Mathematical Geology, Vol. 16, No. 8, 1984

Robustness of Variograms and Conditioning of Kriging Matrices[1]

Phil Diamond[2] and Margaret Armstrong[3]

Current ideas of robustness in geostatistics concentrate upon estimation of the experimental variogram. However, predictive algorithms can be very sensitive to small perturbations in data or in the variogram model as well. To quantify this notion of robustness, nearness of variogram models is defined. Closeness of two variogram models is reflected in the sensitivity of their corresponding kriging estimators. The condition number of kriging matrices is shown to play a central role. Various examples are given. The ideas are used to analyze more complex universal kriging systems.

KEY WORDS: variogram, robustness, kriging, conditioning number.

INTRODUCTION

The theory of robust statistics is relatively recent (see, for example, Huber, 1977, 1981 and the references therein). Geostatisticians have thus only rather recently examined robust procedures. Armstrong and Delfiner (1980), and Cressie and Hawkins (1980) address the problem and provide relevant bibliographies. Attention has largely been centered upon robust estimation of the experimental variogram and tended to ignore the possible sensitivity of further procedures (e.g., kriging) upon the choice of the variogram model γ itself and sensitivity to the configuration of experimental data points. Matheron (1978) gives an amusing and instructive illustration; data were sent to a number of geostatisticians, each of whom independently fitted distinct variogram models from this one data set. Nevertheless, predictive estimates using the diverse models produced almost equivalent results. In some sense, all the models were "close" to each other, and in like fashion, so were the kriging estimators.

[1]Manuscript received 7 April 1983; revised 13 February 1984.
[2]Department of Mathematics, University of Queensland, St. Lucia, Queensland, 4067, Australia. Research performed while on leave at Centre de Geóstatistique et de Morphologie Mathématique, Fontainebleau.
[3]Centre de Géostatistique et de Morphologie Mathématique, École des Mines de Paris, 35 Rue Saint-Honoré, 77305, Fontainebleau, France.

There appears to be a gap here: There is no point whatever in robust estimation of an experimental variogram or generalized covariance if, after such care, predictive estimation procedures (e.g., kriging) based on these are not themselves also robust. As Wilkinson (1971) has pointed out, the primary object of any analysis of such algorithms should be to expose the sources of potential instabilities. Hampel (1971) has rigorously defined a notion of robustness for the estimation of statistical parameters, but no one seems to have taken the further step of examining the robustness of predictive algorithms based upon the preliminary robust estimates.

This note addresses itself to this and a number of related problems in a specifically geostatistical context. First, the intuitively appealing notion of "similar" variogram models is made a little more precise and the closeness of kriging estimators for such variograms, in some sense near each other, is also examined. Then, changes in the sampling configuration that affect kriging is scrutinized. Finally, although some doubts about the usefulness of universal kriging have been raised (Armstrong, 1982) the method is considered for completeness.

As might be expected, various error estimates involve condition numbers of the kriging matrix. It must be emphasized that these estimates are rather conservative. This is because they must encompass exceptional and nasty cases, usually contrived, as well as the more common cases. A number of numerical examples are given to illustrate the principles involved.

δ NEIGHBORHOODS OF VARIOGRAMS AND POINT ESTIMATION KRIGING

Huber (1977) has observed that the term "robust" is often used ambiguously and defines it as signifying insensitivity to small perturbations in data or assumptions. This same sense of the word will be used herein.

Definition

Let $\mathfrak{A}$ be the class of valid variograms $\gamma(r)$ in $\mathbf{R}^n$ expressed in terms of the isotropic variable r. We say a valid variogram $g(r)$ in $\mathfrak{A}$ lies in the δ neighborhood $N_\delta(\gamma)$ of $\gamma(r)$ is

$$1 - \delta < g(r)/\gamma(r) < 1 + \delta, \quad 0 < \delta < 1, 0 < r < \infty \tag{1}$$

This may be rewritten as

$$N_\delta(\gamma) = \{g \in \mathfrak{A}: |g/\gamma - 1| < \delta\}$$

where $|\cdot|$ is the sup norm on continuous functions from $(0, \infty)$ to itself.

The δ neighborhoods endow $\mathfrak{A}$ with a topology which is Hausdorff. This is unimportant in the present context because we are not interested in taking limits of sequences of variograms, although such a procedure would well arise in using

Hampel's (1971) definition. The point is that the definition gives precise quantified form to the intuitive notion of variograms which are, in one sense or another, close together.

