Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 5 (8): 1405-1407, 2005
ISSN 1812-5654

Semivariogram Fitting with a Simple Optimizing Algorithm

B. Kushavand, H. Aghababaee, M.J. Mohammadzadeh and B. Alizadeh
Department of Mining Engineering, Sahand University of Technology, Tabriz, Iran

Abstract: One of the most essential and important stages in a geostatistical study is variogram modeling from the experimental semivariogram values. In order to achieve accurate results, it is suggested to find the variogram parameter value by auto fitting technique. In this study a simple and fast algorithm based on least square and weight least square has been presented. In first step, the objective function defined, then the parameters have been limited at an acceptable range and finally a stochastic algorithm has minimized the objected function.

Key words: Semivariogram, weighted least square, stochastic algorithm

INTRODUCTION

Semivariogram or covariance inference provides a set of experimental values $\hat{\gamma}(h_k)$ or $\hat{C}(h_k)$ for a finite number of lags, h_k, k=1,...,K and directions. Continues functions must be fitted to these experimental values, so as to deduce semivariogram or covariance values for any possible lag h required by interpolation algorithms and also to smooth out sample fluctuations[1].

The first stage is determining the type of semivariogram model by comparing the experimental semivariogram diagram with basic semivariogram models diagrams[1,2].

There after, its parameter should be estimated. Here is a sample semivariogram model that is a linear combination of Spherical model and Nugget effect:

$$\gamma(a,c,c_0,h) = \begin{cases} \left[c_0 + c\left(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\left(\frac{h}{a}\right)^3 \right) \right] & h_k \le a \\ c + c_0 & h_k > a \end{cases} \quad (1)$$

If the experimental semivariogram were according to sample semivariogram model (1), the aim is estimate of the best a, c, c_0 values. However, some of these parameters must be at acceptable limits, for example nugget effect (c_0) must be positive, to reach these conditions some variable conventions needed. If there is no limitations to parameters, for some experimental semivariograms, the algorithm will estimate unacceptable values to parameters, like negative nugget effect.

Two common criteria that used in automatic fitting are the least square and weight least square that are sum of squares of differences between experimental $\hat{\gamma}(h_k)$ and model $\gamma(a, c, c_0, h_k)$ semivariogram value.

Objective function: At least square method the purpose is to estimate a, c, c_0 values to minimize the below function:

$$f_{LS}(a,c,c_0) = \sum_{k=1}^{K} \left[\hat{\gamma}(h_k) - \gamma(a,c,c_0,h_k) \right]^2 \quad (2)$$

h_k: The lag distances
$\hat{\gamma}(h_k)$: Experimental semivariogram
$\gamma(a, c, c_0\ h_k)$: semivariogram model.

In addition, at Weight Least Square there is a weight coefficient, $w(h_k)$ at relation (2) as below:

$$f_{WLS}(a,c,c_0) = \sum_{k=1}^{K} w(h_k) \left[\hat{\gamma}(h_k) - \gamma(a,c,c_0,h_k) \right]^2 \quad (3)$$

Such as:

$$w(h_k) = \frac{N(h_k)}{\left[\gamma(a,c,c_0,h_k)\right]^2} \quad (4)$$

$N(h_k)$: The number of pair samples which have h_k distance from each other.

Variable conventions: There are so many different ways to limit a variable to an acceptable range but to reach the maximum speed, the simple conventions will be better. Table 1 shows some of these easiest conventions to use.

Algorithm: There are many ways for solving the constrained nonlinear model[3-5].

But between all these ways, the chosen way should have the following specification:

- Since the functions are so complicate, the method should not need differentiation.

Corresponding Author: Behrang Kushavand, Sahand University of Technology, Sahand New Town, Tabriz, Iran
Fax: +984123444311

Table 1: Constrains and useful conventions

Constrain	Convention
$x \geq 0$	$x' = abs(x)$
	$x' = x^2$
$-1<x<1$	$x' = \sin x$
	$x' = \cos x$
$0<x<1$	$x' = \sin^2 x$
	$x' = \cos^2 x$
$l \leq x \leq u$	$x' = l+(u-l)\sin^2 x$

- Algorithm should contain high convergence speed.
- Designing of the computer program should be easy.
- Algorithm is always converging to global minimization.

