *reproducing kernel* of the inner product space F because it has the "reproducing property"

$$\big(k(P,Q),f(Q)\big)_Q=f(P)\qquad\text{for any } f\in F\,. \tag{53}$$

For the proof of the reproducing property we have

$$\begin{aligned}\big(k(P,Q),f(Q)\big)_Q&=\\&=\left(\sum_{pq}(\mathbf{R}^{-1})_{pq}\,f_p(P)\,f_q(Q),\sum_i a_i\,f_i(Q)\right)_Q=\\&=\sum_{pq,i}a_i(\mathbf{R}^{-1})_{pq}\,f_p(P)(f_q,f_i)=\\&=\sum_{pi}a_i\,f_p(P)\sum_q(\mathbf{R}^{-1})_{pq}\,R_{qi}=\\&=\sum_{pi}a_i\,f_p(P)\,\delta_{pi}=\sum_{pi}a_i\,f_i(P)=f(P)\,.\end{aligned} \tag{54}$$

A simpler form of the reproducing kernel can be obtained by replacing the original base functions f_i with an equivalent set of orthogonal base functions ϕ_i, i.e. a set such that

$$(\phi_i,\phi_j)=0\qquad\text{for } i\neq j\,. \tag{55}$$

The positive definite symmetric matrix $\mathbf{R}$ has n positive eigenvalues λ_i with corresponding orthonormal eigenvectors $\mathbf{u}_i$, such that $\mathbf{R}\mathbf{u}_i=\lambda_i\mathbf{u}_i$, and can be diagonalized according to

$$\mathbf{R}=\mathbf{U}\mathbf{D}\mathbf{U}^{T}\,, \tag{56}$$

where $\mathbf{U}$ is an orthogonal matrix with $\mathbf{u}_i$ as its ith column and $\mathbf{D}$ is a diagonal matrix with $D_{ii}=\lambda_i$. An orthogonal set of functions ϕ_i can be introduced with the help of the relations

$$\phi_i=\sum_i U_{ji}\,f_j\,, \tag{57}$$

or in matrix form

$$\boldsymbol{\varphi}=\mathbf{U}^{T}\mathbf{f}\qquad\&\qquad\mathbf{f}=\mathbf{U}\boldsymbol{\varphi}\,, \tag{58}$$

where $\boldsymbol{\varphi}=[\phi_1\ \phi_2\ \ldots\ \phi_m]^T$.

For the same function f we have

$$f=\mathbf{f}^T\mathbf{a}=\boldsymbol{\varphi}^T\mathbf{U}^T\mathbf{a}=\boldsymbol{\varphi}^T\mathbf{c}=\sum_i c_i\phi_i\ , \tag{59}$$

where we have set $\mathbf{c}=\mathbf{U}^T\mathbf{a}$, for the coefficients c_i of f with respect to the new base functions ϕ_i. Inversion of equation (56) gives

$$\mathbf{R}^{-1}=\mathbf{U}\mathbf{D}^{-1}\mathbf{U}^T\ . \tag{60}$$

Taking equations (58) and (59) into account, equation (52) becomes

$$k(P,Q)=\mathbf{f}(P)^T\mathbf{R}^{-1}\mathbf{f}(Q)=\boldsymbol{\varphi}(P)^T\mathbf{U}\mathbf{R}^{-1}\mathbf{U}^T\boldsymbol{\varphi}(Q)=$$

$$=\boldsymbol{\varphi}(P)^T\mathbf{D}^{-1}\boldsymbol{\varphi}(Q)=\sum_i\frac{1}{\lambda_i}\phi_i(P)\phi_i(Q)\ . \tag{61}$$

Setting $\sigma_i^2=\frac{1}{\lambda_i}$ for the positive eigenvalues of $\mathbf{R}^{-1}$, we obtain the representation of the reproducing kernel in terms of orthogonal base functions

$$k(P,Q)=\sum_i\sigma_i^2\phi_i(P)\phi_i(Q)\qquad\text{with }\sigma_i^2=\frac{1}{\lambda_i}\ . \tag{62}$$

Noting that

$$\|\phi_i\|^2=(\phi_i,\phi_i)=\left(\sum_j U_{ji}\,f_j\,,\sum_k U_{ki}\,f_k\right)=$$

$$=\sum_{jk}U_{ji}(f_j,f_k)U_{ki}=\sum_{jk}U_{ji}R_{jk}U_{ki}=$$

$$=(\mathbf{U}^T\mathbf{R}\mathbf{U})_{ii}=D_{ii}=\lambda_i=\sigma_i^{-2}\ , \tag{63}$$

an orthonormal set of base functions is the set with elements

$$\psi_i=\frac{\phi_i}{\|\phi_i\|}=\sigma_i\phi_i \tag{64}$$

with $\|\psi_i\|=1$. The representation of the reproducing kernel with respect to the orthonormal base functions is

$$k(P,Q)=\sum_i\psi_i(P)\psi_i(Q)\ . \tag{65}$$

6. Prediction

Before the interpolation problem can be attacked from a stochastic point of view, it is necessary to introduce the appropriate tools from the theory of prediction of random variables. While the term estimation refers to the determination of estimates for fixed (non-random) variables from random data, prediction refers to the case where both data and parameter to be estimated (predicted) are random variables. More precisely, given the sample values x_1, x_2, …, x_n of n corresponding random variables, it is required that a prediction is to be made for the unknown sample value y of a corresponding random variable. All samples refer to the same outcome. Following the usual convention we use the same notation for random variables and their sample values, since the difference will be always easily understood from the context.