The usual intuitive idea (based on visual comparison) of measuring the difference between γ and g by $|\gamma(r) - g(r)|$ is inadequate. For example, if $g(r)$ is the spherical model and $\gamma(r)$ the Gaussian, then $|\gamma(r) - g(r)| \to 0$ as $r \to 0+$ for all possible combinations of parametric values. But the behavior at the origin is very different, with important consequences for prediction based upon the variograms, and $g(r)/\gamma(r) \to \infty$ as $r \to 0+$. Although two individual models may appear very similar in so far as $\sup_r |\gamma(r) - g(r)|$ is small, their relative behavior as measured by $|\gamma(r) - g(r)|/g(r)$ may be very different, and it is this which places them in different δ neighborhoods.

Suppose we are faced with stationary kriging involving point estimation with the variogram γ. The block estimation case is very similar. Let there be N samples z_i at x_i, $i = 1, 2, \ldots, N$. In the notation of Delfiner (1982a,b), the kriging system is

$$\sum_{j=1}^{N} \lambda_j \gamma(x_i - x_j) + \mu = \bar{\gamma}(x_i, V), \quad i = 1, 2, \ldots, N$$

$$\sum \lambda_j = 1$$

where $\bar{\gamma}(x_i, V) = V^{-1} \int_V \gamma(x_i - x)\, dx$. In matrix notation, $\Gamma X = B_\gamma$, where $\Gamma = [\gamma_{ij}] = [\gamma(x_i - x_j)]$, $X^T = (\lambda_1, \ldots, \lambda_N, \mu)$ and $B_\gamma^T = [\bar{\gamma}(x_i, V)]$. Consider the corresponding system for $g \in N_\delta(\gamma)$, $GX_g = B_g$, with $G = [g_{ij}]$, $B_g^T = [\bar{g}(x_i, V)]$. From (1), integration over V gives, for $i = 1, \ldots, N$

$$(1 - \delta)\, \bar{\gamma}(x_i, V) < \bar{g}(x_i, V) < (1 + \delta)\, \bar{\gamma}(x_i, V) \qquad \text{whence} \tag{2}$$

$$(1 - \delta)\, ||B_\gamma|| < ||B_g|| < (1 + \delta)\, ||B_\gamma|| \tag{3}$$

where $||B||$ is some $L(p)$ vector norm, usually with $p = 1, 2$ or ∞ (see Appendix). Similarly

$$(1 - \delta)\, \gamma_{ij} < g_{ij} < (1 + \delta)\, \gamma_{ij}$$

implies that

$$(1 - \delta)\, ||\Gamma|| < ||G|| < (1 + \delta)\, ||\Gamma|| \tag{4}$$

where $||G||$ is the matrix norm consistent with $||Y||$, defined by $||G|| = \sup_{||Y||=1} ||GY||$ (Wilkinson, 1967). To see that (4) is in fact true, recall that the variogram values g_{ij} are all of like sign, say $g_{ij} \geq 0$ for specificity, as are the γ_{ij}. If $Y^T = [y_j]$, then $GY = [\Sigma\, g_{ij} y_j]$ assumes its maximum at vectors whose elements are all nonnegative. Consequently, for each summand

$$(1 - \delta)\, \gamma_{ij} y_j \leq g_{ij} y_j \leq (1 + \delta)\, \gamma_{ij} y_j$$

with equality only if $y_j = 0$ and eq. (4) follows. If all $g_{ij} \leqslant 0$, similar results follow. Write

$$X_g = X + \Delta X, \quad B_g = B_\gamma + \Delta B, \quad G = \Gamma + \Delta\Gamma$$

From (1) and (2)

$$||\Delta\Gamma|| \leqslant \delta\, ||\Gamma|| \tag{5}$$

$$||\Delta B|| \leqslant \delta\, ||B_\gamma|| \tag{6}$$

while the kriging system for γ gives

$$||B_\gamma|| \leqslant ||\Gamma||\ ||X|| \tag{7}$$

The kriging system for g is rewritten as

$$(\Gamma + \Delta\Gamma)\,(X + \Delta X) = B_\gamma + \Delta B$$

for which

$$(\Gamma + \Delta\Gamma)\,(\Delta X) + (\Delta\Gamma)\,X = \Delta B$$

$$\Delta X = [I + \Gamma^{-1}(\Delta\Gamma)]^{-1}\ \Gamma^{-1}\,[\Delta B - (\Delta\Gamma)\,X] \tag{8}$$

