manuscripta geodaetica Vol. 9 (1984) 159-167

KRIGING AND COLLOCATION - A COMPARISON

A. DERMANIS

SUMMARY

A simple comparison of the Kriging estimation techniques with the method of collocation widely used in geodesy is carried out. Similarities and differences between the two methods are pointed out and explained on the basis of their corresponding underlying stochastic assumptions.

1. INTRODUCTION

A collection of methods for prediction or interpolation of function values originating in geostatistics and called kriging (after Krige) have been presented by Blais (1982).

Kriging is an example of the developement of similar ideas in different fields of applied science where they appear under particular names thus giving rise to a scientific babel which contributes to the lack of interdisciplinary communication. The other example, of course, is the familiar to the geodetic world method of collocation.

Here we shall give a comparison between collocation and the various kriging techniques with the exception of some non-linear extensions of the latter. The familiar extensions of collocation where the data are more general than point-values of an underlying function will also be excluded from such a comparison.

Both methods follow the principle of unbiased minimum error variance prediction when the underlying function is considered as a second order stochastic process with known, or estimated, covariance function. However a

deterministic interpretation is also possible, where both methods are simply viewed as interpolation methods.

We refer only to Blais (1982) where further references to the literature can be found.

2. STANDARD PREDICTION TECHNIQUES AND COLLOCATION

Let f_i, $i = 1,2,\ldots,n$ be a set of known realizations of respective random variables. An estimate $\hat{f}$ of the realization f of a respective new random variable, correlated to the former ones, can be uniquely obtained if the following three criteria are combined:

i) Linearity

$$\hat{f} = \sum_i \lambda_i f_i + b \quad , \tag{1}$$

ii) Unbiasedness

$$E\{\hat{f}\} = E\{f\} \Rightarrow E\{\varepsilon\} = 0 \quad , \tag{2}$$

for the prediction error

$$\varepsilon = \hat{f} - f \tag{3}$$

iii) Minimum error variance

$$\sigma_\varepsilon^2 = E\{\varepsilon^2\} = \text{minimum} \quad . \tag{4}$$

Using matrix notation the derived prediction is

$$\hat{f} = \mathbf{C}_{f\mathbf{f}} \, \mathbf{C}_{\mathbf{ff}}^{-1} \, \mathbf{f} \tag{5}$$

in the particular case of zero-mean random variables

$$E\{f_i\} = 0 \Rightarrow E\{\mathbf{f}\} = \mathbf{0} \, , \quad E\{f\} = 0 \quad . \tag{6}$$

Here $\mathbf{f}$ is the vector of the f_i values and $\mathbf{C}_{f\mathbf{f}}$, $\mathbf{C}_{\mathbf{ff}}$ the covariance matrices

of the respective subscript vectors.

In the most general case of random variables with non-zero means

$$E\{f_i\} = m_i \neq 0, \quad E\{\mathbf{f}\} = \mathbf{m} \neq \mathbf{0}, \quad E\{f\} = m \neq 0 \tag{7}$$

the prediction becomes

$$\hat{f} - m = \mathbf{C}_{f\mathbf{f}} \, \mathbf{C}_{\mathbf{ff}}^{-1} \, (\mathbf{f} - \mathbf{m}) \quad . \tag{8}$$

It must be noted that equation (5) can also be derived by a priori setting $b = 0$ in the linearity condition (1). In such a case however, equation (8) cannot follow as a generalization of (5); the assumption of a not necessarily zero additive term is in this case essential as it will be shown in the following.

The realization of the prediction requires the knowledge of the means and covariances for all related random variables.

Collocation in its most simple form is a straightforward application of equation (8) to the case where f_i and f are the values of a stochastic process $f(\mathbf{x})$, i.e.,

$$f_i = f(\mathbf{x}_i), \qquad f = f(\mathbf{x}) \tag{9}$$

at points $\mathbf{x}_i$ and $\mathbf{x}$ of its domain of definition. In the most general case, not considered here, f_i, f can be values resulting from the application of appropriate functionals on the stochastic process $f(\mathbf{x})$. The elements of the covariance matrices $\mathbf{C}_{f\mathbf{f}}$, $\mathbf{C}_{\mathbf{ff}}$ are computed from the known covariance function $c(\mathbf{x},\mathbf{x}')$ of the stochastic process $f(\mathbf{x})$, while the means m, $\mathbf{m}$ from the corresponding known mean function $m(\mathbf{x})$.

For Gaussian processes the prediction $\hat{f}$ given by equation (8) coincides with the conditional expectation of f given the values f_i, which is an unbiased minimum error variance prediction of f without the restriction of linearity in equation (1).