The random variables with known sample values constitute a random vector

$$\mathbf{x}=[x_1\ x_2\ \cdots\ x_n]^T \tag{66}$$

with known mean vector

$$\boldsymbol{\mu}_\mathbf{x}=E\{\mathbf{x}\} \tag{67}$$

and known covariance matrix

$$\mathbf{C}_{\mathbf{xx}}=E\{(\mathbf{x}-\boldsymbol{\mu}_\mathbf{x})(\mathbf{x}-\boldsymbol{\mu}_\mathbf{x})^T\}\ . \tag{68}$$

The random variable y, to be predicted, has also known mean $\mu_y=E\{y\}$ and known variance $\sigma_y^2=E\{(y-\mu_y)^2\}$. However, what makes prediction possible and meaningful is the knowledge of the cross-covariance matrix

$$\mathbf{C}_{y\mathbf{x}}=E\{(y-\mu_y)(\mathbf{x}-\boldsymbol{\mu}_\mathbf{x})^T\} \tag{69}$$

which expresses the stochastic dependence between $\mathbf{x}$ and y.

If $\tilde{y}$ is the prediction of y, the prediction error is

$$\varepsilon=\tilde{y}-y\ . \tag{70}$$

The optimality criterion for the determination of the prediction is the minimization of the mean square error

$$R=E\{\varepsilon^2\}=\min\ . \tag{71}$$

Predictions which minimize the mean square error are called *best* predictions. The best prediction is given by the conditional mean of y

$$\tilde{y}=E\{y|x\} \tag{72}$$

and its computation requires that the joint distribution of $\mathbf{x}$ and y is known. When only means, variances and covariances are known, we must restrict ourselves to two types of linear prediction:

Homogeneous linear prediction

$$\tilde{y}=\mathbf{d}^T\mathbf{x}\ , \tag{73}$$

and *inhomogeneous linear prediction*

$$\tilde{y}=\mathbf{d}^T\mathbf{x}+\kappa\ . \tag{74}$$

An other restriction may be posed by requiring that the prediction is to be *unbiased* in the sense that its mean value is

identical with the known mean value of the predicted random variable

$$E\{\tilde{y}\}=E\{y\}\,. \tag{75}$$

From the combination of the above, four types of prediction result:

inhomBLIP
(Best inhomogeneous Linear Prediction):
with minimum mean square error among all linear inhomogeneous predictions.

inhomBLUP
(Best inhomogeneous Unbiased Linear Prediction):
with minimum mean square error among all linear inhomogeneous predictions which are also unbiased.

homBLIP
(Best homogeneous Linear Prediction):
with minimum mean square error among all linear homogeneous predictions.

homBLUP
(Best homogeneous Unbiased Linear Prediction):
with minimum mean square error among all linear homogeneous predictions which are also unbiased.

The mean square error in the inhomogeneous case is

$$R=E\{(\mathbf{d}^T\mathbf{x}+\kappa-y)^2\}=$$

$$=\sigma_y^2+\mathbf{d}^T\mathbf{C}_{\mathbf{xx}}\mathbf{d}-2\mathbf{C}_{y\mathbf{x}}\mathbf{d}+(\boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{d}+\kappa-\mu_y)^2 \tag{76}$$

and in the homogeneous case

$$R=E\{(\mathbf{d}^T\mathbf{x}-y)^2\}=\sigma_y^2+\mathbf{d}^T\mathbf{C}_{\mathbf{xx}}\mathbf{d}-2\mathbf{C}_{y\mathbf{x}}\mathbf{d}+(\boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{d}-\mu_y)^2\,. \tag{77}$$

The condition for unbiasedness in the inhomogeneous case is

$$\mathbf{d}^T\boldsymbol{\mu}_{\mathbf{x}}+\kappa-\mu_y=0 \tag{78}$$

and in the homogeneous case

$$\mathbf{d}^T\boldsymbol{\mu}_{\mathbf{x}}-\mu_y=0\,. \tag{79}$$

We shall only sketch the derivations of the four predictions which are based on the standard techniques of minimization directly or under conditions with the use of Lagrange multipliers, for the determination of $\mathbf{d}$ and κ:

inhomBLIP: minimize R of equation (76).

inhomBLUP: minimize R of equation (76) subject to condition (78).

homBLIP: minimize R of equation (77).

homBLUP: minimize R of equation (78) subject to condition (79).

All four types of prediction can be written in the common form

$$\tilde{y}=\alpha\,\mu_y+\mathbf{C}_{y\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}(\mathbf{x}-\alpha\,\boldsymbol{\mu}_{\mathbf{x}}) \tag{80}$$

where the parameter α varies according to the type of prediction as follows:

InhomBLIP = inhomBLUP:

$$\alpha=1 \tag{81}$$

homBLIP:

$$\alpha=\frac{\boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x}}{1+\boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{C}_{\mathbf{xx}}^{-1}\boldsymbol{\mu}_{\mathbf{x}}} \tag{82}$$

homBLUP:

$$\alpha=\frac{\boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x}}{\boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{C}_{\mathbf{xx}}^{-1}\boldsymbol{\mu}_{\mathbf{x}}} \tag{83}$$

In the particular case where $\boldsymbol{\mu}_{\mathbf{x}}=\mathbf{0}$ and $\mu_y=0$, all prediction types coincide and we have

InhomBLIP=inhomBLUP=homBLIP=homBLUP:

$$\tilde{y}=\mathbf{C}_{y\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{x}\,. \tag{84}$$

This prediction is homogeneous and unbiased, but it cannot be improved by removing the restrictions of homogeneity and unbiasedness. We shall refer to it as best linear (unbiased) prediction.

7. Prediction in stochastic processes

The application of prediction to the solution of the interpolation problem is possible only if the available data $s_i=L_i(f)$ and the value $f(P)$ of the function f at any point can be considered as sample values of corresponding random variables. A function such that its values at any point are random variables is a random function and it is called stochastic process. The sample value of a stochastic process associated with a specific outcome is an ordinary deterministic function.

