

克立格算法的若干思考

王银宏, 彭润民, 王建平, 胡华斌

(中国地质大学, 北京 100083)

[摘要] 克立格法是一种最优、线性(或非线性)、无偏内插估计量的方法。随着信息社会的发展, 克立格法的应用范围越来越广。文中介绍了克立格法相关的计算方法, 并基于 Windows 编写克立格算法流程图, 提出改进克立格算法的若干方法, 包括: ①强化数据预处理; ②将基本滞后距和实验变差函数计算结合起来一并运算; ③防止数据溢出; ④选择快速有效算法求解克立格方程组。

[关键词] 克立格法 算法 流程图 思考

[中图分类号] P628 [文献标识码] A [文章编号] 0495-5331(2004)02-0077-03

0 引言

克立格法最早用于矿产资源储量估算, 在其之前尽管估算方法多种多样, 如多边形法、剖面法、三角形法、以及距离反比法和距离平方反比法等, 但都不同程度的存在问题^[1-3]。为了解决上述储量计算方法的不足, 由南非金矿工程师 D G 克立格提出, 后来由马特隆系统研究并以克立格的名字命名的一种方法, 即克立格法。它是一种最优、线性(或非线性)、无偏内插估计量的方法。本文基于 Windows 编写了克立格算法流程图, 并提出改进克立格算法的若干方法。

1 克立格法相关公式的计算方法

1.1 实验变差函数的计算方法

1.1.1 无距离加权^[4-6]

$$\gamma^*(kh) = \frac{1}{2N(kh)} \sum_{i=1}^{N(kh)} [Z(x_i) - Z(x_i + kh)]^2 \quad (k = 1, 2, \dots, K)$$

其中: h —基本滞后距; $N(kh)$ —距离为 kh 时数据对的数目。

当地质量的位置均落在基本滞后距组成的网格节点上, 可使用无距离加权的计算方法^[6]。

1.1.2 距离加权^[4-6]

$$\gamma^*(kh) = \frac{1}{2 \sum_{i=1}^L \sum_{j=1}^{M_i} d_{ij}^{(k)}} \sum_{i=1}^L \sum_{j=1}^{M_i} d_{ij}^{(k)} [Z(x_i) - Z(x_i + kh)]^2$$

 $h_{ij}]^2$

其中: M_i —距 x_i 的距离为 h_{ij} , 且在一定的距离误差限内的数据对数目;

L —可以组成数据对的 i 的变化范围;

$d_{ij}^{(k)}$ —在一定的距离误差限内的距离权系数。

即:

$$d_{ij}^{(k)} = 1 - \left[\frac{h_{ij} - k \cdot h}{h/2} \right]^2 \quad k = 1, 2, \dots, K$$

当地质量的位置不均落在基本滞后距组成的网格节点上, 应该使用距离加权的计算方法^[6]。

1.2 理论变差函数拟合的计算方法

1.2.1 直接法

用实验变差函数开始两点的连线与纵坐标轴的交点的纵坐标值作为理论变差函数的块金值 C_0 ; 以实验变差函数倒数第二点的值作为拱高值 $C + C_0$; 其对应的横坐标值作为变程值 a 。

1.2.2 公式法^[4-6]

根据加权多项式回归计算理论变差函数的块金值、拱高值和变程值:

$$\begin{cases} C_0 = b_0 \\ a = \sqrt{\frac{-b_1}{3b_2}} \\ C = \frac{2b_1}{3} \sqrt{\frac{-b_1}{3b_2}} \end{cases} \quad \begin{cases} b_1 = \frac{l_{11}l_{22} - l_{21}l_{12}}{l_{11}l_{22} - l_{12}^2} \\ b_2 = \frac{l_{21}l_{11} - l_{11}l_{12}}{l_{11}l_{22} - l_{12}^2} \\ b_0 = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 \end{cases}$$

其中: $y = \gamma(h)$, $x_1 = h$, $x_2 = h^3$;

[收稿日期] 2003-03-10; [修订日期] 2003-04-21; [责任编辑] 曲丽莉。

[基金项目] 国家 973 项目(编号: 199904215) 和国家 973 项目专题(编号: 2001CB409807) 联合资助。

[作者简介] 王银宏(1973年-), 男, 2000年毕业于中国科学技术大学, 获学士学位, 在读博士生, 主要从事计算机和矿床地质研究工作。

$$l_{ik} = \sum_{i=1}^n N_i x_j x_{ki} - \frac{1}{N} \left(\sum_{i=1}^n N_i x_{ji} \right) \left(\sum_{i=1}^n N_i x_{ki} \right) \quad (j, k = 1, 2)$$

$$l_{jj} = \sum_{i=1}^n N_i x_j x_i - \frac{1}{N} \left(\sum_{i=1}^n N_i x_{ji} \right) \left(\sum_{i=1}^n N_i x_{ji} \right) \quad (j = 1, 2)$$

