Mathematical Geology, Vol. 19, No. 6, 1987

Kriging Without Negative Weights[1]

F. Szidarovszky,[2] E. Y. Baafi,[3] and Y. C. Kim[4]

Under a constant drift, the linear kriging estimator is considered as a weighted average of n *available sample values. Kriging weights are determined such that the estimator is unbiased and optimal. To meet these requirements, negative kriging weights are sometimes found. Use of negative weights can produce negative block grades, which makes no practical sense. In some applications, all kriging weights may be required to be nonnegative. In this paper, a derivation of a set of nonlinear equations with the nonnegative constraint is presented. A numerical algorithm also is developed for the solution of the new set of kriging equations.*

KEY WORDS: nonnegative kriging weights, kriging equations, nonlinear optimization, numerical algorithm.

INTRODUCTION

In linear kriging, the estimation of an average grade of a block $Z(V)$ is considered as a weighted average of n available sample values $Z(x_i)$, $i = 1, 2, \ldots, n$, i.e.

$$Z^*(V) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{1}$$

Weights λ_i ($i = 1, 2, \ldots, n$) usually are determined such that the kriging estimator is unbiased and optimal (i.e., with minimal mean squared error). The unbiasedness condition is equivalent to

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{2}$$

[1]Manuscript received 20 October 1986; accepted 27 January 1987.
[2]Institute of Mathematics and Computer Science, Karl Marx University of Economics, Budapest, Hungary.
[3]Department of Civil and Mining Engineering, The University of Wollongong, Wollongong, NSW, 2500, Australia.
[4]Department of Mining and Geological Engineering, The University of Arizona, Tucson, Arizona 85721, U.S.A.

The mean squared error expressed in terms of the variogram function γ is

$$E\left\{\left[Z(V) - Z^*(V)\right]^2\right\} = -\gamma_{vv} + 2\sum_{i=1}^{n}\lambda_i\gamma_{vi} - \sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_i\lambda_j\gamma_{ij} \quad (3)$$

where

$$\gamma_{ij} = \gamma(x_i - x_j)$$

$$\gamma_{vi} = \frac{1}{V}\int_V \gamma(x - x_i)\,dx$$

$$\gamma_{vv} = \frac{1}{V^2}\iint_{VV} \gamma(x - y)\,dxdy$$

The kriging estimator is described commonly as "optimal" because it has minimal estimation variance (i.e., minimal mean squared error) given in Eq. (3).

Minimization of function (3) under constraint (2) is equivalent to the solution of the linear equations

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{v1} \\ \vdots \\ \gamma_{vn} \\ 1 \end{bmatrix} \quad (4)$$

The weights obtained by solution of these equations are not necessarily nonnegative. Under some conditions, the weights may be required to be nonnegative (Baafi et al., 1986). In such cases, the additional nonnegativity constraints of

$$\lambda_i \geqq 0 \qquad (i = 1, 2, \ldots, n) \quad (5)$$

must be considered and included in the optimization process (Schapp and George, 1981; Barnes and Johnson, 1984). Because any nonnegative number can be rewritten as a square of a real number, this extended optimization problem can be rewritten as

$$\text{minimize } -\gamma_{vv} + 2\sum_{i=1}^{n} a_i^2\gamma_{vi} - \sum_{i=1}^{n}\sum_{j=1}^{n} a_i^2 a_j^2\gamma_{ij}$$

$$\text{subject to } \sum_{i=1}^{n} a_i^2 - 1 = 0 \quad (6)$$

By use of the Lagrange method, this constrained optimization problem can be reformulated as an unconstrained minimization of the Lagrangian defined as

$$
\begin{aligned}
L(a_1, \ldots, a_n, \mu) = -\gamma_{vv} &+ 2 \sum_{i=1}^{n} a_i^2 \gamma_{vi} \\
&- \sum_{i=1}^{n} \sum_{j=1}^{n} a_i^2 a_j^2 \gamma_{ij} \\
&- 2\mu \left(\sum_{i=1}^{n} a_i^2 - 1 \right)
\end{aligned} \tag{7}
$$

