Mathematical Geology, Vol. 29, No. 5, 1997

A New Form of the Cokriging Equations[1]

Andrew E. Long[2] and Donald E. Myers[2]

Myers developed a matrix form of the cokriging equations, but one that entails the solution of a large system of linear equations. Large systems are troublesome because of memory requirements and a general increase in the matrix condition number. We transform Myers's system into a set of smaller systems, whose solution gives the classical kriging results, and provides simultaneously a nested set of lower dimensional cokriging results. In the course of developing the new formulation we make an interesting link to the Cauchy-Schwarz condition for the invertibility of a system, and another to a simple situation of coregionalization. In addition, we proceed from these new equations to a linear approximation to the cokriging results in the event that the cross-variograms are small, allowing one to take advantage of a recent results of Xie and others which proceeds by diagonalizing the variogram matrix function over the lag classes.

KEY WORDS: cokriging, condition number, SVD.

INTRODUCTION

The (univariate) universal cokriging estimator for the intrinsic vector-valued random function $\mathbf{z}$ is given by the equations

$$\mathbf{z}^*(x_0) = \sum_{i=1}^{N} \Gamma_i^T z(x_i)$$

where the weight matrices Γ_i satisfy the conditions

$$\sum_{i=1}^{N} F_l(x_i)\Gamma_i = F_l(x_0), \qquad l = 1, \ldots, p \tag{1}$$

where the p matrices F_l are given by

$$F_l(x) = f_l(x) * I$$

[1]Received 23 January 1996; revised 11 November 1996.

[2]Department of Mathematics, University of Arizona, Tucson, Arizona 85721. e-mail: ael@maya.medctr.luc.edu or myers@math.arizona.edu

and where the $f_l(x)$ are independent functions forming a basis for the drift surface (Myers, 1982). The weight matrices are determined by the $N + p$ sets of equations given by the constraints (1), and the sets of linear equations

$$\sum_{i=1}^{N} V(x_i - x_j)\Gamma_j + \sum_{l=1}^{p} F_l(x_i)\mu_l = V(x_i - x),$$
$$i = 1, \ldots, N \tag{2}$$

V is the variogram matrix function, and the μ_l are matrices of Lagrange multipliers. Note that this is precisely the form of the universal kriging equations, where scalar quantities have been replaced by matrices.

The practitioner's role is to model the variograms and cross-variograms, select the degree of drift (in the situation of polynomial drift, or the type of drift functions otherwise), and to determine the size and degree of anisotropy of the cokriging neighborhoods. The estimates then are given by the classical equations as given, formulated as a matrix system by Myers (1982, 1992).

MOTIVATION: WHY A NEW FORMULATION?

The issues that dictated a new form for the cokriging equations were primarily issues of linear algebra. How could we reduce the size of the matrices involved? How could we improve the condition numbers of our systems?

The most important single element in the solution of the cokriging equations is the matrix solver. Carr and Myers (1990) discussed different equation solvers for cokriging programs, and decided (at that time) on Gaussian elimination. In their first cokriging code, Carr, Myers, and Glass (1985) used a slower, iterative algorithm which minimized memory use (which was more important at that time than it is now). The program "cokrige," which was developed by the Geostatistics Group of the Mathematics Department, University of Arizona, for adaptation into the Geo-EAS pantheon of programs (but never formally incorporated into the Geo-EAS package), used Gaussian elimination. (It is available from the authors.)

Early in a recent study of our own, we encountered trouble when generating maps with "cokrige:" it seems that we were using "too many variables" or "too many sites," which led to estimates which were obviously poor (e.g., estimates many orders of magnitudes higher or lower than any data values); yet there was no indication from the program of any problem. Gaussian elimination, although a good method in many instances, was a poor selection for us because of the dangers posed by both the size and the conditioning of the matrices we needed to (formally) invert. Because the modeling process is understood poorly, and the risk of creating large ill-conditioned matrices sufficiently high, we were inspired to write a new program, selecting another and safer algorithm for the matrix inversion: the SVD, in double precision.

McCarn and Carr (1992) compare Gaussian elimination, LU decomposition, and, to a lesser extent, the SVD, in the computation of the kriging weights, as well as the effect of numerical precision used and the advantages of iterative improvement. They give the number of operations for the three methods (Table 1). They also discuss the value of using only a small number of neighbors, to reduce round-off error, suggesting 10–20 neighbors for local neighborhoods.