1.3 克立格估值和方程组

1.3.1 普通克立格估值

$$z_k^* = \sum_{i=1}^n \lambda_i z_i$$

表示任一待估块段 V 的真值 Z_V 的估计值 Z_k^* 是估计邻域内 n 个信息值 Z_i 的线形组合, λ_i 为估值权系数。

1.3.2 普通克立格方程组

$$\begin{cases} \sum_{i=1}^n \lambda_i c(x_i, x_j) - \mu = c(x, x_j) & j = 1, 2, \dots, n \\ \sum_{i=1}^n \lambda_i = 1 \end{cases}$$

其中: $c(x_i, x_j)$ 一点 x_i 与点 x_j 之间的协方差函数值; μ 一求条件极值的拉格朗日乘数。

以上列出的是普通克立格方程组, 另外, 泛克立格、指示克立格、对数克立格等计算方法以及以上所有方法的详细解释请参见相关资料^[1-8], 它们的程序流程图如图 1。

2 改进克立格算法的建议

1) 加强数据预处理, 增强程序报错功能。在数据预处理阶段进行相关性、统计特征及特异值分析, 确定合理的数学模型。在所处理的原始地质数据中, 尤其对重复样品和同一坐标位置有两个以上的

品位值进行报错, 并提示用户检查自己的原始地质数据。如果不这样, 计算机所计算的各坐标方向的基本滞后距趋于零, 导致克立格估值不正确。

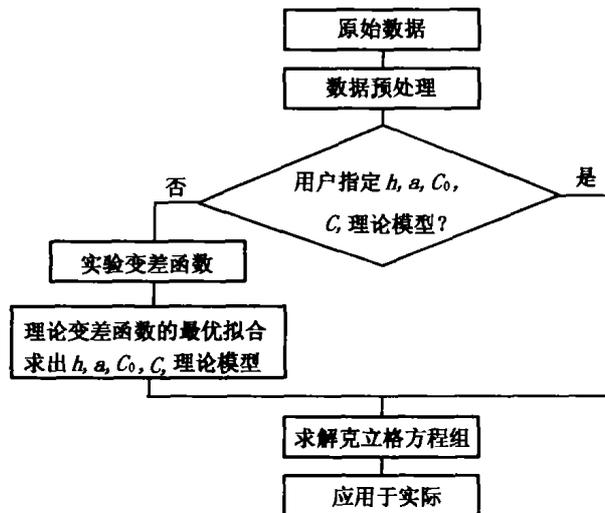


图 1 克立格程序流程图

2) 将基本滞后距和实验变差函数计算结合起来一并运算, 提高程序运行效率。各坐标方向的基本滞后距是在角度误差限内, 所有数据中该坐标方向任意两点间的最短距离; 而计算实验变差函数的 $N(kh)$ 时, 要多次用到任意两点间的距离。因此可以动态申请一个数组, 将第一次计算的结果存放在该数组中, 而后计算 $N(kh)$ 时需要的距离直接在数组中取存, 这样大大节省程序的运行时间。以下是基于 Visual C++ 平台使用 C++ 语言编写的代码片段^[9]:

```

double dx, dy, tempx = 1.0E + 38, tempy = 1.0E + 38;
double ** disMatrix;
disMatrix = NULL;
disMatrix = new double* [m_inNum];
for(i=0; i < m_inNum; i++)
{ disMatrix[i] = new double[m_inNum]; } // 动态申请数组 disMatrix
for(i=0; i < m_inNum; i++)
{ for(j=i+1; j < m_inNum; j++)
{ dx = fabs(m_OriDat[0][i] - m_OriDat[0][j]); // x 轴两点间的距离
dy = fabs(m_OriDat[1][i] - m_OriDat[1][j]); // y 轴两点间的距离
disMatrix[i][j] = sqrt(dx * dx + dy * dy); // disMatrix 上三角阵存放第 i, j 两点间距离
if(dx == 0) // disMatrix 下三角阵存放第 i, j 两点间所成角度
disMatrix[j][i] = 1.5707963; // π/2.0 ≈ 1.5707963, 与 x 轴夹角 90 度
else
disMatrix[j][i] = atan(dy/dx); // 与 x 轴夹角
if(disMatrix[j][i] <= AngError && disMatrix[i][j] != 0.0 && disMatrix[i][j] < tempx)
tempx = disMatrix[i][j]; // 计算 x 坐标方向基本滞后距
}
}
  