Thus, the first-order necessary conditions are

$$\frac{\partial L}{\partial a_i} = 4a_i \left(\gamma_{vi} - \sum_{j=1}^{n} a_j^2 \gamma_{ij} - \mu \right) = 0 \qquad (i = 1, 2, \ldots, n) \tag{8}$$

$$\frac{\partial L}{\partial \mu} = -2 \left(\sum_{i=1}^{n} a_i^2 - 1 \right) = 0 \tag{9}$$

Consider Eq. (8). If $a_i \neq 0$, then necessarily

$$\sum_{j=1}^{n} a_j^2 \gamma_{ij} + \mu = \gamma_{vi} \tag{10}$$

that is, the ith equation of the original kriging system (4) must be satisfied. If $a_i = 0$, the ith equation of Eq. (8) is obviously true. Equation (9) is equivalent to the original condition (2), because $a_i^2 = \lambda_i$.

Summarizing the above observations, we obtain the following formulation. Let $\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*$ be the optimal nonnegative weights from solving Eqs. (8) and (9), and let $\lambda_{i_1}^*, \ldots, \lambda_{i_r}^*$ denote the strictly positive weights among them where $r \leq n$. For weights λ_j^* $(j \neq i_1, \ldots, i_r)$, Eq. (8) above is satisfied because coefficients λ_j^* $(j \neq i_1, \ldots, i_r)$ are zeros. For the strictly positive weights $\lambda_{i_1}^*, \ldots, \lambda_{i_r}^*$, Eq. (4) of the original kriging system also is satisfied. These positive weights can be obtained by using the original kriging method for sample points $x_{i_1}, \ldots, x_{i_r}$.

This important observation is the theoretical basis of a numerical method for finding optimal nonnegative weights. Details of the method are discussed in the next section.

THE NUMERICAL ALGORITHM

Results of the previous section suggest the following method:

Step 1. Generate all possible subsets $\{i_1, \ldots, i_r\}$ of the set $\{1, \ldots, n\}$.

Step 2. In the case of each subset $\{i_1, \ldots, i_r\}$ apply the original version of the kriging method based on sample points $x_{i_1}, \ldots, x_{i_r}$ and variogram $\gamma(h)$. If all the weights obtained are nonnegative, then calculate the corresponding kriging variance by using the formula

$$V_{i_1 i_2, \ldots, i_r} = -\gamma_{vv} + 2 \sum_{k=1}^{n} \lambda_{i_k} \gamma_{vik} - \sum_{k=1}^{r} \sum_{l=1}^{r} \lambda_{i_k} \lambda_{i_l} \gamma_{i_k} i_l \tag{11}$$

Step 3. Accept subset $\{i_1, \ldots, i_r\}$ and corresponding weights $\overset{*}{\lambda}_{i_r}$ which give the smallest error variance defined by Eq. (11) the optimal solution.

The main difficulty in applying this method is the fact that for each subset $\{i_1, \ldots, i_r\}$ we have to repeat the whole kriging estimation process. Because the number of nonempty subsets of $\{1, \ldots, n\}$ equals $(2^n - 1)$, the number of kriging systems to be solved also is equal to $(2^n - 1)$. For example, if $n = 6$, $2^6 - 1 = 63$ solutions are required; if $n = 10$, then $2^n - 1 = 1023$ solutions are necessary. Furthermore $2^n - 1$ tends to infinity exponentially, making computations unattractive even with a digital computer.

However, a systematic generation of all subsets $\{i_1, \ldots, i_r\}$ and use of a numerical technique known as "inversion by blocks" (Szidarovszky and Yakowitz, 1978) can be shown to make the above method very efficient, so that we can generate all possible subsets $\{i_r, \ldots, i_r\}$ and obtain optimal weights and kriging variances without repeating the entire kriging process.