provided the matrix $I + \Gamma^{-1}(\Delta\Gamma)$ is nonsingular. This is guaranteed by the following assumption

$$||\Gamma^{-1}||\ ||\Delta\Gamma|| < 1 \tag{9}$$

for then the expansion

$$[I + \Gamma^{-1}(\Delta\Gamma)]^{-1} = I - \Gamma^{-1}(\Delta\Gamma) + [\Gamma^{-1}(\Delta\Gamma)]^2 - \cdots$$

is valid. Note further that

$$\begin{aligned} ||[I + \Gamma^{-1}(\Delta\Gamma)]^{-1}|| &\leqslant 1 + ||\Gamma^{-1}||\ ||\Delta\Gamma|| + (||\Gamma^{-1}||\ ||\Delta\Gamma||)^2 + \cdots \\ &= 1/(1 - ||\Gamma^{-1}||\ ||\Delta\Gamma||) \end{aligned}$$

Substitution in (8) gives

$$||\Delta X|| \leqslant ||\Gamma^{-1}||\,(||\Delta B|| + ||\Delta\Gamma||\ ||X||)/(1 - ||\Gamma^{-1}||\ ||\Delta\Gamma||) \tag{10}$$

Introducing the usual notation for the condition number of a matrix, $\kappa(\Gamma) = ||\Gamma||\ ||\Gamma^{-1}||$ and further substituting the relations (5), (6), and (7) into (10) produces

$$\begin{aligned} ||\Delta X|| &\leqslant ||\Gamma^{-1}||\,(\delta\,||B_\gamma|| + \delta\,||\Gamma||\ ||X||)/(1 - \delta\,||\Gamma||\ ||\Gamma^{-1}||) \\ &\leqslant ||\Gamma^{-1}||\,(\delta\,||\Gamma||\ ||X|| + \delta\,||\Gamma||\ ||X||)/[1 - \delta\kappa(\Gamma)] \\ &= 2\delta\,||X||\ \kappa(\Gamma)/(1 - \delta\kappa(\Gamma)) \end{aligned}$$

and so we have

$$||\Delta X||/||X|| \leqslant 2\delta\kappa(\Gamma)/[1 - \delta\kappa(\Gamma)] \tag{11}$$

Observe that the weaker inequality $||\Gamma^{-1}(\Delta\Gamma)|| < 1$ is sufficient for the invertibility of $I + \Gamma^{-1}(\Delta\Gamma)$, but this follows from (9) which, in turn, holds provided that

$$\delta\kappa(\Gamma) < 1 \tag{12}$$

The conclusion to be drawn from this analysis is that, for any valid variogram γ and any $\epsilon > 0$, there is a δ neighborhood $N_\delta(\gamma)$ of variograms which produces a relative error $||\Delta X||/||X||$ in kriging that is no greater than ϵ provided $\delta < \min\,[1/\kappa(\Gamma),\ \epsilon/(2+\epsilon)\,\kappa(\Gamma)]$. Suppose $\delta\kappa(\Gamma) < 1$, and put $2\delta\kappa(\Gamma)/(1 - \delta\kappa(\Gamma)) < \epsilon$, and certainly $||\Delta X||/||X|| < \epsilon$ by (11). This estimate for δ will, on the whole, be extremely conservative, but could almost be attained with carefully constructed perturbations ΔB and $\Delta\Gamma$, and thus be theoretically possible for $g \in N_\delta(\gamma)$ (Wilkinson, 1967, p. 73-75). Consequently the robustness, with regard to small perturbations of the variogram γ, each perturbed model viewed as a variogram in $N_\delta(\gamma)$ evaluated at the same experimental points x_i, ultimately depends on the condition number $\kappa(\Gamma)$ of the kriging matrix Γ corresponding to the original unperturbed variogram γ. Clearly a change in the configuration of data points x_i, while leaving the function γ unchanged will, however, perturb Γ and B in the kriging equation. This effect is discussed below.