```

```

else
if( (1.5707963 - disMatrix[j][i]) <= AngError && disMatrix[i][j] != 0.0 && disMatrix[i][j] < tempy)
tempy = disMatrix[i][j];          //计算 y 坐标方向基本滞后距
} } .....
if( disMatrix)                    //释放 disMatrix
{ for(i=0; i < m_inNum; i++) delete disMatrix[i]; delete disMatrix; }

```

3) 避免大数“吃”小数,防止数据溢出。在使用公式法拟合理论变差函数时,要注意一个很大的数与一个很小的数相运算,同时计算机中浮点型变量在内存中占4个字节,变量数的范围为 $-2^{31} \sim 2^{31} - 1$,当计算的数据量很大时,很可能产生数据溢出。在加权多项式回归中计算 b_1 和 b_2 时,开始设置一个阈值,当 b_1 和 b_2 中的分子或分母大于该值时,分子或分母同除以一个数,以防止数据溢出。

4) 选择快速有效算法求解克立格方程组。求解高维方程组是克立格估值的一个重要内容,因此建议使用高效的列主消元法^[1-3],并将其编写成一个子模块,供系统调用。其算法如下^[10]:

设二维数组A储存增广矩阵(A,b)的元素。列主元消去法算法为:

1) 消去过程

对 $k = 1, 2, \dots, n - 1$,作:

- ① 在 $a_{ik} (i = k, \dots, n)$ 中选主元,设为 a_{rk} ;
- ② 若 $|a_{rk}| < \varepsilon$,则消去法失败,停止计算;
- ③ 若 $r \neq k$,则对 $j = k, \dots, n + 1$ 作 $a_{kj} \leftrightarrow a_{rj}$;
- ④ 对 $i = k + 1, \dots, n$,计算

$$a_{ik} \leftarrow \frac{a_{ik}}{a_{rk}}$$

对 $j = k + 1, \dots$,计算 $a_{ij} \leftarrow a_{ij} - a_{ik} a_{rj}$

2) 回代过程

- ① 若 $|a_{nn}| < \varepsilon$,则消去法失败,停止计算;
- ② 对 $k = n, n - 1, \dots, 1$,计算

$$a_{k,n+1} \leftarrow \frac{a_{k,n+1} - \sum_{j=k+1}^n a_{kj} a_{j,n+1}}{a_{k,k}}$$

最后,解向量存放在数组A的最后一列。

CONSIDERATIONS OF THE KRIGING ALGORITHM

WANG Yin-hong, PENG Run-min, WANG Jian-ping, HU Hua-bin
(China University of Geosciences, Beijing 100083)

Abstract: The Kriging method is the best, linear (or non-linear), and unbiased estimate method. With the instant development of society, the Kriging method is used more and more widely. In this paper, the author introduces correlative computing method of Kriging, compiles flow diagram of the Kriging algorithm based on Windows, and puts forward methods of improving the Kriging algorithm. The methods include: (1) intensifying data preprocessing; (2) calculating the basic lag distance and the experimental variogram together; (3) avoiding data overflow; (4) choosing fast and effective algorithm to solve the Kriging equation group.

Key words: Kriging method, algorithm, flow diagram, considerations

3 总结

在应用克立格法进行估值时,要根据具体情况和研究目的选择相应的算法。尤其当处理的数据量很大而对其精度要求又不是很高时,要选择省时的快速算法,并尽量减少克立格方程组的维数,在尽可能短的时间内计算出结果。最后还要注重对克立格估算结果中可能出现的负值进行研究。

感谢翟裕生院士通读全文,并感谢中国地质调查局发展中心向运川高级工程师在程序实现过程中给予的大力支持与帮助。

[参考文献]

- [1] 侯景儒,黄竞先.地质统计学在固体矿产资源/储量分类中的应用[J].地质与勘探,2001,37(6):61-66.
- [2] 侯景儒,尹镇南,李维明,等.实用地质统计学[M].北京:地质出版社,1998.
- [3] 杨晓雷,胡海生.克立格法在采场品位估计中的应用[J].地质与勘探,1996,32(6):28-31.
- [4] 孙洪泉.实用地质学程序集[M].北京:地质出版社,1997.
- [5] 王仁铎,胡光道.线性地质统计学[M].北京:地质出版社,1988.
- [6] 石广仁.地学中的计算机应用新技术[M].北京:石油工业出版社,1999.
- [7] C. V. Deutsch. A. G. Journel, GSLIB, Geostatistical Software Library and User's Guide[M]. Oxford University press, 1992.
- [8] 胡小荣,俞茂宏,用线性规划法求解克立格估值权系数的研究[J].地质与勘探,2001,37(3):53-57.
- [9] Visual C++ 6.0 技术内幕(希望图书创作室译)[M].北京:北京希望电子出版社,2001.
- [10] 黄其明,廖鸿志,数值计算方法及其程序设计[M].昆明:云南大学出版社,1995.