

克里金估值中条带效应的校正

李少华*^① 王勇标^{①②} 李 君^①

刘显太^③ 王 军^③ 龚蔚青^③

(^①长江大学地球科学学院,湖北荆州 434023; ^②中海石油(中国)有限公司湛江分公司,广东湛江 524057; ^③中国石化胜利油田分公司地质科学研究院,山东东营 257015)

李少华,王勇标,李君,刘显太,王军,龚蔚青. 克里金估值中条带效应的校正. 石油地球物理勘探, 2012, 47(6): 978~983

摘要 矿业和石油勘探行业中的待分析数据多呈条带状分布,采用克里金插值易出现两个端点数据权值偏大的情况,即“条带(String)效应”。本文在传统克里金方法的基础上,根据目标变量与条件数据之间的距离对权值进行校正,使离待估点越近的数据点的权值越大,越远的数据点的权值越小。这种采用距离约束的克里金方法(DCSK)校正了简单克里金算法中的条带效应,利用 FORTRAN 编程语言实现了该方法并进行了实例验证。试验结果证实 DCSK 方法不仅消除了传统克里金方法引起的条带效应,而且其估计方差基本不变或略有增加,从而使得估计结果更接近实际情况。

关键词 条带效应 简单克里金 距离约束克里金 权值

中图分类号: TE319 **文献标识码**: A

1 引言

“克里金技术”起源于矿业,从其诞生至今已有 50 多年的历史,现还处在持续发展和完善之中。该技术的应用范围不断扩大,在地质学、水文学、土壤科学、气象学、农业、遥感、石油、生态、海洋、资源与环境等领域得到了广泛应用^[1~3]。

鉴于矿业和石油行业中的待分析数据主要是通过钻孔或钻井获取,因此数据常呈条带状分布,许多人对此做了大量的研究^[4~9]。在数据呈条带状分布时,克里金插值会出现“条带(String)效应”,即在有限的条带两端的数据通常会获得偏大的权值,这样会导致插值结果的不合理。为了消除条带效应,本文采用距离约束法对克里金估值得到的权值进行校正,在估值方差基本不增加时保证了估值的合理性,并以计算实例对该改进方法进行阐述。

2 简单克里金原理^[10]

从统计的意义上说,克里金插值法是从变量相

关性和变异性出发,在有限区域内对区域化变量的取值进行无偏最优估计。而简单克里金法是在区域化变量 $Z(x)$ 的数学期望为已知的情况下建立的一种插值法。由于已知 $E[Z(x)] = E[Z(x+h)] = m$, 若令 $Y(x) = Z(x) - m$, 则 $E[Y(x)] = E[Z(x) - m] = 0$, 其协方差 $E[Y(x), Y(y)] = \text{Cov}(x, y)$, 那么对 $Z(x_0)$ 的估计可转化为对 $Y(x_0)$ 的估计,只要求出 $Y^*(x_0)$ 就可以得到 $Z^*(x_0)$ 。即

$$Y^*(x_0) = \sum_{i=1}^n \lambda_i Y(x_i) \quad (1)$$

估计方差

$$\sigma_E^2 = E[Y(x_0) - Y^*(x_0)]^2 = \text{Cov}(x_0, x_0) - 2 \sum_{i=1}^n \lambda_i \text{Cov}(x_i, x_0) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov}(x_i, x_j) \quad (2)$$

为使 σ_E^2 达到最小,按求极值原理,对 λ_i 求偏导数,即

$$\frac{\partial \sigma_E^2}{\partial \lambda_i} = -2 \text{Cov}(x_i, x_0) + 2 \sum_{j=1}^n \lambda_j \text{Cov}(x_i, x_j) = 0 \quad (3)$$

从而可得简单克里金方程组

$$\sum_{j=1}^n \lambda_j \text{Cov}(x_i, x_j) = \text{Cov}(x_i, x_0) \quad (4)$$

* 湖北省荆州市长江大学地球科学学院, 434023。Email: yongbiao0110@163.com

本文于 2011 年 9 月 20 日收到,最终修改稿于 2012 年 10 月 8 日收到。

本项研究受国家科技重大专项(2011ZX05011-001)和国家自然科学基金项目(41272136)联合资助。

对其求解,即可得简单克里金权系数 $\lambda_i (i=1, \dots, n)$, 同时算出简单克里金估计方差

$$\sigma_k^2 = \text{Cov}(x_0, x_0) - \sum_{i=1}^n \lambda_i \text{Cov}(x_i, x_0) \quad (5)$$