Consider first the kriging equations based on the points $\{i_1, \ldots, i_r\}$. A simple rearranging of the rows and columns gives

$$\begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & \gamma_{i_1 i_1} & \gamma_{i_1 i_2} & \cdots & \gamma_{i_1 i_r} \\ 1 & \gamma_{i_2 i_1} & \gamma_{i_2 i_2} & \cdots & \gamma_{i_2 i_r} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \gamma_{i_r i_1} & \gamma_{i_r i_2} & \cdots & \gamma_{i_r i_r} \end{bmatrix} \begin{bmatrix} \mu \\ \lambda_{i_1} \\ \lambda_{i_2} \\ \vdots \\ \lambda_{i_r} \end{bmatrix} = \begin{bmatrix} 1 \\ \gamma_{vi_1} \\ \gamma_{vi_2} \\ \vdots \\ \gamma_{vi_r} \end{bmatrix} \tag{12}$$

that is, $\mathbf{G}\boldsymbol{\lambda} = \boldsymbol{\gamma}$. Here $\mathbf{G}$ represents a $(r + 1) \times (r + 1)$ matrix, $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ are $(r + 1)$-dimensional vectors. Observe that the diagonal elements of matrix $\mathbf{G}$ are equal to zero.

Let $\boldsymbol{\lambda}_0$ denote the solution of this system.

Assume first that a new point $x_{i_{r+1}}$ is included into kriging; then the extended kriging equations can be rewritten as

$$\begin{bmatrix} \mathbf{G} & \mathbf{g} \\ g^T & \gamma_{i_{r+1}i_{r+1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \lambda_{i_{r+1}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma} \\ \gamma_{vi_{r+1}} \end{bmatrix} \tag{13}$$

where $\mathbf{g}^T = (1, \gamma_{i_{r+1}i_1}, \gamma_{i_{r+1}i_2}, \ldots, \gamma_{i_{r+1}i_r})$ is a $1 \times (r + 1)$ vector.

The inverse of the coefficient matrix of system (13) can be given in block form (Szidarovszky and Yakowitz, 1978)

$$\begin{bmatrix} \mathbf{X} & \mathbf{y} \\ \mathbf{y}^T & s \end{bmatrix} \quad \text{where} \tag{14}$$

$$s = \frac{-1}{\mathbf{g}^T\mathbf{G}^{-1}\mathbf{g}} \quad \text{or} \quad s = \frac{-1}{\mathbf{g}^T\mathbf{m}}$$

$$\mathbf{y} = -\mathbf{G}^{-1}\mathbf{g}s \quad \text{or} \quad \mathbf{y} = -\mathbf{m}s$$

$$\mathbf{X} = \mathbf{G}^{-1} + \frac{1}{s}\mathbf{y}\mathbf{y}^T \quad \text{when} \quad \mathbf{m} = \mathbf{G}^{-1}\mathbf{g} \tag{15}$$

and the solution of the system of equations given by Eq. (13) can be expressed in the form

$$\lambda_{i_{r+1}} = s(\gamma_{vi_{r+1}} - \mathbf{g}^T\boldsymbol{\lambda}_0) \tag{16}$$

$$\begin{aligned} \boldsymbol{\lambda} &= \boldsymbol{\lambda}_0 - \mathbf{G}^{-1}\mathbf{g}\lambda_{i_{r+1}} \\ &= \boldsymbol{\lambda}_0 - m\lambda_{i_{r+1}} \end{aligned} \tag{17}$$

Note that $\lambda_{i_{r+1}}$ is the weight for point $x_{i_{r+1}}$ and $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}$ are the old solution and the vector constructed from the first $r + 1$ components of the old solution of the kriging system.

The new kriging variance is given by the following

$$\begin{aligned} V_{i_1}, \ldots, {}_{i_r i_{r+1}} &= V_{i_1}, \ldots, {}_{i_r} \\ &\quad - (\gamma_{vi_{r+1}} - \mathbf{g}^T\mathbf{G}^{-1}\boldsymbol{\gamma})^2/\mathbf{g}^T\mathbf{G}^{-1}\mathbf{g} \\ &= V_{i_1}, \ldots, {}_{i_r} + (\lambda_{i_{r+1}}^2/s) \end{aligned} \tag{18}$$

Here, $V_{i_1}, \ldots, {}_{i_r}$ is the original kriging variance, prior to adding the new point $x_{i_{r+1}}$.

ADDITION OF A POINT—ALGORITHM 1

If a new point $x_{i_{r+1}}$ is added, the whole estimation method can be replaced by the following algorithm.