They note that the SVD gives results identical to those using Gaussian elimination or LU decomposition for ordinary kriging, but state that "for universal kriging . . . there is a large difference in the solution yielded by SVD from that yielded by either Gauss elimination or LU decomposition." This is evidently the result of a failure to address an issue of scaling in their code, that is, the functions used to model the drift were not scaled properly. Note that the two parts of the cokriging matrix in (3),

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

are independent: scaling the variables related to V does not affect F, and *vice versa.* If rows and columns corresponding to F, say, are allowed to get larger than the V portion of the matrix, the condition number will increase artificially (in the sense that scaling would have prevented any problems). This could happen if the functions used were simple monomials (like xy), and the geographical coordinates were orders of magnitude larger than the variogram values contained in V. The drift functions and the variogram matrix values should be scaled so as to be on the same order of magnitude.

This scaling problem is the same as that identified (but not pursued) by O'Dowd (1991), when he reported that the condition number of the ordinary kriging system went up with a linear increase in the sills of the variogram

Table 1. Operation Counts for Different Equation Solvers (McCarn and Carr, 1992)

Gaussian elimination	LU decomposition	SV decomposition
$\text{ops} = \dfrac{2N^3}{3}$	$\dfrac{N^3}{3} \le \text{ops} \le \dfrac{2N^3}{3}$	$\text{ops} > N^3$

models. This is simply a result of having a column (and row) of fixed values (ones) in the F portion of (3), while the V portion is scaled linearly. Poor conditioning in this example is not a fundamental characteristic of the kriging system, as it can be removed by scaling.

One advantage of using the SVD as a solver is that the condition number of the coefficient matrix shows up immediately as the ratio of the largest and smallest singular values: if A is $N \times N$, then

$$\text{Condition}(A) = \begin{cases} \dfrac{\lambda_1}{\lambda_N}, & \lambda_N \neq 0 \\ \infty & \lambda_N = 0 \end{cases}$$

The condition number should be reported, especially when it is high, because it serves as a handy diagnostic to indicate whether the results may be useful. If the coefficient matrix is noninvertible to machine precision, then the option should be given to proceed with the pseudo-inverse (which is obtained from the SVD, and leads to a least-squares solution for the projection of the right-hand side onto the residual column space of the matrix).

THE NEW FORMULATION IN THE TWO VARIABLE CASE

First consider the case of cokriging with two variables. Myers's formulation, that is the system of size $2(N + p) \times 2(N + p)$, is given explicitly in this case by

$$\begin{bmatrix} V & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} \Gamma \\ \mu \end{bmatrix} = \begin{bmatrix} V_0 \\ F_0 \end{bmatrix} \tag{3}$$

where the elements of V are the block variogram matrices (which, at the risk of confusion, also will be termed V) made up of the variograms and cross-variogram of the two variables for each pair of data locations, and F is the matrix function of p linearly independent F_l matrix functions (whose coefficients are determined in the cokriging process). V is only conditionally positive definite (and hence not invertible, in general) (Myers, 1992).

On the right-hand side is the "column matrix" of variogram matrices referenced to the site x_0, the location at which the estimate is desired; and similarly for F_x. This is represented by

$$\begin{bmatrix} V(x_1 - x_1) & V(x_1 - x_2) & \cdots & V(x_1 - x_N) & F_1(x_1) & \cdots & F_p(x_1) \\ V(x_2 - x_1) & V(x_2 - x_2) & \cdots & V(x_2 - x_N) & F_1(x_2) & \cdots & F_p(x_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ V(x_N - x_1) & V(x_N - x_2) & \cdots & V(x_N - x_N) & F_1(x_N) & \cdots & F_p(x_N) \\ F_1(x_1) & F(x_2) & \cdots & F_1(x_N) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ F_p(x_1) & F_p(x_2) & \cdots & F_p(x_N) & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_N \\ \mu_1 \\ \vdots \\ \mu_p \end{bmatrix} = \cdot \begin{bmatrix} V(x_1 - x_0) \\ V(x_2 - x_0) \\ \vdots \\ V(x_N - x_0) \\ F_1(x_0) \\ \vdots \\ F_p(x_0) \end{bmatrix} . \tag{4}$$

where the subscripts refer to the data points determining the distance used by the matrix variogram function.

The trick to transforming these into the new formulation was discovered by inspecting the cokriging system in the situation of zero cross-variograms. It is clear in this special case that the columns and rows can be permuted so as to separate the system into two kriging systems (which would be better solved separately, from the standpoint of memory, efficiency, and matrix condition; furthermore, the inversion could be carried out in parallel).

Thus, we simply permute the rows and columns of this large matrix (4) so that the variograms (diagonal elements of the block matrices of V) and cross-variogram (off-diagonal elements) get separated. For two variable cokriging, define a permutation matrix P such that

$$P \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

or

$$P \equiv [\mathbf{e}_1 \quad \mathbf{e}_{N+p+1} \quad \mathbf{e}_2 \quad \mathbf{e}_{N+p+2} \quad \cdots \quad \mathbf{e}_{N+p} \quad \mathbf{e}_{2(N+p)}]$$

where $\mathbf{e}_i$ is the Euclidean unit vector with 1 in the ith place, and zeros elsewhere. The generalization is obvious for other numbers of variables, and the result is the same, in the sense that the variables are separated similarly.