When the stochastic process $f(\mathbf{x})$ has a non-zero but unknown mean function equation (8) cannot be directly applied. In practice, the required mean values m, $\mathbf{m}$ are replaced by estimates $\hat{m}$, $\hat{\mathbf{m}}$ resulting from some estimate $\hat{m}(\mathbf{x})$ of the unknown mean function $m(\mathbf{x})$.

The usual way of obtaining such an estimated mean function is to follow the procedure called trend removal where the unknown trend function $m(\mathbf{x})$ is

restricted to be of the form

$$m(\mathbf{x}) = \sum_{k=1}^{q} a_k \varphi_k(\mathbf{x}) = \mathbf{a}^T \boldsymbol{\varphi}(\mathbf{x}) = \boldsymbol{\varphi}(\mathbf{x})^T \mathbf{a} \qquad (10)$$

i.e., a linear combination of some selected base functions $\varphi_k(\mathbf{x})$. The optimal values of the coefficients a_k are obtained by solving the least squares problem ($q < n$)

$$\mathbf{f} = \boldsymbol{\Phi}\,\mathbf{a} + \mathbf{g}\,, \qquad \mathbf{g}^T \mathbf{C}_{\mathbf{ff}}^{-1}\,\mathbf{g} = \text{minimum} \qquad (11)$$

where

$$\boldsymbol{\Phi}_{ij} = \varphi_j(\mathbf{x}_i), \qquad E\{\mathbf{g}\} = \mathbf{0}, \qquad E\{\mathbf{g}\,\mathbf{g}^T\} = \mathbf{C}_{\mathbf{ff}} \quad . \qquad (12)$$

The solution of (11) is well known from the least squares adjustment theory

$$\hat{\mathbf{a}} = (\boldsymbol{\Phi}^T \mathbf{C}_{\mathbf{ff}}^{-1}\,\boldsymbol{\Phi})^{-1}\,\boldsymbol{\Phi}^T \mathbf{C}_{\mathbf{ff}}^{-1}\,\mathbf{f} \qquad (13)$$

and consequently the estimates to be used in equation (8) are

$$\hat{m} = \boldsymbol{\varphi}(\mathbf{x})^T\,\hat{\mathbf{a}} = \boldsymbol{\varphi}(\mathbf{x})^T\,(\boldsymbol{\Phi}^T \mathbf{C}_{\mathbf{ff}}^{-1}\,\boldsymbol{\Phi})^{-1}\,\boldsymbol{\Phi}^T \mathbf{C}_{\mathbf{ff}}^{-1}\,\mathbf{f} \qquad (14)$$

$$\hat{\mathbf{m}} = \boldsymbol{\Phi}\,\hat{\mathbf{a}} = \boldsymbol{\Phi}\,(\boldsymbol{\Phi}^T \mathbf{C}_{\mathbf{ff}}^{-1}\,\boldsymbol{\Phi})^{-1}\,\boldsymbol{\Phi}^T \mathbf{C}_{\mathbf{ff}}^{-1}\,\mathbf{f} \quad . \qquad (15)$$

3. THE KRIGING SYSTEM

In the Kriging approach conditions (ii) and (iii) of the previous section still hold, but the linearity condition (i) is replaced by

i') $$\hat{f} = \sum_i \lambda_i f_i = \boldsymbol{\lambda}^T \mathbf{f} \quad . \qquad (16)$$

Here b=0 a priori and this is an important difference. In the case of zero means, it follows anyway that $b = 0$, as seen from equation (5) and there is no difference between Kriging and collocation. However, in the case of non zero means the standard prediction according to collocation gives

$$b = m - \mathbf{C}_{\mathbf{f}}\,\mathbf{C}_{\mathbf{ff}}^{-1}\,\mathbf{m} \neq 0 \qquad (17)$$

as can be seen from equation (8).

In kriging where a priori $b = 0$, combination of (i'), (ii) and (iii) gives the so called kriging system

$$\mathbf{C_{ff}}\,\boldsymbol{\lambda} - \mathbf{C}_{\mathbf{f}f} = \alpha\,\mathbf{m} \tag{18}$$

$$\mathbf{m}^{\mathsf{T}}\,\boldsymbol{\lambda} = m \tag{19}$$

or

$$\begin{bmatrix} \mathbf{C_{ff}} & \mathbf{m} \\ \mathbf{m}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ -\alpha \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{\mathbf{f}f} \\ m \end{bmatrix} \tag{20}$$

where α is a Lagrange multiplier.