Perturbing γ to $g \in N_\delta(\gamma)$ affects the kriging variance. Denote the kriging variances for γ, g, respectively, by $\sigma_\gamma^2, \sigma_g^2$

$$\begin{aligned}
\sigma_\gamma^2 &= X^T B_\gamma - \bar{\gamma}(V, V) \\
\sigma_g^2 &= X_g^T B_g - \bar{g}(V, V) \\
&= (X^T + \Delta X^T)(B_\gamma + \Delta B) - \bar{g}(V, V) \\
&= \sigma_\gamma^2 + (\Delta X^T) B_\gamma + X^T(\Delta B) + \Delta X^T \Delta B + \bar{\gamma}(V, V) - \bar{g}(V, V)
\end{aligned}$$

It follows that

$$\begin{aligned}
|\sigma_g^2 - \sigma_\gamma^2| &\leqslant 2\delta\,||X||\;||B_\gamma||\,\kappa(\Gamma)/[1 - \delta\kappa(\Gamma)] \\
&\quad + \delta\,||X||\;||B_\gamma|| + 2\delta^2\,||X||\;||B_\gamma||\,\kappa(\Gamma)/[1 - \delta\kappa(\Gamma)] \\
&\quad + \bar{\gamma}(1 - \bar{g}/\bar{\gamma}) \\
&\leqslant 2\delta(1+\delta)\,||\Gamma||\;||X||^2\,\kappa(\Gamma)/[1 - \delta\kappa(\Gamma)] \\
&\quad + \delta\,||\Gamma||\;||X||^2 + \delta\bar{\gamma}
\end{aligned}$$

$$|\sigma_g^2 - \sigma_\gamma^2| \leqslant \delta\{\bar{\gamma} + ||\Gamma||\;||X||^2 + 2\,||\Gamma||\;||X||^2\,(1+\delta)\,\kappa(\Gamma)/[1 - \delta\kappa(\Gamma)]\} \tag{13}$$

Example 1

Consider the following illustration of instability of the Gaussian model under perturbation of sill and parameter. In the plane, let sampling points be at positions (-0.4, 0), (0.4, 0), and (0.39, 0.1). In practice, these last two points would be combined, but this example, and the next, are contrived to illustrate the mechanisms at work and to show how bad things can get even in the simplest of systems.

Take the original variogram γ to be Gaussian with sill at 1 and parameter $a = 1$, while g is Gaussian with sill at 1 and parameter $a = 1.1$. All computations krige a 1.0×1.0 block. Observe that g and γ are in the same δ neighborhood, and kriging is not especially sensitive to parameter changes in this case. Indeed, if we write $\gamma(r) = 1 - e^{-r^2}$, $g(r) = c(1 - e^{-r^2/a^2})$, then

$$F(r) = g(r)/\gamma(r) = c(1 - e^{-r^2/a^2})/(1 - e^{-r^2})$$

only takes on values between c and c/a^2 because F is monotonic in $0 < r < \infty$, as can be checked by differentiation $F'(r)$ having the sign of $1 - (1/a)$. Thus, with g and γ as above, the ratio increases from 1/1.1 to 1.1 as r goes from 0 to ∞, and $\delta = 0.1$. The corresponding change in the third weight is about 14%.

Example 2

Consider the following comparison between the stability of Gaussian and spherical models when a small nugget effect is added.

Let the sampling configuration be as in the example above. We take the perturbation g to be the original variogram γ plus the nugget effect of 0.01. The nugget effect actually takes g out of $N_\delta(\gamma)$ because $\lim_{r \to 0^+} g(r)/\gamma(r)$ is unbounded. However, at distances not too far away from the origin, the variograms appear "close," with δ seemingly around 1%. Nonetheless, the Gaussian gives very unstable behavior whereas the spherical model is relatively robust.

(a) γ Gaussian, with sill at 1 and parameter $a = 1$, while g is Gaussian with $a = 1$ but sill at 0.99 with a nugget effect of 0.01.

(b) γ is spherical, with sill at 1 and range $3^{1/2}$ corresponding to the effec-

Table 1.

Points	γ kriging weights	g kriging weights
(-0.4, 0)	0.4982	0.4985
(0.4, 0)	0.3970	0.4112
(0.39, 0.1)	0.1048	0.0903
σ_K^2	0.0184	0.0146

Table 2a.

Points	γ kriging weights	g kriging weights
(−0.4, 0)	0.4982	0.4954
(0.4, 0)	0.3970	0.3253
(0.39, 0.1)	0.1048	0.1793
σ_K^2	0.0184	0.0224

Table 2b.

Points	γ kriging weights	g kriging weights
(−0.4, 0)	0.4824	0.4806
(0.4, 0)	0.2721	0.2716
(0.39, 0.1)	0.2454	0.2478
σ_K^2	0.0961	0.0988

tive range of the Gaussian model above being $3^{1/2}$, whereas g is spherical with range $3^{1/2}$ but sill 0.99 with nugget effect once more of 0.01.