Define

$$X \equiv P \begin{bmatrix} V & F \\ F^T & 0 \end{bmatrix} P^T \equiv \begin{bmatrix} K_1 & C \\ C & K_2 \end{bmatrix}$$

where K_1 and K_2 represent the coefficient matrices of the kriging systems for the two variables, and C represents the cross-variogram matrix given by off-diagonal terms of the variogram matrix model. The inverse of the matrix X is given simply in terms of the matrix inverses of K_1 and K_2 (which are needed to get the kriging results) and the matrix inverses of two other $(N + p) \times (N + p)$ matrices:

$$M_1 \equiv I - K_1^{-1} C K_2^{-1} C \tag{5}$$

and

$$M_2 \equiv I - K_2^{-1} C K_1^{-1} C \tag{6}$$

The form of these matrices and their consequences of their invertibility suggest a link to the Cauchy–Schwartz condition, which, for a pair of variables, is

$$\sigma_{12}^2 \leq \sigma_1^2 \sigma_2^2$$

where σ_{12} is the covariance of the two, and on the right-hand side are the variances σ_1^2 and σ_2^2. Rewrite that as

$$m_1 \equiv 1 - (\sigma_1^2)^{-1}(\sigma_{12})(\sigma_2^2)^{-1}(\sigma_{12})$$

with the condition that

$$m_1 \geq 0$$

Comparing m_1 and (5) shows that the kriging matrices are playing the roles of the variances (appropriately enough, as the variogram is the decomposition of the variance) and the cross-variogram matrix is playing the role of the covariance. The Cauchy–Schwartz condition, reflected in the inequality guarantees strict positive definiteness in a 2×2 matrix, which guarantees unique solvability of the system. How is the inequality reflected in this matrix case?

M_1 is noninvertible if and only if there exists nonzero $\mathbf{x}$ such that

$$M_1 \mathbf{x} = \mathbf{0}, \quad \text{if and only if } K_1^{-1} C K_2^{-1} C\mathbf{x} = \mathbf{x}$$

This does not happen if

$$1 - \|K_1^{-1} C K_2^{-1} C\|_2 \geq 0$$

and similarly for the case of (6). Hence, the coefficient matrix is invertible (provided the kriging systems are invertible) only if the largest singular values of the matrices $K_1^{-1}CK_2^{-1}C$ and $K_2^{-1}CK_1^{-1}C$ are less than 1.

One can gain some appreciation for this by starting with two independent variables, where C is zero; then the two matrices $K_1^{-1}CK_2^{-1}C$ and $K_2^{-1}CK_1^{-1}C$ also are zero. Now, as correlation is "increased," via the cross-variogram, resulting in an increase in the matrix norm of C, the singular values of $K_1^{-1}CK_2^{-1}C$ and $K_2^{-1}CK_1^{-1}C$ move continuously on the real line, out from zero (the degenerate "singular value" of the zero matrix). At some point, the largest singular value (and hence the norm of these matrices) may increase beyond 1, at which time the system will no longer be invertible for all right-hand sides.

If the kriging matrices and M_1 and M_2 are invertible, then inverting X is easy:

$$\begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} K_1 & C \\ C & K_2 \end{bmatrix} = \begin{bmatrix} I & K_1^{-1}C \\ K_2^{-1}C & I \end{bmatrix}$$

and

$$\begin{bmatrix} I & -K_1^{-1}C \\ -K_2^{-1}C & I \end{bmatrix} \begin{bmatrix} I & K_1^{-1}C \\ K_2^{-1}C & I \end{bmatrix} = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}$$

Let

$$A_1 \equiv K_1^{-1}C \quad \text{and} \quad A_2 \equiv K_2^{-1}C \tag{7}$$

then

$$X^{-1} = \begin{bmatrix} M_1^{-1} & 0 \\ 0 & M_2^{-1} \end{bmatrix} \begin{bmatrix} I & -A_1 \\ -A_2 & I \end{bmatrix} \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix}$$

COMPUTATION OF ESTIMATES

The kriging estimates are

$$\mathbf{z}_k^*(x_0) = \begin{bmatrix} V_{10}^T & 0 \\ 0 & V_{20}^T \end{bmatrix} \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

whereas the cokriging estimates are given by

$$\mathbf{z}_c^*(x_0) = \begin{bmatrix} V_{10}^T & C_0^T \\ C_0^T & V_{20}^T \end{bmatrix} \begin{bmatrix} M_1^{-1} & 0 \\ 0 & M_2^{-1} \end{bmatrix} \begin{bmatrix} I & -A_1 \\ -A_2 & I \end{bmatrix} \cdot \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