3 条带效应

在矿业和石油行业中,沿钻孔或钻井获取的数据通常是最主要的源数据。这些数据呈条带状分布,在应用克里金插值时经常遇到以下两种情况:①条带状数据被地质体或地层的边界截断成有限的条带数据(图 1a);②条带状数据被局部搜索范围截断成有限的条带数据(图 1b)。

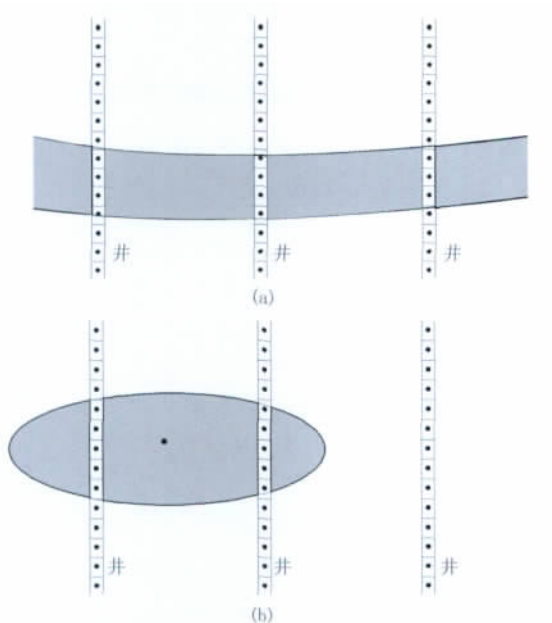


图 1 有限的条带数据

(a) 数据被地层边界截断; (b) 数据被搜索范围截断

图 1a 和图 1b 中的这种沿井得到的数据叫做条带数据。当用克里金进行插值时,如果使用的数据是有限的条带数据,那么经常看到在条带数据外围的数据比其他位置的数据具有更高的权值。图 2 显示的这些权值就体现了“条带效应”^[11,12],这些权值是由克里金估值得到的,估值时采用球状变差函数模型,变程等于估计点到样品点的距离,块金效应为 0。从图中可明显地看到:条带数据中离估计点最远两端的样品点获得的权值偏大,而离估计点较近的位置的权值较小。这些权值在理论上说是正确的,产生的原因主要是因为克里金估值中的遍历性假设,也就是认为最外端的数据点与其他数据点相比

冗余度更低,因而权值较大。这是不符合实际情况的,特别是当数据的分布不平稳、端点数据明显大于或小于中间数据时,需要对条带效应进行校正。

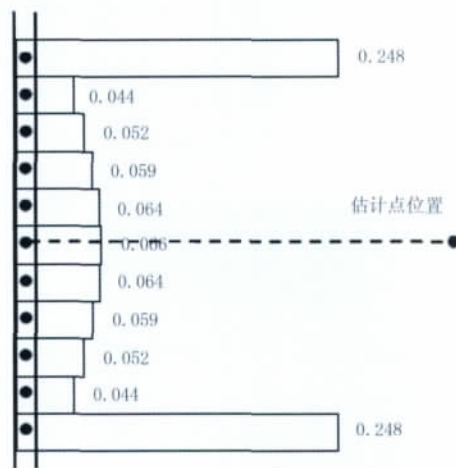


图 2 传统克里金引起的条带效应

4 简单克里金算法的改进

4.1 距离约束克里金法(DCSK)

为了校正条带效应,就要对权值进行较合理的约束。尤其是当权值被分配给一个与未知点位置的距离有一定顺序关系的条带数据时,应该采用这样的约束:使这个条带数据中与估计位置距离最近数据点获得最大的权值,第二近数据点获得次大的权值,以此类推,离估计位置最远的数据点被分配一个最小的权值。此即距离约束克里金的主要思想^[12]。

假设在一个条带数据中有相邻的 n 个位于位置 $u_i (i=1, 2, \dots, n)$ 处排成一列的数据点,采用距离约束克里金方法对未知位置 u_0 处的未取样本 Z 的变量值进行估计,其具体方法如下。

对条带数据 $Z(u_i) (i=1, \dots, n)$, 在待估点 $Z(u_0)$ 处的距离约束简单克里金(DCSK)的估计量是

$$Z_{\text{DCSK}}^*(u_0) = \sum_{i=1}^n \lambda_{\text{DCSK},i}(u_0) Z(u_i) + \left[1 - \sum_{i=1}^n \lambda_{\text{DCSK},i}(u_0) \right] m \quad (6)$$