Observe that the 1% change induces a similar perturbation in the third weight in the spherical model, but over 70% change in that weight where the Gaussian model is concerned. The Gaussian model itself is ill-conditioned and this is exacerbated by the configuration. If $\gamma(r) = 1 - e^{-r^2}$, here $\kappa(\Gamma) \approx 97$, whereas the spherical model is not too ill-conditioned. Further, the configuration of sampling points is bound to be bad for conditioning. If the obvious is done and the last two points combined, the respective changes in the combined weights turns out to be only 0.6% (Gaussian), 0.4% (spherical).

PERTURBED SAMPLING CONFIGURATION

Suppose that the configuration C of experimental points $x_i \in \mathbf{R}^n$, $i = 1, 2, \ldots, N$ is perturbed to a configuration C' with positions $x_i + \Delta x_i$, but the underlying variogram model remains unchanged. Consider only the case of point estimation kriging: block estimation is very similar. Denote the stationary kriging system for the unperturbed grid C by

$$\Gamma_c X = B_c$$

where $\Gamma_c = [(\gamma_{ij}) = \gamma(x_i - x_j)]$, $B_c = [\bar{\gamma}(x_i, V)]$; and that of C' by

$$(\Gamma_c + \Delta\Gamma)(X + \Delta X) = B_c + \Delta B$$

Here, the elements of $\Gamma_c + \Delta\Gamma$ are $\gamma(x_i - x_j + \Delta x_i - \Delta x_j)$ and those of $B_c + \Delta B$ are $\bar{\gamma}(x_i + \Delta x_i, V)$.

Further assume that γ is everywhere differentiable, except that, at the origin, the partial derivatives exist from the right and are upper continuous. Several assumptions concerning approximations will be made as they arise. Note that, formally, as in the derivation of (10)

$$||\Delta X|| \leqslant ||\Gamma_c^{-1}|| \, (||\Delta B|| + ||\Delta\Gamma|| \, ||X||)/(1 - ||\Gamma_c^{-1}|| \, ||\Delta\Gamma||) \quad (14)$$

Now

$$\gamma(x_i - x_j + \Delta x_i - \Delta x_j) = \gamma_{ij} + (\Delta x_i - \Delta x_j)^T \gamma'(c_{ij}) \quad (15)$$

where c_{ij} lies between $x_i - x_j$ and its perturbation, all expressed either as vector arguments or as isotropic variables. In eq. (15), the symbol γ' denotes the n vector of the partial derivatives $(D_k\gamma)$. Write $\Delta_{ij} = \Delta x_i - \Delta x_j$ and use the approximation (not unreasonable in a geostatistical situation) that

$$D_k\gamma(c_{ij}) \approx D_k\gamma(x_i - x_j), \quad k = 1, 2, \ldots, n$$

It follows that, within the limits of this approximation

$$\Delta\Gamma = [\Delta_{ij}^T \gamma'(x_i - x_j)] = (\Delta_{ij}^T \gamma'_{ij}) \quad (16)$$

and note that the diagonal elements are no longer necessarily zero.

As for the elements of $B_c + \Delta B$

$$\bar{\gamma}(x_i + \Delta x_i, V) = V^{-1} \int_V \gamma(x_i + \Delta x_i - x)\, dx$$

$$\approx \bar{\gamma}(x_i, V) + V^{-1} \int_V \Delta x_i^T \gamma'(x_i - x)\, dx$$

again using the approximation above. So

$$|\bar{\gamma}(x_i + \Delta x_i, V) - \bar{\gamma}(x_i, V)| \leqslant ||\Delta x_i|| \, V^{-1} \int_V ||\gamma'(x_i - x)|| \, dx$$

Assume that with respect to the configuration C and the volume V

$$||\gamma'(x_i - x)|| \leqslant L, \quad x \in V \qquad \text{and} \quad i = 1, 2, \ldots, N$$

Since then $\int_V ||\gamma'(x_i - x)|| \, dx \leqslant LV$ and, putting $D_{\max} = 2 \max_i ||\Delta x_i||$ it follows that

$$||\Delta B|| \leqslant LD_{\max}/2 \quad (17)$$

Note that

$$||B_c|| \leqslant ||\Gamma_c||\ ||X|| \qquad ||X|| \leqslant ||\Gamma_c^{-1}||\ ||B|| \quad \text{and so}$$

$$||\Delta X||/||X|| \leqslant \kappa(\Gamma_c)\,(||\Delta B||/||B_c|| + ||\Delta\Gamma||\ ||\Gamma_c^{-1}||/(1 - \kappa(\Gamma_c)\ ||\Delta\Gamma||/||\Gamma_c||) \tag{18}$$