From the Most optimization methods, it seems that the Stochastic Algorithms are more suitable for achieving above points. Figure 1 shows the general flowchart of this kind of algorithms[3-5].

One of the useful methods of this kind of algorithms is Random walk method with direction exploitation[5].

In this algorithm, at first the constrained nonlinear programming model has been converted to an abstract non-liner programming model. The convention that has been used at this algorithm is abs(X).

Now, different steps of this algorithm for Weighted Least Square are shown as follow:

- Choose the repetition number of algorithm m and Length Step e.
- Suggest first values to a, c, c_0.
- i =1
- Define new variables:

$$a' = abs(a)$$
$$c' = abs(c)$$
$$c'_0 = abs(c_0)$$

- $f = f_{WLS}(a', c', c'_0)$
- Produce the monotonic stochastic digits: $-1<r_1, r_2, r_3<1$
- $R = \sqrt{r_1^2 + r_2^2 + r_3^2} \Rightarrow \begin{matrix} r_1 = r_1/R \\ r_2 = r_2/R \\ r_3 = r_3/R \end{matrix}$
- $\bar{a} = a + e \times r_1, \bar{c} = c + e \times r_2, \bar{c}_0 = c_0 + e \times r_3$
- $a' = abs(\bar{a}), c' = abs(\bar{c}), c'_0 = abs(\bar{c}_0)$
- $\bar{f} = f_{WLS}(a', c', c'_0)$
- If $\bar{f} > f$, go to step 14
- $a = \bar{a}, c = \bar{c}, c_0 = \bar{c}_0$ and $f = \bar{f}$
- Go to 8
- i = i+1

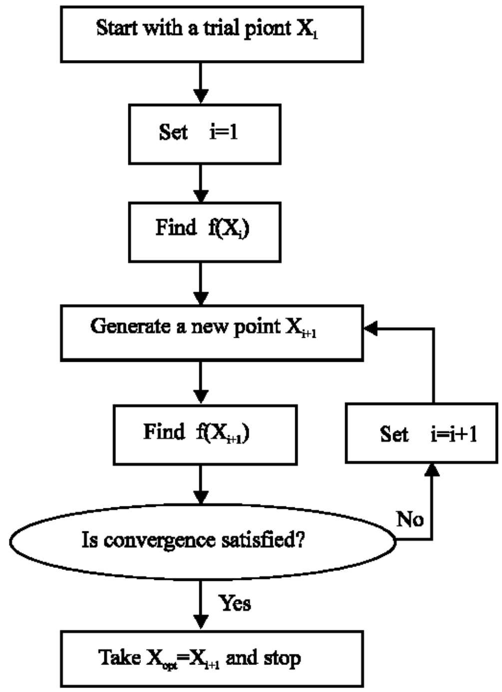

Fig 1: Flowchart of general stochastic algorithm

- If i<m, go to step 6
- Stop, The answers are:
  $abs(a), abs(c), abs(c_0)$
  $f = f_{WLS}[abs(a), abs(c), abs(c_0)]$

Example: The data used at this study were collected by the Swiss Federal Institute of Technology at Lausanne[1]. Table 2 and Fig. 2 show the experimental semivariogram that has been calculated from Zn grades. Figure 2 indicate that the spherical model with nugget effect is more appropriate for Zn semivariogram at this region. Therefore, the objective function is:

$$\text{Min} \quad f_{WLS}(a,c,c_0) = \sum_{k=1}^{K} w(h_k)\left[\hat{\gamma}(h_k) - \gamma(h_k,a,c,c_0)\right]^2$$

St: $$w(h_k) = \frac{N(h_k)}{[\gamma(a,c,c_0,h_k)]^2}$$

and

$$\gamma(h_k,a,c,c_0) = \begin{cases} c_0 + c\left[\frac{3}{2}\frac{h_k}{a} - \frac{1}{2}\left(\frac{h_k}{a}\right)^3\right] & h_k \leq a \\ c_0 + c & h_k > a \end{cases}$$

Table 2 contains the values of $\hat{\gamma}(h_k), N(h_k), h_k$.