The stochastic characteristics of a stochastic process $f(P)$ are essentially known when the joint distribution of the random variables $f(P_1)$, $f(P_2)$, …, $f(P_n)$ is known for any set of points P_1, P_2, …, P_n. The application of the prediction equations of the previous section, however, require only the knowledge of means, variances and covariances. We shall therefore assume that known about the stochastic process $f(P)$ are its mean function $m(P)$, defined point-wise by

$$m(P)=E\{f(P)\}\,, \tag{85}$$

and its covariance function

$$C(P,Q)=E\{[f(P)-m(P)][f(Q)-m(Q)]\}\,. \tag{86}$$

We shall consider here the case where $m(P)=0$, which is always possible by replacing the original function $f(P)$

with the difference $f(P)-m(P)$. However it must be noted that the results obtained in this case using the prediction equations (84) are the same with those obtained without reduction from the inhomBLIP=inhomBLUP type of prediction, while homBLIP and homBLUP give modified results.

From the covariance function $C(P,Q)$ we can compute the variances and covariances of the random variables $f_P=f(P)$, s_1, s_2, …, s_n:

$$\sigma^2(f_P)=C(P,P) \tag{87}$$

$$\sigma(s_i,s_j)=E\{L_i(f)L_j(f)\}=L_iL_jE\{f(P_i)f(P_j)\}=$$

$$=L_iL_jC(P_i,P_j) \tag{88}$$

$$\sigma^2(s_i)=\sigma(s_i,s_i) \tag{89}$$

$$\sigma(f_P,s_i)=E\{f(P)L_i(f)\}=L_iE\{f(P)f(P_i)\}=L_iC(P,P_i)\,. \tag{90}$$

We can now apply equation (84) with $y=f(P)$ and $\mathbf{x}=\mathbf{s}$, to obtain the prediction, which for the specific sample function is an interpolation,

$$f(P)=\mathbf{c}(P)^T\mathbf{C}^{-1}\mathbf{s} \tag{91}$$

where we have set

$$\mathbf{C}=\mathbf{C_{ss}}=E\{\mathbf{s}\mathbf{s}^T\} \tag{92}$$

and

$$\mathbf{c}(P)=\mathbf{C}_{\mathbf{s}\,f(P)}=E\{\mathbf{s}\,f(P)\}\,. \tag{93}$$

The matrices appearing in the prediction equation (91) have elements

$$C_{ij}=L_iL_jC(P_i,P_j) \tag{94}$$

$$c_i(P)=L_iC(P,P_i)\,. \tag{95}$$

8. Relation between deterministic interpolation and stochastic prediction

Comparison of equations (91), (94), (95) with equations (43), (49), (50), respectively, shows a remarkable resemblance, the only difference being that the reproducing kernel $k(P,Q)$ is now replaced by the covariance function $C(P,Q)$. We may therefore conclude that:

Best linear (unbiased) prediction of the sample value of a function modeled as a stochastic process gives identical results with deterministic minimum norm interpolation, provided that the covariance function of the stochastic process is identical with the reproducing kernel associated with the minimized norm.

Now this conclusion is similar to a well known result from classical adjustment theory with a finite number of unknown parameters:

Best linear unbiased estimation when the observations are modeled as sample values of random variables gives identical estimates with deterministic weighted least squares, provided that the covariance matrix of the observations is identical with the inverse of the weight matrix used in the minimized quadratic form.

The above similarity becomes even stronger if equation (52) is interpreted as a propagation law for "inverse weights" and the reproducing kernel is seen as an "inverse weight function". In particular if the stochastic process $f(P)$ has sample functions in the function space F of all linear combinations of the given base functions f_1, f_2, …, f_m, it can be represented in the form

$$f(P)=a_1f_1(P)+a_2f_2(P)+\ldots+a_mf_m(P)=\mathbf{f}(P)^T\mathbf{a}\,, \tag{96}$$

where now the coefficients a_i are random variables with zero means and known covariance matrix

$$\mathbf{C_{aa}}=E\{\mathbf{a}\mathbf{a}^T\}\,. \tag{97}$$

The covariance function of the stochastic process $f(P)$ can be determined from the covariance matrix $\mathbf{C_{aa}}$ using the law of propagation of covariances

$$C(P,Q)=E\{f(P)f(Q)\}=E\{\mathbf{f}(P)^T\mathbf{a}\mathbf{a}^T\mathbf{f}(Q)\}=$$

$$=\mathbf{f}(P)^TE\{\mathbf{a}\mathbf{a}^T\}\mathbf{f}(Q)=\mathbf{f}^T(P)\mathbf{C_{aa}}\mathbf{f}(Q)\,. \tag{98}$$

Comparison with equation (52) leads to the conclusion that both the stochastic prediction approach and the deterministic interpolation approach give identical results when the covariance matrix $\mathbf{C_{aa}}$ of the coefficients coincides with the inverse of their weight matrix $\mathbf{R}$, i.e., when

$$\mathbf{C_{aa}}=\mathbf{R}^{-1}\,. \tag{99}$$

The elements of $E^m_{\mathbf{R}}$ with inner product defined by equation (32), are *m*-vectors and can be considered as functions with domain of definition the integers 1, 2, …, *m*, i.e., $a_i=a(i)$, and the same inner product

$$\big(a(i),b(i)\big)=\mathbf{a}^T\mathbf{R}\mathbf{b}=\sum_{pq}a(p)b(q)\,. \tag{100}$$

The reproducing kernel is in this case given by $k(i,j)=$ $=(\mathbf{R}^{-1})_{ij}$ as it follows from the reproducing property.

$$\big(k(i,j),a(j)\big)_j=\sum_{pq}R_{pq}k(i,p)a(q)=$$

$$=\sum_{pq}(\mathbf{R}^{-1})_{ip}R_{pq}a(q)=\sum_i\delta_{iq}a(q)=a(i)\,. \tag{101}$$

In the same way that the vector **a** corresponds to the "function" $a(i)$, the "reproducing matrix" $\mathbf{R}^{-1}$ corresponds to the reproducing kernel $k(i,j)$. The above mentioned identity of results between stochastic and deterministic approach, is guaranteed in this finite dimensional case from the identity (99) between reproducing matrix and covariance matrix. In the most general infinite dimensional case, to be examined next, the identity must hold between the reproducing kernel function and the covariance function, which functions are in a certain sense infinite entities.