By (16), $||\Delta\Gamma|| \leqslant D_{\max}\,||(\gamma'_{ij})||$ which is no greater than $NLD_{\max}$ in $L(1)$ norm or $N^{1/2}LD_{\max}$ in $L(2)$ norm. Thus the estimate (18) can be made more explicit, if somewhat coarsened, as

$$||\Delta X||/||X|| \leqslant \kappa(\Gamma_c)\,(LD_{\max}/2||B_c|| + NLD_{\max}\,||\Gamma_c^{-1}||/(1 - \kappa(\Gamma_c)\,NLD_{\max}\,||\Gamma_c||^{-1}). \tag{19}$$

The kriging variance of the configuration C can be written as $\sigma_c^2 = X^T B_c - \bar{\gamma}(V, V)$ and that of the perturbed configuration

$$\sigma_{c'}^2 = (X^T + \Delta X^T)\,(B_c + \Delta B) - \bar{\gamma}(V, V)$$

since the underlying variogram model is assume to remain the same. Hence

$$|\sigma_c^2 - \sigma_{c'}^2| \leqslant ||\Delta X||\ ||B_c|| + ||X||\ ||\Delta B|| + ||\Delta X||\ ||\Delta B||$$

which can be somewhat laboriously expressed in terms of $\kappa(\Gamma_c)$ and estimates above involving the kriging matrix and derivatives, as well as the perturbation, without reference to X. The really important point to note is that again, the conditioning of the kriging system is the most important factor in the error estimates. In the case above both the initial configuration and the underlying variogram determine the conditioning number.

Example 3

Let the sampling points in the configuration C be at positions (−0.4, 0), (0.4, 0), (0.41, 0.1), and that of C' be at (−0.4, 0), (0.4, 0), (0.39, 0.1). Again using a Gaussian model $\gamma(r) = 1 - e^{-r^2}$, the relations depicted in Table 3 hold.

Table 3.

λ_1	λ_2	λ_3	σ_K^2
	C		
0.4998	0.4757	0.0245	0.0186
	C'		
0.4982	0.3970	0.1048	0.0184

The condition number of almost 100 goes some way toward explaining why a perturbation of $2\frac{1}{2}\%$ in the third coordinate alters the weight λ_3 by some 328%. Even for the numerically unstable Gaussian model, if the last two positions are combined the change in the combined weight reduces to about 15%, corresponding to a configuration change of 1.3%. Moreover, in larger systems this inherent instability would be compounded by additional points and increasing dimension of the kriging system.

UNIVERSAL KRIGING

Although problems involving indeterminacy of both the drift and the underlying variogram arise in universal kriging (Armstrong, 1982), it is nonetheless theoretically instructive to examine this more complex kriging system for robustness. We restrict the discussion to point estimation of the random value $Z(x)$ from sample values $z(x_1), \ldots, z(x_N)$, as a linear combination $\Sigma\, \lambda_i z(x_i)$ under the hypothesis that $Z(x)$ is weakly stationary with covariance function $K(x, y)$. Here $Z(x)$ has a decomposition into a deterministic drift $m(x)$ and a weakly stationary random function $Y(x)$ with zero mean

$$Z(x) = m(x) + Y(x)$$

Drift is expressed as a linear combination $\Sigma\, a_j f^j(x)$, where the functions $f^j(x)$ are monomials of degree no greater than p. The kriging equations form an $N + k(p)$ degree system, where $k(p)$ is the number of monomials in the space variables of degree $\leqslant p$. In matrix notation, $\Gamma\lambda = B$, where

$$\Gamma = \begin{vmatrix} K & F \\ F^T & 0 \end{vmatrix} \quad \begin{matrix} \lambda^T = [\lambda_1, \ldots, \lambda_N, \quad a_1, \ldots, a_{k(p)}] \\ B^T = [K(x_i, x_0), \ldots, K(x_N, x_0), \quad f^1(x_0), \ldots, f^{k(p)}(x_0)] \end{matrix}$$

$$F = [f^j(x_i)]_{N \times k(p)} \quad K = [K(x_i, x_j)]_{N \times N}$$

Two types of sensitivity spring to mind.