Step 1. Compute vector $\mathbf{m} = \mathbf{G}^{-1}\mathbf{g}$

Step 2. Compute the following quantities to get the inverse of the extended coefficient matrix

$$s = \frac{-1}{\mathbf{g}^T\mathbf{m}} \qquad y = -\mathbf{m}s \qquad \mathbf{X} = \mathbf{G}^{-1} + \frac{1}{s}\mathbf{y}\mathbf{y}^T \tag{19}$$

Step 3. Compute the new solution by relations

$$\lambda_{i_{r+1}} = s(\lambda_{vi_{r+1}} - \mathbf{g}^T\boldsymbol{\lambda}_0) \tag{20}$$

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}_0 - m\lambda_{i_{r+1}} \tag{21}$$

Step 4. Update the estimation variance

$$V_{i_1}, \ldots, i_r i_{r+1} = V_{i_1}, \ldots, i_r + \lambda_{i_{r+1}}^2/s \tag{22}$$

This algorithm can be justified as follows:

Equations (19) are simple consequences of definition of $\mathbf{m}$ and relations (15). Equation (20) is the same as Eq. (16), but in order to verify relation (21), we have to observe that relations (18)-(20) imply

$$\begin{aligned}(\gamma_{vi_{r+1}} - \mathbf{g}^T\mathbf{G}^{-1}\gamma)^2/\mathbf{g}^T\mathbf{G}^{-1}\mathbf{g} &= (\gamma_{vi_{r+1}} - \mathbf{g}^T\boldsymbol{\lambda}_0)^2/\mathbf{g}^T\mathbf{m} \\ &= -s(\lambda_{i_{r+1}}/s)^2 \\ &= -\lambda_{i_{r+1}}^2/s\end{aligned}$$

DELETION OF A POINT—ALGORITHM 2

Assume that a point $x_{i_{r+1}}$ has to be dropped from the estimation process. Coefficients of system (13), the inverse matrix (14), solutions (20) and (21), and estimation variance $V_{i_1}, \ldots, i_r i_{r+1}$ are known.

$\mathbf{G}^{-1}$, $\boldsymbol{\lambda}_0$, and $V_{i_1}, \ldots, i_r$ can be determined by the following algorithm.

Step 1. Compute vector

$$\mathbf{m} = -(1/s)\,\mathbf{y} \tag{23}$$

Step 2. Compute matrix

$$\mathbf{G}^{-1} = \mathbf{X} - (1/s)\,\mathbf{y}\mathbf{y}^T \tag{24}$$

Step 3. Determine the solution of the smaller problem

$$\boldsymbol{\lambda}_0 = \boldsymbol{\lambda} + \mathbf{m}\lambda_{i_{r+1}} \tag{25}$$

Step 4: Compute the new estimation variance

$$V_{i_1}, \ldots, i_r = V_{i_1}, \ldots, i_r i_{r+1} - (\lambda_{i_{r+1}}^2/s) \tag{26}$$

Observe that both Algorithms (1) and (2) need $0(r^2)$ operations instead of solving kriging Eqs. (12) or (13) by any variant of the Gaussian elimination method, which needs $0(r^3)$ operations.

GENERATION OF NONEMPTY SUBSETS—ALGORITHM 3

Systematic generation of all nonempty subsets $\{i_1, \ldots, i_r\}$ can be illustrated (Fig. 1) where a particular case of $n = 4$ is chosen. All nonempty subsets of $\{1, 2, \ldots, n\}$ are illustrated by vertices of a directed graph (tree), in which directed edges connect vertices.

$$\boxed{\begin{array}{l}\text{drop}\\ j_1, j_2, \ldots, j_s\end{array}} \rightarrow \boxed{\begin{array}{l}\text{drop}\\ j_1, j_2, \ldots, j_s, j_{s+1}\end{array}}$$

where $j_{s+1} \geqq j_l$ ($l = 1, 2, \ldots, s$). Note that in the case of any vertex which corresponds to the subset $\{i_1, \ldots, i_r\}$

$$\{i_1, \ldots, i_r\} \cup \{j_1, \ldots, j_s\} = \{1, \ldots, n\}$$

and these subsets are disjoint. Search Algorithm (3) can be described as follows:

Step 1. Perform the kriging method by using all sample points. If all weights λ_i are nonnegative, stop. If not, go to Step 2.