9. Interpolation with an infinite number of base functions

Let us now return to the interpolation problem and consider the possibility to have an infinite number of base functions f_1, f_2, …, instead of a finite number m as before. Now we cannot consider the set of all linear combinations of the base functions, since the infinite sum $a_1 f_1 + a_2 f_2 + \ldots$ may not be defined at all. With the help of the definition of an inner product and the derived definition of norm we may initially restrict ourselves to finite linear combinations $f(n) = a_1 f_1 + a_2 f_2 + \ldots + a_n f_n$, with any number of terms and finally include all limits of convergent sequences $f(n)$ where convergence is defined with the help of the relevant norm.

A more reasonable approach is to start not with the definition of the base functions but with the definition of a space of functions H, like the function space F of the finite case, which is appropriate for the purposes of interpolation. The elements of H must have characteristics similar to those of the unknown function, e.g. solutions of a certain differential equation. An appropriate model is a *Hilbert space* of functions, which is an infinite dimensional space with inner product, which is furthermore complete, a term meaning that it contains all the limits of convergent sequences of its elements. Two further properties that the Hilbert space must have in order to serve the purposes of interpolation are the properties of separability and of the existence of a reproducing kernel. Separability means that H has a complete orthonormal system of elements, i.e., an orthonormal set of elements ψ_1, ψ_2, …, such that every element f of H can be expressed in the form of an infinite sum $f = a_1\psi_1 + a_2\psi_2 + \cdots$. A reproducing kernel

$$k(P,Q) = \psi_1(P)\,\psi_1(Q) + \psi_2(P)\,\psi_2(Q) + \ldots \tag{102}$$

exists when the evaluation functionals acting on the elements of H are continuous, in which case the existence of the infinite sum in equation (102) is guaranteed. The relevant function space is called in this case reproducing kernel Hilbert space and is denoted by H_k.

In most cases one starts with a known reproducing kernel $k(P,Q)$ which as a function of either P or Q has properties similar to those of the unknown function. The space H_k can be constructed from the known reproducing kernel if a Hilbert space H is known such that $k(P,Q) \in H$ either as a function of P or Q. Usually we deal with continuous functions which are also square integrable and $H = L^2$, where L^2 is the Hilbert space of functions which are square integrable over their domain of definition with inner product defined by equation (27) and denoted by $(\ ,\ )_{L^2}$.

The kernel $k(P,Q)$ defines an integral operator K with positive eigenvalues k_n^2 and corresponding eigenfunctions ϕ_n

$$(K\phi_n)(P) = \big(k(P,\ ),\phi_n\big)_{L^2} = \int k(P,Q)\phi_n(Q)\,dQ = k_n^2\phi_n(P)\,. \tag{103}$$

According to *Mercer's theorem* (Grenander, 1981, Kubackova, 1975), the eigenfunctions ϕ_n are a complete orthonormal system in L^2 and the reproducing kernel accepts the representation

$$k(P,Q) = \sum_n k_n^2\phi_n(P)\phi_n(Q)\,. \tag{104}$$

Setting $\psi_n = k_n\phi_n$, the Hilbert space H_k has the ψ_n functions as a complete orthonormal system and its reproducing kernel is given by

$$k(P,Q) = \sum_n \psi_n(P)\,\psi_n(Q)\,. \tag{105}$$

When the reproducing kernel is known, minimum norm interpolation can be carried out using equation (43) even though the base functions ψ_n remain unknown.

The problem of choosing an appropriate reproducing kernel can be also solved in the infinite dimensional case with the introduction of a stochastic model for the unknown function. The unknown function may be considered a sample function of a corresponding stochastic process $f(P,\omega)$. The argument ω, meaning outcome, is introduced as a convenient means of discrimination between random and non random quantities. If $C(P,Q)$ is the covariance function of the stochastic process $f(P,\omega)$, there exists a Hilbert space H_C with $C(P,Q)$ as its reproducing kernel. Assuming continuity of $C(P,Q)$, which implies square integrability, we can follow the same approach as previously to define a "covariance operator" with eigenvalues σ_n^2 and corresponding orthonormal eigenfunctions ϕ_n,

$$\int C(P,Q)\phi_n(Q)\,dQ = \sigma_n^2\phi_n(P)\,, \tag{106}$$

so that Mercer's theorem gives the representation

$$C(P,Q) = \sum_n \sigma_n^2\phi_n(P)\phi_n(Q)\,. \tag{107}$$

With the introduction of the random coefficients

$$a_n(\omega) = \big(f(P,\omega),\phi_n(P)\big)_{L^2} = \int f(P,\omega)\,\phi_n(P)\,dP\,, \tag{108}$$

we have the *Karhunen-Loeve expansion*

$$f(P,\omega) = \sum_n \sigma_n^2 a_n(\omega)\,\phi_n(P) \tag{109a}$$

with uncorrelated coefficients

$$E\{a_i a_j\}=\delta_{ij}\sigma_i^2 . \qquad (109b)$$

In view of the above equations the eigenvalues σ_n^2 are called "degree variances". We can also use the complete orthonormal system of elements $\psi_n=\sigma_n\phi_n$ in H_C and the coefficients $b_n=\sigma_n a_n$ to obtain the expansion

$$f(P,\omega)=\sum_n b_n(\omega)\psi_n(P), \qquad \text{with } E\{b_i b_j\}=\delta_{ij} . \qquad (110)$$

For a more rigorous treatment of the relation between the stochastic and the deterministic approach see Sansó (1986).