The solution of the kriging system is

$$\boldsymbol{\lambda} = (\mathbf{m}^{\mathsf{T}}\mathbf{C}_{\mathbf{ff}}^{-1}\,\mathbf{m})^{-1}(m - \mathbf{C}_{\mathbf{f}f}^{\mathsf{T}}\,\mathbf{C}_{\mathbf{ff}}^{-1}\,\mathbf{m})\,\mathbf{C}_{\mathbf{ff}}^{-1}\,\mathbf{m} + \mathbf{C}_{\mathbf{ff}}^{-1}\,\mathbf{C}_{\mathbf{f}f} \tag{21}$$

and replacing into (16) the prediction $\hat{f}$ becomes

$$\hat{f} - \hat{s}\,m = \mathbf{C}_{\mathbf{f}f}^{\mathsf{T}}\,\mathbf{C}_{\mathbf{ff}}^{-1}\,(\mathbf{f} - \hat{s}\,\mathbf{m}) \tag{22}$$

where we have set

$$\hat{s} = (\mathbf{m}^{\mathsf{T}}\,\mathbf{C}_{\mathbf{ff}}^{-1}\,\mathbf{m})^{-1}\,\mathbf{m}^{\mathsf{T}}\,\mathbf{C}_{\mathbf{ff}}^{-1}\,\mathbf{f} \quad . \tag{23}$$

Comparison of equation (22) with equation (8) leads to the following conclusion: When the means are non-zero and known, kriging gives results which do not coincide with the standard prediction technique which in geodesy appears under the name of collocation. In fact, the known mean function $m(\mathbf{x})$ is replaced by a "scaled" version $\hat{s}\,m(\mathbf{x})$, the scale factor $\hat{s}$ being determined by a particular trend removal technique on the available data $\mathbf{f}_i$. The last statement follows from the fact that equation (23) is the solution to the trend removal-least squares problem

$$\mathbf{f} = \mathbf{m}\,s + \mathbf{g}\,, \qquad \mathbf{g}^{\mathsf{T}}\,\mathbf{C}_{\mathbf{ff}}^{-1}\,\mathbf{g} = \text{minimum} \tag{24}$$

However, the stochastic assumption in equation (24) is

$$E\{\mathbf{f}\} = \mathbf{m}\,s \tag{25}$$

in discrepancy with the assumption $E\{\mathbf{f}\} = \mathbf{m}$ used in the derivation of the kriging system.

4. STANDARD KRIGING

Standard kriging follows by adding the assumption that

$$m(\mathbf{x}) = m = \text{constant} \tag{26}$$

with m unknown. The kriging system (20) now becomes

$$\begin{bmatrix} \mathbf{C_{ff}} & \boldsymbol{\delta} \\ \boldsymbol{\delta}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ -\alpha m \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{\mathbf{f}f} \\ 1 \end{bmatrix} \tag{27}$$

where $\boldsymbol{\delta}$ is an $n \times 1$ vector with all elements equal to one. If one starts with conditions (i'), (ii), (iii) he arrives at the system

$$\begin{bmatrix} \mathbf{C_{ff}} & \boldsymbol{\delta} \\ \boldsymbol{\delta}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ -\alpha \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{\mathbf{f}f} \\ 1 \end{bmatrix} \tag{28}$$

which can be shown to be equivalent to (27). The solution of either (27) or (28) is

$$\boldsymbol{\lambda} = \mathbf{C_{ff}^{-1}}\,\mathbf{C}_{\mathbf{f}f} + (\boldsymbol{\delta}^T\,\mathbf{C_{ff}^{-1}}\,\boldsymbol{\delta})^{-1}\,(1 - \boldsymbol{\delta}^T\,\mathbf{C_{ff}^{-1}}\,\mathbf{C}_{\mathbf{f}f})\,\mathbf{C_{ff}^{-1}}\,\boldsymbol{\delta} \tag{30}$$

and replacing into (16) we obtain for the prediction

$$\hat{f} - \hat{m} = \mathbf{C}_{\mathbf{f}f}^T\,\mathbf{C_{ff}^{-1}}\,(\mathbf{f} - \hat{m}\,\boldsymbol{\delta}) \tag{31}$$

where

$$\hat{m} = (\boldsymbol{\delta}^T\,\mathbf{C_{ff}^{-1}}\,\boldsymbol{\delta})^{-1}\,\boldsymbol{\delta}^T\,\mathbf{C_{ff}^{-1}}\,\mathbf{f} \tag{32}$$

is the solution of the trend removal - least squares problem

$$\mathbf{f} = \boldsymbol{\delta}\, m + \mathbf{g} \,, \qquad \mathbf{g}^T \mathbf{C}_{\mathbf{ff}}^{-1}\, \mathbf{g} = \text{minimum} \quad . \tag{33}$$

It follows that standard kriging is equivalent with the collocation prediction with estimated means according to equation (8), (14) and (15) when only one base function is taken in equation (10), namely $\varphi_1(\mathbf{x}) = 1$ and $a_1 = m$.