In either situation, one must compute

$$\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \equiv \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

Then the kriging results are given by the simple dot products

$$\mathbf{z}_k^*(x_0) = \begin{bmatrix} V_{10}^T & 0 \\ 0 & V_{20}^T \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} V_{10}^T \delta_1 \\ V_{20}^T \delta_2 \end{bmatrix}$$

and the cokriging estimates can be simplified further:

$$\begin{aligned} \mathbf{z}_c^*(x_0) &= \begin{bmatrix} V_{10}^T & C_0^T \\ C_0^T & V_{20}^T \end{bmatrix} \begin{bmatrix} M_1^{-1} & 0 \\ 0 & M_2^{-1} \end{bmatrix} \begin{bmatrix} I & -A_1 \\ -A_2 & I \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \\ &= \begin{bmatrix} V_{10}^T & C_0^T \\ C_0^T & V_{20}^T \end{bmatrix} \begin{bmatrix} B_1 & -B_1 A_1 \\ -B_2 A_2 & B_2 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \end{aligned}$$

where

$$B_1 \equiv M_1^{-1} \quad \text{and} \quad B_2 \equiv M_2^{-1}$$

All this can be stored in the same size matrix as originally given, once the matrix products have been computed.

That is followed by one last multiplication, so that, in the end, cokriging at a particular site will take twice the computation that kriging requires:

$$\mathbf{z}_c^*(x_0) = \begin{bmatrix} V_{10}^T & C_0^T \\ C_0^T & V_{20}^T \end{bmatrix} \begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix}$$

in a form entirely analogous to that of the kriging estimates, with

$$\begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix} \equiv \begin{bmatrix} B_1 & -B_1 A_1 \\ -B_2 A_2 & B_2 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$$

One of the advantages of this scheme is that the inversion of the $2(N + p) \times 2(N + p)$ matrix is replaced by the inversion of four $(N + p) \times (N + p)$ matrix [the "Cauchy-Schwartz" matrices M_1 and M_2 in addition to the $(N + p) \times (N + p)$ kriging matrices]. The kriging matrices may be calculated anyway, however, in order to compare the kriging and cokriging results; so this is a savings.

This also suggests that experimenting with various cross-variograms is easier than before: the kriging systems are solved once and for all, and then several cross-variograms of interest can be tried for comparison, with much less additional computation than before.

We give a Matlab example in the Numerical Results section, demonstrating that cokriging using Myers's approach involves the inversion of a large matrix with high condition number, whereas the procedure described here requires the inversion of the kriging systems (presumably necessary anyway), which may have condition numbers on the same order, but somewhat smaller, followed by the inversion of two matrices with small (approximately equal to 1) condition numbers.

FIRST-ORDER APPROXIMATION OF COKRIGING IMPROVEMENT

In this section we develop a first-order approximation to the cokriging results which is valid when cross-variograms are relatively small. This may have important consequences when combined with a recent technique of Xie and Myers (1995) and Xie, Myers, and Long (1995); they attempt to diagonalize the variogram matrix, effectively making a change of variables so as to minimize the cross-variogram terms. This has one consequence they mentioned, namely making the cross-variograms easier to model; and, in addition, it makes the system appropriate for the following linear approximation (which is useful with or without their technique).

If $\|A_1A_2\|$ and $\|A_2A_1\|$ are each much less than 1, then

$$B_1 = M_1^{-1} = (I - A_1A_2)^{-1} \approx I + A_1A_2$$

and similarly for B_2. Then

$$\begin{aligned} \mathbf{z}_c^*(x_0) \approx & \begin{bmatrix} V_{10}^T & C_0^T \\ C_0^T & V_0^T \end{bmatrix} \\ & \cdot \begin{bmatrix} I + A_1A_2 & 0 \\ 0 & I + A_2A_1 \end{bmatrix} \begin{bmatrix} I & -A_1 \\ -A_2 & I \end{bmatrix} \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \\ = & \begin{bmatrix} V_{10}^T & C_0^T \\ C_0^T & V_{20}^T \end{bmatrix} \left(\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} + \begin{bmatrix} A_1A_2 & 0 \\ 0 & A_2A_1 \end{bmatrix} \right) \\ & \cdot \left(\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} - \begin{bmatrix} 0 & A_1 \\ A_2 & 0 \end{bmatrix} \right) \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \end{aligned}$$

$$= \left(\begin{bmatrix} V_{10}^T & 0 \\ 0 & V_{20}^T \end{bmatrix} + \begin{bmatrix} 0 & C_0^T \\ C_0^T & 0 \end{bmatrix} \right) \left(\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \right.$$