10. Invariance of interpolation

We have so far used the "points" P, Q, etc. as arguments of the relevant functions. When computations are to be actually carried out we shall have in the place of P, Q, … the vectors $\mathbf{r}_P$, $\mathbf{r}_Q$, … of their coordinates with respect to some fixed coordinate system in the corresponding domain. Since the definition of the coordinate system is more or less arbitrary, an obvious property of a reasonable interpolation method is its independence from the particular coordinate system used. If G is any coordinate transformation which transforms coordinates $\mathbf{r}$ into equally admissible coordinates $G\mathbf{r}$, the interpolation must be invariant under the transformation G. Returning to the general definition of interpolation given by equation (2) the above mentioned invariance can be expressed by

$$f(\mathbf{s},\mathbf{r},\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_n)=f(\mathbf{s},G\mathbf{r},G\mathbf{r}_1,G\mathbf{r}_2,\ldots,G\mathbf{r}_n) . \qquad (111)$$

The group $\mathcal{G}$ of transformations G with respect to which the interpolation must be invariant depends on the particular situation and especially on the type of the domain of the function and the way that the coordinate system is defined. We mention as examples shifts in the time domain, rigid transformations (translations+rotations) in the plane, rotations on the sphere, etc. Such transformations are in fact transformations of the reference frame used for the definition of coordinates.

In the case of minimum norm interpolation described by equation (43), invariance of the interpolation is guaranteed by the invariance of the reproducing kernel

$$k(\mathbf{r}_P,\mathbf{r}_P)=k(G\mathbf{r}_P,G\mathbf{r}_P) . \qquad (112)$$

In the case of best linear prediction described by equation (91), invariance of the prediction is guaranteed by the invariance of the covariance function

$$C(\mathbf{r}_P,\mathbf{r}_P)=C(G\mathbf{r}_P,G\mathbf{r}_P) . \qquad (113)$$

Every such transformation G can be interpreted not only as a transformation of the reference frame while the relevant points retain their physical position, but also as an equivalent physical transformation of the whole network of points while the reference frame remains unmoved. These two situations are in fact numerically indistinguishable. Therefore the invariance of the interpolation may be interpreted in a different way: We require that the interpolation depends in addition to the data values only on the configuration of the data points and not on their particular placement (position + orientation) in the function domain.

On the basis of the above interpretation of interpolation invariance, a deterministic optimality criterion may be used which, as we shall see, give a solution to the problem of the choice of the reproducing kernel. This criterion is the minimization of the mean square error of the interpolation, where the term mean has now the deterministic meaning of the average over the whole domain of definition of the unknown function. More precisely the interpolation should be performed in such a way that if it is repeated for all the possible placements of the network of prediction and data points over the domain of definition, the mean value of the squares of the interpolation errors should be as small as possible. If the configuration of the network is defined through its actual placement

$$\zeta=\{\mathbf{r},\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_n\} \qquad (114)$$

the interpolated value at $\mathbf{r}$ according to equation (3) is $J_{\mathbf{r}}(\mathbf{s},\zeta)$, and if $f(\mathbf{r})$ is the true value of the unknown function, the corresponding interpolation error is

$$\varepsilon(\mathbf{r})=J_{\mathbf{r}}(\mathbf{s},\zeta)-f(\mathbf{r}) . \qquad (115)$$

Denoting with $G\zeta$ any other placement obtained from the original by a transformation G, our problem is to find an interpolator J which is optimal according to

$$M(J)=M_{\mathcal{G}}\{\varepsilon(G\mathbf{r})^2\}=\int_{\mathcal{G}}[J_{G\mathbf{r}}(\mathbf{s},G\zeta)-f(G\mathbf{r})]^2\,dG=\min . \qquad (116)$$

$M(J)$ as defined above is the average over all possible placements of the actual configuration ζ, obtained as the transformation G runs through all the elements of the relevant transformation group $\mathcal{G}$. The solution to this problem is given by equation (43), just like in the case of minimum norm interpolation, the only difference being that the kernel $k(P,Q)$ to be used is now defined by

$$k(\mathbf{r}_P,\mathbf{r}_Q)=M_{\mathcal{G}}\{f(G\mathbf{r}_P)f(G\mathbf{r}_Q)\} . \qquad (117)$$

For a proof see Sansó (1978). As a consequence of the above definition the kernel $k(P,Q)$ satisfies the invariance condition (112) and has the form

$$k(\mathbf{r}_P,\mathbf{r}_Q)=k(|\mathbf{r}_P-\mathbf{r}_Q|) , \qquad (118)$$

where $|\mathbf{r}_P-\mathbf{r}_Q|$ is the "distance" between the points P and Q measured along the geodesic in their domain of definition . Mean square error interpolation, as defined above, is a particular case of minimum norm interpolation, where the norm is defined in accordance with the definition and interpretation of the reproducing kernel given by equation (117).

Similar invariance considerations in the stochastic case of best linear prediction lead to the restriction of the stochastic process model for the unknown function, to a class of (covariance) homogeneous and isotropic stochastic precesses. These are processes satisfying equation (113), or

$$C(\mathbf{r}_P,\mathbf{r}_Q)=C(|\mathbf{r}_P-\mathbf{r}_Q|)\,. \tag{119}$$

Within the class of homogeneous and isotropic stochastic processes, the most interesting ones are the so called ergodic stochastic processes. These are stochastic processes for which averages over samples (expectations) coincide with averages over the domain of definition, so that

$$C(|\mathbf{r}_P-\mathbf{r}_Q|)=E\{f(\mathbf{r}_P)f(\mathbf{r}_Q)\}=M_G\{f(G\mathbf{r}_P)f(G\mathbf{r}_Q)\}\,,$$

$$|\mathbf{r}_P-\mathbf{r}_Q|=\text{const.}\,. \tag{120}$$

From the above equation it follows that in order to determine the covariance function, which is needed for the prediction of the unknown function, the function itself must be known, thus falling in a vicious circle (Lauritzen, 1973). However, an approximation to the covariance function can be found from the available data using statistical sampling techniques. The situation has certain similarities with least squares adjustment where the application of the BLUE criterion requires that the true covariance matrix of the observations is known, while in practice only estimates obtained from sampling with repeated observations is used.