Step 2. Set $s = 0$, $t = 0$, $M = \infty$, $M_1 = \infty$.

Step 3. If at least one of the conditions

$$s = n - 1 \qquad j_s = n \qquad t = n \qquad M_1 > M \qquad K = 1$$

is satisfied, set $K = 0$ and go to Step 4, otherwise go to Step 5.

Step 4. Set $t = j_s$, $s = s - 1$, and go to Step 7.

Step 5. Set $s = s + 1$, $t = t + 1$, $j_s = t$, and go to Step 6.

Step 6. Use Algorithm (2) assuming that the j_{s+1} point is dropped from the estimation process for updating the inverse of the coefficient matrix of the kriging system, new weights, and new estimation variance. Save these quantities together with the corresponding sequence of $\{j_1, \ldots, j_s\}$, if necessary. If all weights are nonnegative, set $K = 1$; otherwise, set $K = 0$ and let M_1 be the new estimation variance. If $M_1 < M$, set $M = M_1$ and save the weights as the optimal nonnegative weights up to this stage. Go to Step 8.

Step 7. Use Algorithm 1 to recall the coefficient matrix, weights, and estimation variance; go to Step 3.

Step 8. If $s = 1$ and $j_s = n$, stop. Otherwise go to Step 3. These steps are illustrated (Fig. 1) as follows. *Step 1* and *Step 2* correspond to the starting point A of the tree.

In Step 1, s is the number of dropped points, which is zero initially. Variable t is used only when we move backward on an edge. In that case, t gives the largest value of $j_1, j_2, \ldots$, which belongs to the endpoint of that edge. In