$$\left. + \begin{bmatrix} A_1A_2 & -A_1(I - A_2A_1) \\ -A_2(I - A_1A_2) & A_2A_1 \end{bmatrix} \right) \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

$$= \mathbf{z}_k^*(x_0) + \left(\begin{bmatrix} V_{10}^T & 0 \\ 0 & V_{20}^T \end{bmatrix} + \begin{bmatrix} 0 & C_0^T \\ C_0^T & 0 \end{bmatrix} \right)$$

$$\cdot \begin{bmatrix} A_1A_2 & -A_1(I - A_2A_1) \\ -A_2(I - A_1A_2) & A_2A_1 \end{bmatrix}$$

$$\cdot \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} + \begin{bmatrix} 0 & C_0^T \\ C_0^T & 0 \end{bmatrix} \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

$$= \mathbf{z}_k^*(x_0) + \begin{bmatrix} V_{10}^T & C_0^T \\ C_0^T & V_{20}^T \end{bmatrix} \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix}$$

$$\cdot \begin{bmatrix} CK_2^{-1}C & -C(I - K_2^{-1}CK_1^{-1}C) \\ -C(I - K_1^{-1}CK_2^{-1}C) & CK_1^{-1}C \end{bmatrix}$$

$$\cdot \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} + \begin{bmatrix} 0 & C_0^T \\ C_0^T & 0 \end{bmatrix} \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

$$= \mathbf{z}_k^*(x_0) + \begin{bmatrix} C_0^TK_2^{-1}\,d_2 \\ C_0^TK_1^{-1}\,d_1 \end{bmatrix} + \begin{bmatrix} V_{10}^T & C_0^T \\ C_0^T & V_{20}^T \end{bmatrix}$$

$$\cdot \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} CK_2^{-1}C & -C(I - K_2^{-1}CK_1^{-1}C) \\ -C(I - K_1^{-1}CK_2^{-1}C) & CK_1^{-1}C \end{bmatrix}$$

$$\cdot \begin{bmatrix} K_1^{-1} & 0 \\ 0 & K_2^{-1} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

Keeping only the terms linear in C (or C_0),

$$\mathbf{z}_c^*(x_0) \approx \mathbf{z}_k^*(x_0) - \begin{bmatrix} d_2^TK_2^{-1}(CK_1^{-1}V_{10} - C_0) \\ d_1^TK_1^{-1}(CK_2^{-1}V_{20} - C_0) \end{bmatrix}$$

or, recalling that the kriging weights are $\Gamma_i = K_i^{-1}V_{i0}$,

$$\mathbf{z}_c^*(x_0) \approx \mathbf{z}_k^*(x_0) - \begin{bmatrix} (d_2)^T K_2^{-1}(C\Gamma_1 - C_0) \\ (d_1)^T K_1^{-1}(C\Gamma_2 - C_0) \end{bmatrix}$$

Making use of the transformed data $(d_i')^T = (d_i)^T K_i^{-1}$,

$$\mathbf{z}_c^*(x_0) \approx \mathbf{z}_k^*(x_0) - \begin{bmatrix} (d_2')^T (CK_1^{-1}V_{10} - C_0) \\ (d_1')^T (CK_2^{-1}V_{20} - C_0) \end{bmatrix}$$

$$\equiv \mathbf{z}_k^*(x_0) - \begin{bmatrix} (d_2'')^T V_{10} - (d_2')^T C_0 \\ (d_1'')^T V_{20} - (d_1')^T C_0 \end{bmatrix}$$

Thus, a calculation of the cokriging approximation requires storing another form of transformed data, but only two vector inner-products for an actual estimate of the cokriging results (with no additional matrix inversions).

This approximation implies that to first order it is the extent to which $C\Gamma_i$ differ from C_0 that determines whether it is worthwhile to cokrige. If

$$C\Gamma_1 = C_0 \quad \text{and} \quad C\Gamma_2 = C_0$$

that is, if

$$CK_1^{-1}V_{10} = C_0 \quad \text{and} \quad CK_2^{-1}V_{20} = C_0$$

then cokriging may provide no improvement. Although it will be interesting to consider under what conditions these hold, we have not yet done so.

THE ELEMENTAL COREGIONALIZATION CASE

The new formulation of the cokriging equations also gives some insight into a simple form of coregionalization. Start with the form of the variogram matrix in the case of a one-structure "coregionalization" (there are quotes around coregionalization because this is "trivially" coregionalized: there is only a single structure):

$$V(h) = \gamma(h) \begin{bmatrix} a & c \\ c & b \end{bmatrix} \equiv \gamma(h) V$$

where the matrix V on the right-hand side is nonnegative definite, and γ is a standard variogram model (conditionally negative definite function). Variables with this type of model are said to be "intrinsically coregionalized" (Helterbrand and Cressie, 1994).