Comparison of equation (120) with equation (117) leads to the following conclusion:

Best linear (unbiased) prediction of the sample value of a function modeled as an ergodic stochastic process gives identical results with deterministic mean square error interpolation.

While both the deterministic and stochastic approach give the same numerical results, it must be emphasized that the relevant models used are fundamentally different, and this difference has a direct effect on the interpretation of the results, especially with respect to their optimality properties. Recently (Sansó, 1986), a stochastic interpretation for the originally deterministic mean square error interpolation has been proposed. It is based on the idea of considering random transformations $G(\omega)$ rather than the deterministic ones used above.

11. The problem of trend or bias

In minimum norm interpolation, an original unknown function f^a is beforehand replaced with a new unknown "disturbing function" $f=f^a-f^0$, with the help of known "reference function" f^0. The choice of reference function f^0 is an essential part of the whole procedure and has a direct effect on the results of the interpolation, since data are either reduced accordingly

$$s_i=s_i^a-s_i^0=L_i(f^a)-L_i(f^0)\,, \tag{121}$$

in the case of linear functionals, or they differ because they result from the linearization of originally non-linear functionals L_i, with the help of different approximate values $s_i^0=L_i(f^0)$

$$s_i=s_i^a-L_i(f^0)=\nabla L_{i,f^0}(f^a-f^0)=L_i(f)\,. \tag{122}$$

One way to deal with this problem is to determine a reference function f^0, or trend as it is more usually called in this context, from the available data, using smoothing least squares interpolation with a small number m of base functions, where of course $m<n$, n being the number of available data. Using equation (26) for the determination of the relevant coefficients, the interpolated trend function becomes

$$f^0(P)=a_1f_1(P)+a_2f_2(P)+\ldots+a_mf_m(P)=\mathbf{f}(P)^T\mathbf{a}=$$

$$=\mathbf{f}(P)^T(\mathbf{F}^T\mathbf{P}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{P}\mathbf{s}\,. \tag{123}$$

Using next minimum norm interpolation for the disturbing function $\delta f=f-f^0$ with reduced data $\delta\mathbf{s}=\mathbf{s}-\mathbf{s}^0$, where $s_i^0=L_i(f^0)$, we obtain according to equation (43)

$$\delta f(P)=\mathbf{k}(P)^T\mathbf{K}^{-1}\delta\mathbf{s}\,, \tag{124}$$

and the total interpolating function $f(P)=f^0(P)+\delta f(P)$ becomes

$$f(P)=\mathbf{f}(P)^T(\mathbf{F}^T\mathbf{P}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{P}\mathbf{s}+\mathbf{k}(P)^T\mathbf{K}^{-1}(\mathbf{s}-\mathbf{s}^0)\,. \tag{125}$$

A different approach can be based on the simultaneous interpolation and trend determination based on a hybrid approach of the form

$$f(P)=a_1f_1(P)+\ldots+a_mf_m(P)+\delta a_1\delta f_1(P)+\ldots+\delta a_q\delta f_q(P)=$$

$$=\mathbf{f}(P)^T\mathbf{a}+\delta\mathbf{f}(P)^T\delta\mathbf{a}=f^0(P)+\delta f(P)\,,\quad \|\delta f\|^2=\min \tag{126}$$

where possibly $q\to\infty$. The relevant equations for the available data

$$\mathbf{s}=\mathbf{F}\mathbf{a}+\delta\mathbf{F}\delta\mathbf{a}\,,\quad \delta\mathbf{a}^T\mathbf{R}\delta\mathbf{a}=\min\quad (R_{ij}=(\delta f_i,\delta f_j))\,, \tag{127}$$

where $F_{ij}=L_i(f_j)$ and $\delta F_{ij}=L_i(\delta f_j)$, are of the same type as the mixed equations of least squares adjustment, with well known solution

$$\mathbf{K}=\delta\mathbf{F}\mathbf{R}^{-1}\delta\mathbf{F}^T \tag{128}$$

$$\mathbf{a}=(\mathbf{F}^T\mathbf{K}^{-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{K}^{-1}\mathbf{s} \tag{129}$$

$$\delta\mathbf{a}=\mathbf{R}^{-1}\delta\mathbf{F}^T\mathbf{K}^{-1}(\mathbf{s}-\mathbf{F}\mathbf{a})=\mathbf{R}^{-1}\delta\mathbf{F}^T\mathbf{K}^{-1}(\mathbf{s}-\mathbf{s}^0)\,. \tag{130}$$

The last term of equation (130) is justified from the fact that

$$(\mathbf{Fa})_i = \sum_j F_{ij} a_j = \sum_j a_j L_i(f_j) = L_i \sum_j a_j f_j = L_i(f^0) = s_i^0 . \tag{131}$$

Replacing $\mathbf{a}$ and $\delta\mathbf{a}$ from equations (129) and (130), respectively, in equation (126) we obtain the interpolation

$$f(P) = \mathbf{f}(P)^T (\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{K}^{-1} \mathbf{s} + \\ + \delta\mathbf{f}(P)^T \mathbf{R}^{-1} \delta\mathbf{F}^T \mathbf{K}^{-1} (\mathbf{s} - \mathbf{s}^0) =$$

$$= \mathbf{f}(P)^T (\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{K}^{-1} \mathbf{s} + \mathbf{k}(P)^T \mathbf{K}^{-1} (\mathbf{s} - \mathbf{s}^0) . \tag{132}$$

Comparison with equation (125) leads to the following conclusion:

Minimum norm interpolation with simultaneous trend determination is equivalent to prior trend determination with smoothing least squares interpolation using weight matrix $\mathbf{P}$*, followed by minimum norm interpolation with properly reduced data, provided that* $\mathbf{P} = \mathbf{K}^{-1}$ *where* $\mathbf{K}$ *is the matrix with elements* $K_{ij} = L_i L_j k(P_i, P_j)$.