式中 $\lambda_{\text{DCSK},i}(u_0) = \lambda_{\text{DCSK},i}$ 是第 i 个样本点的权值,由最小化方差 σ_{est}^2 可求得

$$\min_{\lambda_{\text{DCSK},i}} \sigma_{\text{est}}^2 = \sigma^2 - 2 \sum_{i=1}^n \lambda_{\text{DCSK},i} \text{Cov}[Z(u_i), Z(u_0)] + \sum_{i=1}^n \sum_{j=1}^n \lambda_{\text{DCSK},i} \lambda_{\text{DCSK},j} \text{Cov}[Z(u_i), Z(u_j)] \quad (7)$$

其中权值有如下约束:若 $d_i < d_j$, 则 $\lambda_{\text{DCSK},i} > \lambda_{\text{DCSK},j}$ ($i=1, \dots, n$)。

4.2 DCSK 算法实现

对于上面给出的 DCSK 的最小化问题,可通过简单的非线性、非约束最优化求解。有关 DCSK 算法中计算距离约束的权值和估计方差的方法如下。

对最小化式(7)中的权值进行变换,有

$$\min_{\xi} \sigma_{\text{est}}^2 = \sigma^2 - 2 \sum_{i=1}^n \lambda_{\text{DCSK},i}(\xi) \text{Cov}[X_i, X(u)] + \sum_{i=1}^n \sum_{j=1}^n \lambda_{\text{DCSK},i}(\xi) \lambda_{\text{DCSK},j}(\xi) \text{Cov}(X_i, X_j) \quad (8)$$

式中: ξ 是由解最小化问题得到的一个参数向量;DCSK 权值 $\lambda_{\text{DCSK},i}(\xi)$ ($i=1, \dots, n$)是 ξ 中第 $(n-k_i+1)$ 个最大的元,且 k_i 是数据链中第 i 个数据点到估计点的距离 d_i 的队列, d_i 是所有距离经过升序排列后得到的向量 (d_1, d_2, \dots, d_n) 中的第 i 个元素。

DCSK 方法实现的具体步骤如下:

(1)计算已知点到待估点的一组距离 d_1, d_2, \dots, d_n ,并将这些距离按升序排列,令 k 为排序后的距离的脚标队列;

(2)利用简单克里金(SK)求出一组权值 λ_i ,并对其进行降序排列,令 $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ 且 $\lambda_1 > \lambda_2 > \dots > \lambda_n$;

(3)设 $x_0 = (a_1, a_2, \dots, a_n)$,其中 $a_i = \lambda_{n-k(i)+1}$,且 $k(i)$ 是步骤(1)中 k 的第 i 个元素;

(4)用梯度函数计算 x_0 的梯度 g_0 ;

(5)将 g_0 代入到改进的拟牛顿法(DFP)中,将 x_0 当做 DFP 的初始点,得到 x'_0 ;

(6)将步骤(5)的 x'_0 中的元按升序排列,使得 $x'_0 = (b_1, b_2, \dots, b_n)$,且 $b_1 < b_2 < \dots < b_n$;再令 $x_1 = (b'_1, b'_2, \dots, b'_n)$,且 $b'_1 = b_{n-k(i)+1}$;

(7)计算 x_1 与 x_0 对应函数值差的绝对值,即 $\epsilon = |\sigma_{\text{est}}^2(x_1) - \sigma_{\text{est}}^2(x_0)|$;若 ϵ 满足精度要求则结束循环,将 x_1 赋值给 x_0 ;否则,继续下一步;

(8)将 x_1 赋值给 x_0 ,从步骤(4)开始重复计算,直到 ϵ 满足精度要求为止,这时的 x_1 中的元即是所求的 DCSK 权值。

上述 DCSK 算法过程中用到的改进拟牛顿法(DFP)^[13~15]的计算过程是:给定初始值 x_0 ,为求最优值 x_1 ,通过下降方向 $S_{k+1} = -H_{k+1} \nabla f(x_{k+1})$ 得到迭代公式 $x_{k+1} = x_k + S_k$,当 $\nabla f(x_{k+1})$ 的 L_2 范数小于或等于给定的控制误差时, $x_1 = x_{k+1}$ 即为最优