A striking result made obvious by this new formulation is that cokriging variables modeled as the elemental constituent of the coregionalization case gives the same result as kriging. Matheron (1965) termed this "intrinsic correlation," and also showed that cokriging reduces to kriging (Matheron, 1979).

The form of Myers's equations in this special case is

$$\begin{bmatrix} 0 & \gamma(x_1 - x_2)V & \gamma(x_1 - x_3)V & \cdots & \gamma(x_1 - x_N)V & F_1 \\ \gamma(x_2 - x_1)V & 0 & \gamma(x_2 - x_3)V & \cdots & \gamma(x_2 - x_N)V & F_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(x_N - x_1)V & \gamma(x_N - x_2)V & \gamma(x_N - x_3)V & \cdots & 0 & F_N \\ F_1 & F_2 & F_3 & \cdots & F_N & 0 \end{bmatrix}$$

which one can permute to

$$\begin{bmatrix} a\begin{bmatrix} K & F \\ F^T & 0 \end{bmatrix} & c\begin{bmatrix} K & 0 \\ 0 & 0 \end{bmatrix} \\ c\begin{bmatrix} K & 0 \\ 0 & 0 \end{bmatrix} & b\begin{bmatrix} K & F \\ F^T & 0 \end{bmatrix} \end{bmatrix}$$

whereas the swapped form of the equation is

$$\begin{bmatrix} aK & cK & aF & 0 \\ cK & bK & 0 & bF \\ aF^T & 0 & 0 & 0 \\ 0 & bF^T & 0 & 0 \end{bmatrix} \tag{8}$$

Let

$$\Delta \equiv ab - c^2$$

To invert (8), apply the following operations on the left:

$$\begin{bmatrix} K^{-1} & 0 & 0 & 0 \\ 0 & K^{-1} & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix}$$

$$\begin{bmatrix} \frac{b}{\Delta} I & \frac{-c}{\Delta} I & 0 & 0 \\ \frac{-c}{\Delta} I & \frac{a}{\Delta} I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix}$$

$$\begin{bmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ -aF^T & 0 & I & 0 \\ 0 & -bF^T & 0 & I \end{bmatrix}$$

$$\begin{bmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & \frac{-\Delta}{ab} (F^T K^{-1} F)^{-1} & 0 \\ 0 & 0 & 0 & \frac{-\Delta}{ab} (F^T K^{-1} F)^{-1} \end{bmatrix}$$

$$\begin{bmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & \frac{b}{\Delta} I & \frac{c}{\Delta} I \\ 0 & 0 & \frac{c}{\Delta} I & \frac{a}{\Delta} I \end{bmatrix}$$

$$\begin{bmatrix} I & 0 & \frac{-ab}{\Delta} K^{-1}F & \frac{cb}{\Delta} K^{-1}F \\ 0 & I & \frac{ca}{\Delta} K^{-1}F & \frac{-ab}{\Delta} K^{-1}F \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix}$$

which gives the identity matrix; then the inverse is

$$\begin{bmatrix} \frac{b}{\Delta}(K^{-1}-M) & \frac{-c}{\Delta}(K^{-1}-M) & \frac{1}{a}K^{-1}FD & 0 \\ \frac{-c}{\Delta}(K^{-1}-M) & \frac{a}{\Delta}(K^{-1}-M) & 0 & \frac{1}{b}K^{-1}FD \\ \frac{1}{a}(K^{-1}FD)^T & 0 & \frac{-1}{a}D & \frac{-c}{ab}D \\ 0 & \frac{1}{b}(K^{-1}FD)^T & \frac{-c}{ab}D & \frac{-1}{b}D \end{bmatrix}$$

where

$$D = D^T \equiv (F^T K^{-1} F)^{-1}$$

and

$$M = M^T \equiv K^{-1}FDF^TK^{-1}$$

(both are nonnegative definite, at least).

Now consider the estimates, which is where this case gets interesting (or rather, so uninteresting, as the cokriging results reduce to the kriging results!):

$$\begin{bmatrix} \Gamma_{c1} & \gamma_2 \\ \gamma_1 & \Gamma_{c1} \\ \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix} = \begin{bmatrix} \frac{b}{\Delta}(K^{-1}-M) & \frac{-c}{\Delta}(K^{-1}-M) & \frac{1}{a}K^{-1}FD & 0 \\ \frac{-c}{\Delta}(K^{-1}-M) & \frac{a}{\Delta}(K^{-1}-M) & 0 & \frac{1}{b}K^{-1}FD \\ \frac{1}{a}(K^{-1}FD)^T & 0 & \frac{-1}{a}D & \frac{-c}{ab}D \\ 0 & \frac{1}{b}(K^{-1}FD)^T & \frac{-c}{ab}D & \frac{-1}{b}D \end{bmatrix} \cdot \begin{bmatrix} aK_0 & cK_0 \\ cK_0 & bK_0 \\ aF_0 & 0 \\ 0 & bF_0 \end{bmatrix}$$