An extreme case of trend removal is the case where only one base function $\mu(P)$ is used ($m=1$), i.e.

$$f(P) = \alpha\, \mu(P) + \delta f(P) , \qquad \|\delta f\|^2 = \min . \tag{133}$$

In this case $\mathbf{F}$ is a vector with elements $F_i = L_i(\mu)$, or

$$\mathbf{F} = \boldsymbol{\mu} = [L_1(\mu)\; L_1(\mu) \cdots L_n(\mu)]^T . \tag{134}$$

Application of equations (128) through (132) gives the solution

$$\alpha = (\boldsymbol{\mu}^T \mathbf{K}^{-1} \boldsymbol{\mu})^{-1} \boldsymbol{\mu}^T \mathbf{K}^{-1} \mathbf{s} \tag{135}$$

$$s_i^0 = \alpha\, L_i(\mu) \qquad \Rightarrow \qquad \mathbf{s}^0 = \alpha\, \boldsymbol{\mu} \tag{136}$$

$$f(P) = \alpha\, \mu(P) + \mathbf{k}(P)^T \mathbf{K}^{-1} (\mathbf{s} - \alpha\, \boldsymbol{\mu}) . \tag{137}$$

If we assume that the above coefficient is not completely unknown but a prior estimate α_0 is available with weight p_α, equation (135) must be replaced with the corresponding well known solution to the problem of observation equations with weights on the unknowns

$$\alpha = \alpha_0 + (\boldsymbol{\mu}^T \mathbf{K}^{-1} \boldsymbol{\mu} + p_\alpha)^{-1} \boldsymbol{\mu}^T \mathbf{K}^{-1} (\mathbf{s} - \alpha_0 \boldsymbol{\mu}) . \tag{138}$$

The interpolation is given by equation (137) with the value of α given above.

Of special interest is the case where $\alpha_0 = 0$ and $p_\alpha = 1$, in which case equation (138) becomes

$$\alpha = (1 + \boldsymbol{\mu}^T \mathbf{K}^{-1} \boldsymbol{\mu})^{-1} \boldsymbol{\mu}^T \mathbf{K}^{-1} \mathbf{s} . \tag{139}$$

When stochastic prediction is used similar problems arise, where now the problem of the unknown trend function is replaced with the problem of the unknown mean function of the stochastic process. In a similar way the original unknown function is considered to be a sample function of a stochastic process $f^a(P,\omega)$ whose mean function $m(P)$ is taken to be equal to the reference function $f^0(P)$ used for the linearization. In this way a zero mean stochastic process

$$f(P,\omega) = f^a(P,\omega) - m(P) = f^a(P,\omega) - f^0(P,\omega)$$

is obtained and best linear (unbiased) prediction can be applied using equation (91). Note that this reduction is allowed only in the case of prediction of the type inhomBLIP=inhomBLUP. In the case where prediction of the type homBLIP (which by the way is biased) or homBLUP is used such a subtraction of mean values has an immediate effect on the prediction results as can easily be seen from equations (82) and (83) used in combination with equation (80).

When the true mean function $m(P)$ of the stochastic process is different from the used reference function $f^0(P)$, the unbiased prediction of the type inhomBLIP=inhomBLUP leads to biased results, in the same way that errors with non-zero means in adjustment lead to biased parameter estimates. This bias problem can be treated in a way similar to that in the case of deterministic interpolation. Instead of the reference function $f^0(P)$, or in addition to it, an unknown but properly parametrized mean function

$$m(P) = a_1 m_1(P) + a_2 m_2(P) + \ldots + a_k m_k(P) \quad (k < n) \tag{140}$$

can be used, where $m_i(P)$ are known (base) functions and a_i are unknown coefficients that can be estimated, either prior to, or simultaneously with the prediction using the BLUE criterion.

We shall finally look into the case of prediction of the type homBLIP and homBLUP in the case where the mean function m(P) of the stochastic process is assumed to be known. In this case we can apply equations (80), (81) and (82) with $y = f(P)$, $\mathbf{x} = \mathbf{s}$ to obtain

$$f(P) = \alpha\, m(P) + \mathbf{c}(P)^T \mathbf{C}^{-1} (\mathbf{s} - \alpha\, \mathbf{m}) \tag{141}$$

where $\mathbf{c}(P)$ and $\mathbf{C}$ are as defined by equations (92) and (93), respectively,

$$m(P) = E\{f(P)\} \tag{142}$$

and

$$\mathbf{m} = \boldsymbol{\mu}_\mathbf{s} = E\{\mathbf{s}\} \tag{143}$$

with elements

$$m_i = E\{s_i\} = E\{L_i(f)\} = L_i(E\{f\}) = L_i(m) . \tag{144}$$

The coefficient α has the following values, according to the type of prediction:

homBLUP:

$$\alpha = (\mathbf{m}^T \mathbf{C}^{-1} \mathbf{m})^{-1} (\mathbf{m}^T \mathbf{C}^{-1} \mathbf{s}) \tag{145}$$

homBLIP:

$$\alpha=(1+\mathbf{m}^T\mathbf{C}^{-1}\mathbf{m})^{-1}(\mathbf{m}^T\mathbf{C}^{-1}\mathbf{s}) \qquad (146)$$

Comparison of equations (141), (145) and (146) with equations (137), (135) and (139), respectively, shows (once more!) a remarkable resemblance. There are only two differences: (a) the reproducing kernel $k(P,Q)$ has been replaced by the covariance function $C(P,Q)$ and (b) the single base function $\mu(P)$ used for the trend removal has been replaced by the known mean function of the relevant stochastic process. And once more, it must be emphasized that this is only a numerical equivalence of results, which have however very different interpretations, due to the fundamental differences of the underlying models.