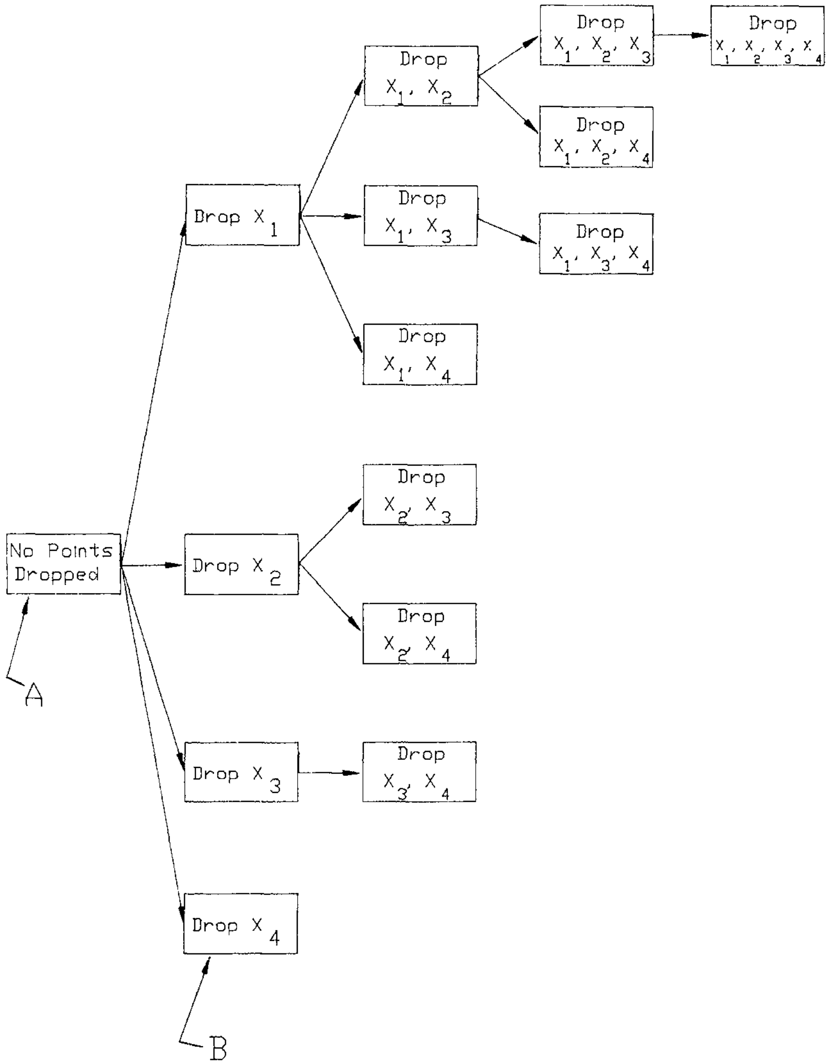

Fig. 1. Illustration of Algorithm 3.

the case of the initial point, we did not move backward. Consequently, we chose $t = 0$. M is the least estimation variance found in previous cases using only nonnegative weights, and M_1 is the estimator variance belonging to the vertex under consideration.

If at least one negative weight exists the estimation variance belonging to the initial vertex must not be considered in the minimization process because it does not belong to feasible solutions. Thus, in this case, M and M_1 should be set as ∞.

Step 3 checks whether a move forward or backward occurs. If $s = n - 1$ points have been dropped, by dropping an additional point only the empty set can be obtained. If $j_s = n$, no additional point can be dropped which has a larger subscript than j_s. If $t = n$, all points which can be reached by a single forward step from this point have been examined before. If $M_1 > M$, the estimation variance corresponding to this point is larger than the best one found so far.

In the first three cases, we cannot proceed forward along the tree, because no additional point exists or they have been searched earlier. In the last case, no reason exists to drop additional points, because the resulting estimation variance would be even worse. Step 4 gives a backward step, and Step 5 gives a forward step. Steps 6 and 7 update coefficients and the estimation variance. Step 8 checks whether the algorithm terminates or not.

If all quantities corresponding to all of the vertices cannot be saved because of limited memory, *Step 7* should be modified.

Step 9. By using Algorithm (1), update the coefficient matrix of the kriging system, weights, and estimation variance. Go to Step 3.

IMPLEMENTATION OF THE DEVELOPED ALGORITHMS

A computer program ZKRIG was developed to implement the above three algorithms. Program ZKRIG is a block kriging program which uses the above algorithms to generate optimal nonnegative weights (Fig. 2).

In order to reduce the tree search (Algorithm 3), program ZKRIG has an option not to search below a minimum number of a combination of sample points. This minimum number which is input by the user will depend on both the drilling density and size of the block to be estimated. Note that no precise rule gives that minimum number. It depends on the user's subjective judgement and past experience.

CONCLUSION

In linear kriging, negative weights are sometimes unavoidable. To guarantee nonnegative kriging weights, an additional constraint $\lambda_i \geq 0$ ($i = 1, 2, \ldots, n$) must be considered and included in the kriging process. A numerical algorithm has been presented for the solution of the new set of kriging equations. The developed algorithm generates optimal nonnegative weights from a set of sample points. The method selects a subset of samples from the available samples such that weights of the subset are all nonnegative and also satisfy the unbiasedness condition of kriging. Also, the algorithm ensures the least esti-