This reduces to exactly the kriging weights, for example,

$$[\Gamma_{c1} \quad \gamma_2] = \left[\frac{ab - c^2}{\Delta}(K^{-1} - M)K_0 + K^{-1}FDF_0 \right.$$
$$\left. \cdot \frac{-bc}{\Delta}(K^{-1} - M)(K_0 - K_0)\right]$$

becomes

$$\begin{aligned}[\Gamma_{c1} \quad \gamma_2] &= [(K^{-1} - M)K_0 + K^{-1}FDF_0 \quad 0] \\ &= [K^{-1}((I - FDF^TK^{-1})K_0 + FDF_0) \quad 0] \\ &= [K^{-1}K_0 - K^{-1}FDF^TK^{-1}K_0 + K^{-1}FDF_0 \quad 0] \qquad (9)\end{aligned}$$

Note the first term in the vector right-hand side: it is the kriging weight, which is seen by solving just one block of the cokriging system:

$$\Gamma_{c1} = K^{-1}K_0 - K^{-1}F\mu$$

where

$$\mu = -D(F_0 - FK^{-1}K_0)$$

which gives a result identical to the first element of (9).

Thus, there is absolutely no change in the estimates by cokriging in this case, as the weights do not change. That is especially interesting and important because one method proposed for determining a valid model for the cross-variogram of two variables is to use a model which is a nested combination of models of the variograms: if the variograms have the same models (type and sill), however, the situation reduces directly to this case, and one sees immediately that cokriging need not be attempted at all.

One can reach the same conclusion (with a lot less calculation!) via an argument about the form of the variogram matrix function: recall that the variogram estimator can be written as:

$$V^*(\mathbf{h}) = \frac{1}{2N_h}D^T(\mathbf{h})D(\mathbf{h})$$

where D is the dataset of paired differences. In this simple case that indicates that

$$\frac{1}{2N_h}D^T(\mathbf{h})D(\mathbf{h}) = \gamma(\mathbf{h})\begin{bmatrix} a & c \\ c & b \end{bmatrix} = \gamma(\mathbf{h})Q\Lambda Q^T$$

This says that by merely transforming the pair difference data via

$$D' = DQ$$

(which is equivalent to the same transformation on the original data, as the D are just linear combinations of the original data vectors), the sample variogram matrix will have been diagonalized, that is the sample variogram matrix for the transformed data will have the form

$$V'^{*}(\mathbf{h}) = \gamma(\mathbf{h})\Lambda$$

Oddly enough, Helterbrand and Cressie (1994) report differences in estimates in the situation of an intrinsically coregionalized cokriging, citing similar claims in the Summer 1992 issue of *Geostatistics: An interdisciplinary Geostatistics Newsletter* (available from the authors).

GENERALIZATION: MORE THAN TWO VARIABLES

The new formulation of the cokriging equations generalizes, but not elegantly. For example, in the three variable situation, one may permute as before and multiply through by the kriging system matrix inverses to get

$$\begin{bmatrix} I & A_{12} & A_{13} \\ A_{21} & I & A_{23} \\ A_{31} & A_{32} & I \end{bmatrix}$$

Multiplying through by the inverse in the first two variables, as given,

$$\begin{bmatrix} B_1 & -B_1A_{12} & 0 \\ -B_2A_{21} & B_2 & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} I & A_{12} & A_{13} \\ A_{21} & I & A_{23} \\ A_{31} & A_{32} & I \end{bmatrix} = \begin{bmatrix} I & 0 & \alpha_1 \\ 0 & I & \alpha_2 \\ A_{31} & A_{32} & I \end{bmatrix}$$

followed by

$$\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -A_{31} & -A_{32} & I \end{bmatrix} \begin{bmatrix} I & 0 & \alpha_1 \\ 0 & I & \alpha_2 \\ A_{31} & A_{32} & I \end{bmatrix} = \begin{bmatrix} I & 0 & \alpha_1 \\ 0 & I & \alpha_1 \\ 0 & 0 & I - A_{31}\alpha_1 - A_{32}\alpha_2 \end{bmatrix}$$

so that one must invert the matrix $I - A_{31}\alpha_1 - A_{32}\alpha_2$, which is again $(N + p) \times (N + p)$.

Induction on this process implies that in each situation one gets the solution of a single chain $(1, \ldots, k - 1, k\text{-way})$ of cokriging equations, from the kriging example all the way up to the k-way situation, each of which usually will be of interest.