12. Some final remarks

In order to preserve the unity of the presentation, we have consciously avoided much of the terminology used for these methods in geodesy and elsewhere. It is now time that a connection is established:

Minimum norm interpolation and inhomogeneous best linear (unbiased) prediction with zero-mean stochastic processes are, of course, the deterministic and stochastic counterparts of the famous collocation method.

Trend removal techniques come from relevant stochastic models in the analysis of time series.

Interpolation or prediction with simultaneous trend removal has a certain resemblance with least squares collocation, where deterministic and stochastic or function-related parameters (signals) appear simultaneously in the same observational model. However the deterministic parameters in least squares collocation have a specific physical meaning in contrast to the lack of any such interpretation for the coefficients of the trend function.

Hybrid interpolation is in a certain sense related to Tikhonov's method of regularization for ill-posed problems.

Homogeneous best linear unbiased prediction is related to the kriging method used in geostatistics.

References

Bjerhammar, A. (1985): "A Robust Approach to Global Problems in Physical Geodesy", Bulletin Geodesique, vol. 59, pp. 303-315.

Davis, P.J. (1975): "*Interpolation and Approximation*", Dover.

Dermanis, A. (1976): "*Probabilistic and Deterministic Aspects of Linear Estimation in Geodesy*", Ohio State University Department of Geodetic Science, Report No. 244, Columbus.

Dermanis, A. (1978): "Adjustment of Geodetic Observations in the Presence of Signals", Erice Summer School 1978, Bollettino di Geodesia e Scienze Affini, XXXVIII (1979), no. 4, pp. 513-539.

Dermanis, A. (1983): "Theory and Application of Collocation in Surveying". In: M. Unguendoli (ed.): '*Techniche moderne di analisi dei dati geodetici con particolare riguardo alla collocazione'*, pp. 37-67, Editrice CLUEB, Bologna.

Dermanis, A. (1984): "Kriging and Collocation - A Comparison", Manuscripta Geodaetica, vol. 9, 159-167.

Grenander, U. (1981): "*Abstract Inference*", Wiley.

Krarup, T. (1969): "*A Contribution to the Mathematical Foundation of Physical Geodesy*", Danish Geodetic Institute Publ. No. 44, Copenhagen.

Kubackova, L. (1975): "Probability Model for Studying the Spectral Anomalous Geophysical Fields", Studia Geophysica et Geodaetica, vol. 19, 1-7.

Lauritzen, S. (1973): "*The Probabilistic Background of Some Statistical Methods in Physical Geodesy*", Danish Geodetic Institute, Publ. No. 48, Copenhagen.

Meissl, P. (1975): "Elements of Functional Analysis". In: Brosowski and Martensen (eds.): 'Methoden und Verfahren der mathematischen Physik' vol. 12, pp. 19-78, B.I.-Wissenschaftsverlag, Mannheim.

Meschkowski, H. (1962): "*Hilbertsche Raume mit Kernfunktion*", Springer.

Moritz, H. (1980): "*Advanced Physical Geodesy*", Wichmann Verlags, Karlsruhe.

Parzen, E. (1959): "Statistical Inference on Time Series by Hilbert Space Methods". Reprinted in E. Parzen (1967): *'Time Series Analysis Papers*', pp. 251-282, Holden-Day.

Sansó, F. (1978): "The minimum Mean Square Estimation Error Principle in Physical Geodesy", 7th Symposium on Mathematical Geodesy, Assissi, June 1978. Bollettino di Geodesia e Scienze Affini, XXXIX (1980), no. 2, pp. 111-129.

Sansó, F. (1986): "Statistical Methods in Physical Geodesy". In: H. Sünkel (ed.): '*Mathematical and Numerical Techniques in Physical Geodesy'*, Lecture Notes in Earth Sciences, vol. 7, pp. 49-155, Springer.

Sansó, F., R. Barzaghi and C.C. Tscherning (1986): "Choice of Norm for the Density Distribution of the Earth", Geophysical Journal of the Royal Astronomical Society, vol. 87, pp. 123-141.

Schaffrin, B. (1983): "*Model Choice and Adjustment Techniques in the Presence of Prior Information*", Ohio State University Department of Geodetic Science and Surveying, Report No. 351, Columbus.

Schaffrin, B. (1985): "*Das geodätische Datum mit stochastischer Vorinformation*", Habilitationsschrift, Universität Stuttgart. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 313, München.

Tikhonov, A. and V. Arsenin (1977): "*Solutions of Ill-Posed Problems*", Wiley.

Tscherning, C.C. (1979): "Comparison of Some Methods for the Detailed Representation of the Earth's Gravity Field", XVII General Assembly of IUGG, Canberra.

Tscherning, C.C. (1985): "Current Problems in Gravity Field Approximation", 1st Hotine-Marussi Symposium, Rome, June 1985.

Tscherning, C.C. (1985): "Local Approximations of the Gravity Potential by Least Squares Collocation". In K.P. Schwarz (ed.): 'Proceedings of the International Summer School on Local Gravity Field Approximation, Beijing Aug.-Sept. 1984'. The University of Calgary Division of Surveying Engineering, Publication No. 60003.

Tscherning, C.C. (1986): "Functional Methods for Gravity Field Approximation". In: H. Sünkel (ed.): 'Mathematical and Numerical Techniques in Physical Geodesy', Lecture Notes in Earth Sciences, vol. 7, pp.4-47, Springer.