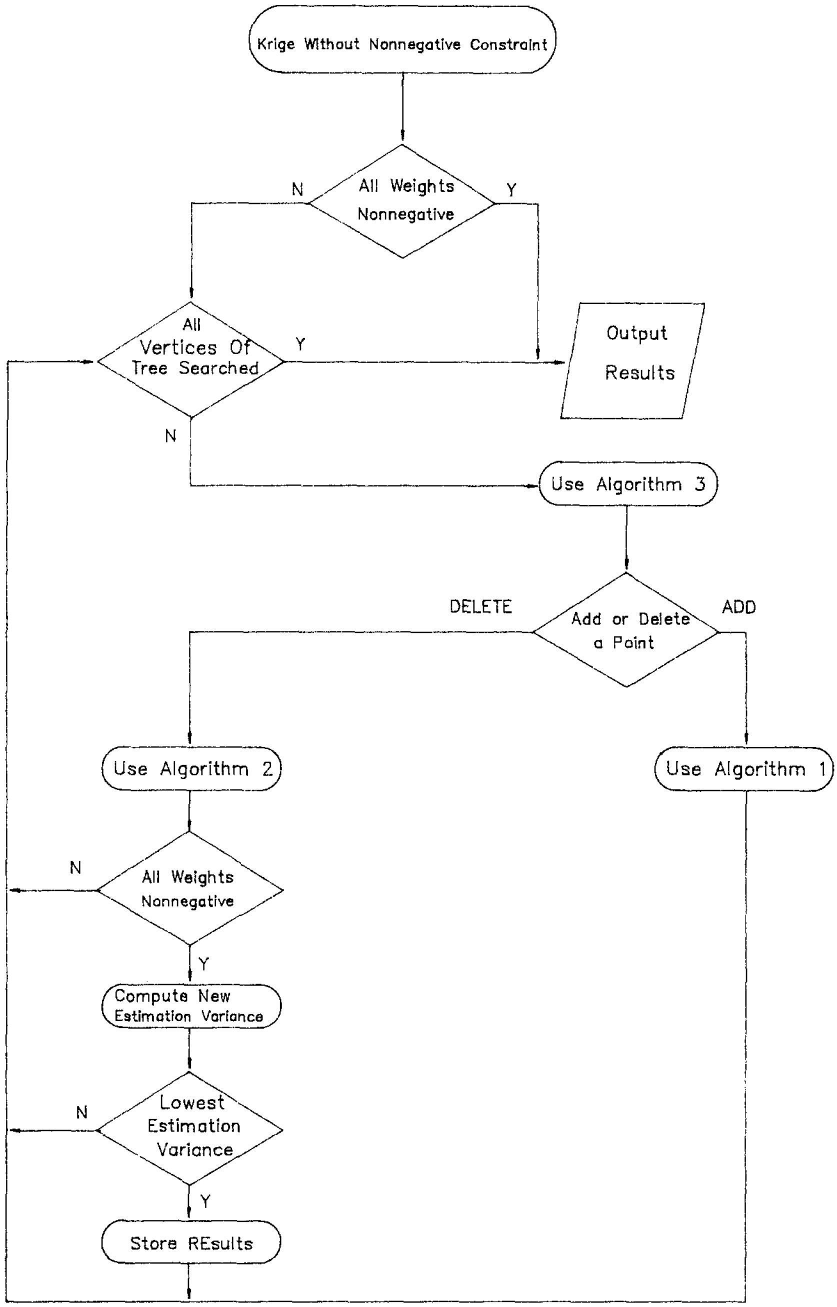

Fig. 2. Simplified flow chart of program ZKRIG.

mation variance among all possible combinations of sample points whose weights are all nonnegative. The developed algorithm utilizes the solution technique to the standard kriging system of equations, usually without nonnegative weight constraint. This means that a computer program of the standard linear kriging system only requires slight modifications in implementation of the nonnegative constraint.

REFERENCES

Baafi, E. Y.; Kim, Y. C.; and Szidarovszky, F., 1986, on Nonnegative Weights of Linear Kriging Estimation: Min. Eng., vol. 38, n. 6, p. 437–442.

Barnes, R. J. and Johnson, T. B., 1984, Positive Kriging, in G. Verly, M. David, A G. Journel, and A. Marechal (Eds.) Geostatistics for Natural Resource Characterization, NATO ASI Series, p. 231–244.

Journel, A. G. and Huijbregts, Ch. J., 1978, Mining Geostatistics: Academic Press, New York, 600 p.

Schapp, W. and George, J. D. St., 1981, On Weights of Linear Kriging Estimator: Trans. Inst. Min. Metall. (Sect. A: Min. Ind.), v. 90, p. A25–A27.

Szidarovszky, F. and Gershon, M., 1984, Multi-Objective Design of Optimal Drill-Hole Sites in the Mining Industry: Proceedings of the 18th APCOM, Institute of Mining and Metallurgy, London, p. 783–790.

Szidarovszky, F. and Yakowitz, S., 1978, Principles and Procedures of Numerical Analysis: Plenum Publishers, New York, 331 p.