Table 2. Comparison of (Co)-Kriging Weights

Cokriging	System	Kriging	System	First-order	System
−1.1323	−0.0228	−1.1205	0	−1.1205	−0.0225
0.2238	−0.0039	0.2559	0	0.2259	−0.0039
1.9085	0.0267	1.8946	0	1.8946	0.0264
0.0002	0.0001	−0.0001	0	−0.0001	0.0001
0.0114	0.0055	0	0.0052	0.0113	0.0052
0.0020	0.4208	0	0.4208	0.0019	0.4208
−0.0133	0.5737	0	0.5740	−0.0132	0.5740
0.0001	0.0102	0	0.0102	0.0001	0.0102

NUMERICAL RESULTS

The following results are taken from a Matlab script (available from the authors) that demonstrates the ideas in this paper. We computed the cokriging weights for a small example case comparing Myers's system and our new formulation. The cokriging weights obtained from both systems were exactly the same. In the following table (Table 2) we also give the kriging weights for this small sample system, followed by the weights obtained using the linear approximation (we used small cross-variograms so that we could demonstrate this approximation in a situation where its use is appropriate).

In Table 3, we show that the similar matrices were actually better conditioned than the matrix (C) used in Myers's system:

Table 3. Better Conditioned Small Matrices

Matrix	Condition number
C	3578
K_1	3267
K_2	35.3597
$I - AB$	1.0181
$I - BA$	1.0181

CONCLUSIONS

This new formulation of the cokriging equations has the following features, which we feel justify its study and use:

- simultaneous kriging estimates;
- a chain of (sub-)cokriging estimates;

- ease of comparison of, and experimentation with, different cross-variogram models;
- a linear approximation to cokriging, requiring no additional matrix inversions; and
- smaller and better conditioned systems of equations in intermediate steps.

We have not discovered yet the ideal cokriging method, which would permit the solution of a k - way cokriging system by giving the results of *all* $(k - 1)$-way, $(k - 2)$-way, . . . , and 1-way (kriging) systems as well. In that situation, one might simply cokrige all variables, and, based on cross-validation results of each subset of cokrigings, select that combination which does best according to some *a priori* criterion. Although short of that goal, this new formulation leads to one set of $k - 1$, $k - 2$, . . . , $k-$way, and all kriging solutions in the process of cokriging a set of k variables. If the goal were, say, the estimation of the concentration of nitrates (n_1) in groundwater, and there were $k - 1$ other variables (such as sodium and chloride) which one suspected might help improve the estimates of nitrate via cokriging, then one could order them as $n_2, n_3, \ldots, n_k$ and cokrige with this method so as to get the results of

- kriging for nitrate (as well as kriging for each of the n_i);
- cokriging for nitrate with sodium;
- cokriging for nitrate with sodium and chloride;

⋮

- and cokriging for nitrate with sodium, and all other variables.

REFERENCES

Carr, J. R., and Myers, D. E., 1990, Efficiency of different equation solvers in cokriging: Computers & Geosciences, v. 16, no. 5, p. 705–716.

Carr, J. R., Myers, D. E., and Glass, C. E., 1985. Cokriging—a computer program: Computers & Geosciences, v. 11, no. 2, p. 111–127.

Helterbrand, J. D., and Cressie, N., 1994, Universal cokriging under intrinsic coregionalization: Math. Geology, v. 26, no. 2, p. 197–204.

Matheron, G., 1965, Les Variables Régionalisées et Leur Estimation: Masson et Cie, Paris, 306 p.

Matheron, G., 1979, Recherche de simplification dans un probleme de cokrigeage, N-628: Centre de Géostatistique, Fontainebleau, unpaginated.

McCarn, D. W., and Carr, J. R., 1992, Influence of numerical precision and equation solution algorithm on computation of kriging weights: Computers & Geosciences, v. 18, no. 9, p. 1127–1167.

Myers, D. E., 1982, Matrix formulation of co-kriging: Math. Geology, v. 14, no. 3, p. 249–257.

Myers, D. E., 1992, Kriging, cokriging, radial basis functions and the role of positive definiteness: Computer Math. Applications, v. 24, no. 12, p. 139–148.

O'Dowd, R. J., 1991, Conditioning of coefficient matrices of ordinary kriging: Math. Geology, v. 23, no. 5, p. 721–740.

Xie, T., and Myers, D. E., 1995, Fitting matrix valued variogram models by simultaneous diagonalization (part I: Theory): Math. Geology, v. 27, no. 7, p. 867–875.

Xie, T., Myers, D. E., and Long, A. E., 1995, Fitting matrix valued variogram models by simultaneous diagonalization (part II: Application): Math. Geology, v. 27, no. 7, p. 877–888.