(i) Robustness associated with δ neighborhoods of the covariance K and the sampling grid. This leads to arguments and estimates essentially of the same character as those of previous sections;

(ii) Robustness associated with the drift; for example, attempting to increase the degree to $p + 1$ and adding monomials $f^{k(p)+1}, \ldots, f^{k(p+1)}$. For definiteness it is this precise perturbation which will be considered below.

The new system, with drift now being

$$m_1(x) = m(x) + \sum_{j=k(p)+1}^{k(p+1)} a_j f^j(x)$$

is given by

$$\begin{vmatrix} \Gamma & H \\ H^T & 0 \end{vmatrix} \begin{vmatrix} \lambda + \Delta\lambda \\ a \end{vmatrix} = \begin{vmatrix} B \\ h \end{vmatrix} \tag{20}$$

with $H = [f^j(x_i)]$, $i = 1, \ldots, N; j = k(p) + 1, \ldots, k(p + 1)$, $h^T = [f^{k(p)+1}(x_0), \ldots, f^{k(p+1)}(x_0)]$. Recall that $\Gamma\lambda = B$ and it follows at once that

$$\Gamma(\Delta\lambda) + Ha = 0 \tag{21}$$

$$H^T(\lambda + \Delta\lambda) = h \tag{22}$$

Note that in general H is not a square matrix and it is meaningless to speak of anything but a *pseudoinverse*. That is, given the system $HU = V$, where U has $[k(p + 1) - k(p)]$ elements and V is an N vector, the pseudoinverse H^+ is a unique matrix satisfying $U = H^+V$ such that U is the minimum length solution of the least-squares problem: minimize $||HU - V||$ (Lawson and Hanson, 1974). The condition number of H is $\kappa(H) = ||H||\ ||H^+||$.

From (21), $\Delta\lambda = -\Gamma^{-1}Ha$, and so

$$||\Delta\lambda|| \leqslant ||\Gamma^{-1}||\ ||H||\ ||a|| \tag{23}$$

and since $||\Gamma||\ ||\lambda|| \geqslant ||B||$

$$||\Delta\lambda||/||\lambda|| \leqslant \kappa(\Gamma)||H||\ ||a||/||B|| \tag{24}$$

From (22), $H^T(\lambda - \Gamma^{-1}Ha) = h$, whence

$$a = H^+\Gamma\lambda - H^+\Gamma H^{+T} h$$

$$||a|| \leqslant ||H^+||\ ||\Gamma||\ ||\lambda|| + \kappa(H)||\Gamma||\ ||h|| \tag{25}$$

$$||a||/||\lambda|| \leqslant ||H^+||\ ||\Gamma|| + \kappa(H)||\Gamma||^2\ ||h||/||B|| \tag{26}$$

Using (23) and (24)

$$||\Delta\lambda||/||\lambda|| \leqslant \kappa(\Gamma)\,\kappa(H)\,(1 + ||\Gamma||\ ||H||\ ||h||/||B||) \tag{27}$$

Together, (26) and (27) give the relative change in the weights of the universal kriging estimator for $Z(x)$ introduced by adding a higher degree monomial set to the expression for the drift $m(x)$. The product of two condition numbers of the matrices Γ and H are essentially involved in the estimates of the perturbations that are introduced with the monomial set. This suggests that the potential instability of the kriging system is increased whenever a better approximation to the drift is sought. This practical consideration is possibly a third problem of universal kriging, to be added to those arising from the indeterminacies mentioned above.

DISCUSSION

We have set out to distinguish between

(i) robust procedures which are used to estimate experimental variograms, and
(ii) the overall robustness of prediction algorithms, such as kriging, which use these variograms.

In regression, the very procedures themselves guarantee that data sets which are near to each other produce regression predictions (e.g., the experimental variogram) within the span of their design vectors which are also close to each other, though the regression equations may be very different in form. In stark contrast, if subsequent kriging is applied to the same raw data but using variograms in different δ neighborhoods, very different kriging weights (and thus markedly different predictions of block or point grades) will result. This is because "nearness" in the regression sense is ultimately based upon a least-squares concept of distance, no matter what refinements are used to enhance robustness.