By applying the given stochastic algorithm which is presented here and starting by initial answer:

a = 1.4, c = 70, c_0 = 10

Table 2: Experimental semivariogram values for Zn

Lag h_k	Number of pairs N (h_k)	Semivariogram $\hat{\gamma}(h_k)$
0.024	386	11.335
0.104	310	23.170
0.206	498	24.019
0.297	828	37.591
0.391	1288	39.449
0.496	1384	54.084
0.601	1076	44.481
0.700	1510	62.433
0.800	1832	63.876
0.896	1418	85.392
1.002	1862	68.895
1.099	2630	79.701
1.201	1620	97.075
1.298	2524	88.579
1.398	2214	79.835
1.494	2582	75.570
1.597	2186	90.147
1.700	2186	95.732
1.795	2740	76.766

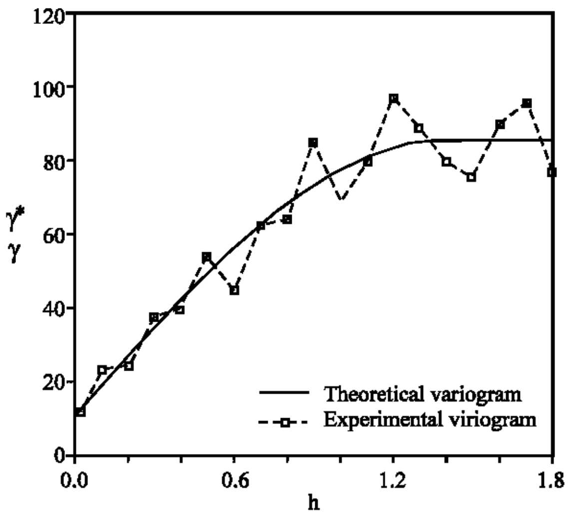

Fig 2: Diagram of experimental and fitted theoretical semivariogram: Model=Spherical + Nugget Effect: a = 1.362, c = 74.802, c_0 = 10.572

With initial parameters:

$$m = 10000,\ e = 0.001$$

The results of the algorithm will be as:

$$a = 1.362,\ c = 74.802,\ c_0 = 10.572$$

The theoretical and experimental semivariogram diagram of the Zn data is plotted in Fig. 2.

CONCLUSIONS

In this study, an appropriate method has been represented for finding the theoretical semivariogram parameters.

The algorithm will be repeated at least m counts, so in applying this stochastic algorithm, correct choosing of m is so important such that with increasing m, optimized estimation for theoretical semivariogram model parameters can be calculated but the running time of program will be more longer. The second parameter of algorithm denoted by e (the length of step) in which decreasing of this parameter result in increase of accuracy and decrease in the rate of algorithm.

In order to achieve accurate result, it is suggested to repeat the algorithm frequently with different initial values. Nested structures can also be modeled by this algorithm, in which the number of parameters that must be found is increased, leading in decrease of the rate of algorithm.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Dr. Gregoire Dubois, Radioactivity Environmental Monitoring group Emissions and Health Unit of ITALY for his constructive advices during the course of this study.

REFERENCES

1. Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. New York, Oxford, pp: 87-107.
2. Jurnel, A.G. and Ch.J. Huijberegts, 1978. Mining Geostatistics. Academic Press, pp: 148-195.
3. Bazara, M.S. and C.M. Shetty, 1979. Nonlinear Programming: Theory and Algorithms. John Wiley and Sons, pp: 332-434.
4. Chong, E.K.P. and S.H. Zak, 2001. An Introduction to Optimization. John Wiley and Sons, pp: 365-433.
5. Rao, S.S, 1984. Optimization: Theory and Application. Wiley Eastern Ltd., New Delhi, pp: 259-265.