5. UNIVERSAL KRIGING

Universal kriging follows by adding the assumption that the mean function $m(\mathbf{x})$ has the form of equation (10) with unknown coefficients a_i. Since

$$\mathbf{m} = \boldsymbol{\Phi}\, \mathbf{a} \,, \qquad m = \boldsymbol{\varphi}^T \mathbf{a} \tag{34}$$

the kriging system becomes in this case

$$\mathbf{C}_{\mathbf{ff}}\, \boldsymbol{\lambda} - \alpha\, \boldsymbol{\Phi}\, \mathbf{a} = \mathbf{C}_{\mathbf{f}f} \tag{35}$$

$$\mathbf{a}^T \boldsymbol{\Phi}^T \boldsymbol{\lambda} = \mathbf{a}^T \boldsymbol{\varphi} \quad . \tag{36}$$

The above equation can be replaced by the sufficient condition

$$\boldsymbol{\Phi}^T \boldsymbol{\lambda} = \boldsymbol{\varphi} \tag{37}$$

or

$$\begin{bmatrix} \mathbf{C}_{\mathbf{ff}} & \boldsymbol{\Phi} \\ \boldsymbol{\Phi}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ -\alpha\, \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{\mathbf{f}f} \\ \boldsymbol{\varphi} \end{bmatrix} \tag{38}$$

The solution of (38) is

$$\boldsymbol{\lambda} = \mathbf{C}_{\mathbf{ff}}^{-1}\, \mathbf{C}_{\mathbf{f}f} + \mathbf{C}_{\mathbf{ff}}^{-1}\, \boldsymbol{\Phi}\, (\boldsymbol{\Phi}^T \mathbf{C}_{\mathbf{ff}}^{-1}\, \boldsymbol{\Phi})^{-1}\, (\boldsymbol{\varphi} - \boldsymbol{\Phi}^T \mathbf{C}_{\mathbf{ff}}\, \mathbf{C}_{\mathbf{f}f}) \tag{39}$$

and replacing into (16) we obtain for the prediction

$$\hat{f} - \varphi^T \hat{a} = C_f^T \, C_{ff}^{-1} \, (f - \Phi \hat{a}) \tag{40}$$

where we have set

$$\hat{a} = (\Phi^T C_{ff}^{-1} \Phi)^{-1} \Phi^T C_{ff}^{-1} f \quad . \tag{41}$$

From equation (41) follows that a is the solution of the trend removal - least squares problem

$$f = \Phi a + g \, , \qquad g^T C_{ff}^{-1} g = \text{minimum} \tag{42}$$

Comparison of equations (40), (41) with equations (8), (13) shows that universal kriging gives identical results with the standard prediction (collocation) with estimated means, where the mean function is first estimated from the available data f_i by means of a trend removal technique of the type of equations (10) and (11).

6. CONCLUSIONS

As far as their basic assumptions are concerned, kriging differs from collocation in the form of the linearity condition for the predicted value. Collocation allows an additive constant to the linear combination of the available values.

As far as results of prediction are concerned, standard kriging and universal kriging coincide, under similar assumptions on the form of the unknown mean function, with collocation using estimated means, where the mean function is estimated by a preceding trend removal technique. On the other hand, when the mean function is known, the results of kriging differ from those of collocation. The known means of collocation are replaced by scaled means where a scale factor is estimated by a preceding trend removal technique. This difference between kriging and collocation is a direct consequence of the difference in the linearity conditions mentioned above.

The above comparisons are independent of any assumptions about the nature of the covariance function $c(x,x')$ of the stochastic process $f(x)$. In collocation it is usually, but not necessarily, assumed that the covariance function is homogeneous and isotropic, i.e.,

$$c(\mathbf{x},\mathbf{x}') = c(\| \mathbf{x} - \mathbf{x}' \|) \tag{43}$$

In the so called transitive kriging a homogeneous but not isotropic covariance function is assumed, i.e.,

$$c(\mathbf{x}, \mathbf{x}') = c(\mathbf{x}' - \mathbf{x}) \quad . \tag{44}$$

The non-linear extensions of the kriging method have not been considered since they cannot be compared with the linear method of collocation.

REFERENCES

Blais, J.A.R. (1982): "Synthesis of Kriging Estimation Methods", *Manuscripta Geodaetica*, vol. 7, no. 4, pp. 325-353.

Address:

Prof. Dr. Athanasios Dermanis

Department of Geodesy and Surveying
University of Thessaloniki
University Box 503
540 06 Thessaloniki, GREECE