On the other hand, nearness of two variograms is defined in terms of the relative difference between them $|(\gamma - g)/g|$. This distance defines the δ neighborhoods. Resultant kriging will give similar results for near sets of data only if it proceeds from variograms in the same δ neighborhood. Even then the robustness of the kriging predictor depends heavily upon the condition number of the kriging matrix. So, as the above examples have illustrated, circumstances can arise in which some classes of variograms (e.g., the Gaussian) produce major differences in subsequent kriging from small variations in the data. This phenomenon is produced by high condition numbers.

Unlike ordinary regression, minor perturbations in the data alone may give rise to major effects in the final kriged values. However apt robust procedures may be for the estimation of experimental variograms, such procedures cannot be regarded as guaranteeing the robustness of the entire predictive algorithm. For this, some such notion of nearness, as expressed by the δ-neighborhood concept, is necessary. Too, close attention must be given to the conditioning of the kriging matrices which arise from the fitted variogram.

APPENDIX

The following standard formulas and definitions for vector and matrix norms are included for completeness so as to make this note self-contained.

Definition 1

If $x = (x_i)$ is an n vector, the $L(1)$, $L(2)$, and $L(\infty)$ norms of x are defined by

(a) $||x||_1 = \Sigma_{i=1}^{n} |x_i|$, the $L(1)$ norm.

(b) $||x||_2 = (\Sigma_{i=1}^n x_i^2)^{1/2}$, the $L(2)$ or Euclidean norm.
(c) $||x||_\infty = \max_{1 \leq i \leq n} |x_i|$, the $L(\infty)$, infinity or uniform norm.

Each of these is a special case of the $L(p)$ norm

$$||x||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \qquad p \geq 1$$

Definition 2

A subordinate or induced matrix norm $||A||_p$ is defined by

$$||A||_p = \max_{x \neq 0} ||Ax||_p / ||x||_p = \max_{||x||=1} ||Ax||_p$$

where usually $p = 1, 2$, or ∞.

Explicit Formulas

If $A = (a_{ij})$, then

$$||A||_1 = \max_j \sum_i |a_{ij}|$$

$$||A||_\infty = \max_i \sum_j |a_{ij}|$$

Although there is no simple explicit expression for $||A||_2$, the condition number $\kappa(A) = ||A||_2 \, ||A^{-1}||_2$ in terms of $L(2)$ norm on square matrices may be expressed as

$$\kappa(A) = |\lambda_{\max}| / |\lambda_{\min}|$$

where $\lambda_{\max}$, $\lambda_{\min}$ are the eigenvalues of, respectively, maximum and minimum modulus.

A concise treatment, with proof, can be found in Broyden (1975), and many examples can be found in Atkinson (1978).

REFERENCES

Armstrong, M., 1982, The problems with universal kriging: Centre de Géostatisque, Fountainebleau, N-784.

Armstrong, M. and Delfiner, P., 1980, Towards a more robust variogram: A case study on coal: Centre de Géostatisque, Fontainebleau, N-671.

Atkinson, K. E., 1978, An introduction to numerical analysis, John Wiley & Sons, New York.

Broyden, C. G., 1975, Basic matrices: Macmillan, London.

Cressie, N. and Hawkins, D. H., 1980, Robust estimators of the variogram: I: Jour. Math. Geol., v. 12, p. 115–126.

Delfiner, P., 1982a, Excerpt from Basic Introduction to Geostatistics: Centre de Géostatisque: Fontainebleau, C-95.

Delfiner, P., 1982b, The intrinsic model of order k: Centre de Géostatisque, Fontainbleau, C-97.[4]

Hampel, F. R., 1971, A general qualitative definition of robustness, Ann. Math. Stat., v. 42, p. 1887–1896.

Huber, P. J., 1977, Robust statistical procedures, Regional Conference Series in Applied Mathematics, SIAM, No. 27, Pennsylvania.

Huber, P. J., 1981, Robust statistics: John Wiley & Sons, New York.

Lawson, C. L. and Hanson, R. J., 1974, Solving least squares problems: Prentice-Hall, New Jersey.

Matheron, G., 1978, Estimer et choisir: Centre de Géostatisque, Fontainebleau, Fascicule 7.

Wilkinson, J. H., 1967, The solution of ill-conditioned linear equations, Ralston, A. and Wilf, H. S. (Eds.), Mathematical methods for digital computers, vol. 2: John Wiley & Sons, New York.

Wilkinson, J. H., 1971, Modern error analysis, SIAM, v. 13, p. 548–568.

[4]Bibliographical note: Delfiner (1982a,b) are to be published in Delfiner, P., "Geostatistics and Kriging," John Wiley & Sons, New York.