值; H_{k+1} 可通过 DFP 修正公式

$$H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{S_k S_k^T}{y_k^T S_k} \quad (9)$$

求得。其中 $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$, $S_k = x_{k+1} - x_k$,计算时可置 $H_1 = I$ (单位矩阵)。

5 计算实例

通过一个简单实例,对距离约束简单克里金的计算效果进行验证。假设有 7 个呈条带状分布的样品点,分别位于 $(1,0), (2,0), (3,0), (4,0), (5,0), (6,0), (7,0)$,对处于 $(4,20)$ 位置的点进行估计,估值时用到基台值为 1、变程为 40 的球状变差函数模型,得到简单克里金和距离约束简单克里金的权值及对应的估计方差(图 3)。

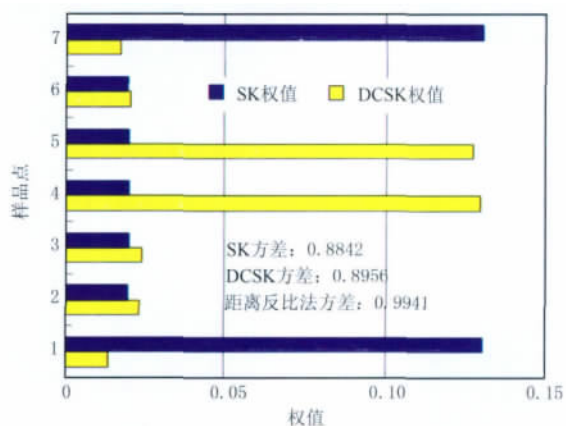


图 3 块金效应为 0 时的权值和估计方差

从图 3 可看到:根据简单克里金法(SK)得到的权值具有明显的两端大、中间小的条带效应,而根据改进的距离约束简单克里金法(DCSK)得到的权值消除了这种条带效应;DCSK 的估计方差比 SK 的估计方差稍大,但比距离加权反比法的估计方差小。因此 DCSK 方法不仅消除了 SK 的条带效应,且具有较高的估计精度。对于普通克里金可采用类似方式处理条带效应。另外,还针对球状模型和指数模型不同块金值情况下 DCSK 与 SK 的权值进行了对比。从表 1 中可看出块金效应越大时产生的条带效应越小,不同理论变差函数模型计算的权值差异不大。

图 4 为改变条件数据个数时采用 DCSK 对待估点(位于中间)进行估值得到的条件数据权值分布,从中可看出 DCSK 具有很好的稳定性,在不同条件数据个数时权值分配的趋势保持一致,对两端

表 1 不同块金效应得到的克里金权值

样品点	球 状 模 型						指 数 模 型					
	块金值=0		块金值=0.2		块金值=0.8		块金值=0		块金值=0.2		块金值=0.8	
	SK	DCSK	SK	DCSK	SK	DCSK	SK	DCSK	SK	DCSK	SK	DCSK
(1,0)	0.1305	0.0137	0.0650	0.0377	0.0316	0.0307	0.1035	0.0075	0.0583	0.0213	0.0242	0.0223
(2,0)	0.0196	0.0233	0.0480	0.0486	0.0311	0.0311	0.0143	0.0193	0.0356	0.0363	0.0231	0.0231
(3,0)	0.0200	0.0241	0.0395	0.0489	0.0308	0.0311	0.0145	0.0199	0.0259	0.0373	0.0225	0.0231
(4,0)	0.0201	0.1299	0.0370	0.0647	0.0307	0.0316	0.0146	0.1028	0.0233	0.0571	0.0223	0.0242
(5,0)	0.0200	0.1273	0.0395	0.0641	0.0308	0.0316	0.0145	0.0992	0.0259	0.0560	0.0225	0.0242
(6,0)	0.0196	0.0206	0.0480	0.0387	0.0311	0.0308	0.0143	0.0144	0.0356	0.0280	0.0231	0.0224
(7,0)	0.1305	0.0172	0.0650	0.0382	0.0316	0.0308	0.1035	0.0104	0.0583	0.0232	0.0242	0.0224

条件数据的权值都能进行修正,对于中间位置以外待估点的权值分配也能起到较好的校正作用。如图 5a和图 5b 所示,待估点的位置与条件数据的某一端点平行(垂向),DCSK 对另一端点条件数据的权值能很好修正;图 5c 中待估点不在中间也不在两端的情况下 DCSK 对两个端点条件数据的权值也进行了修正。

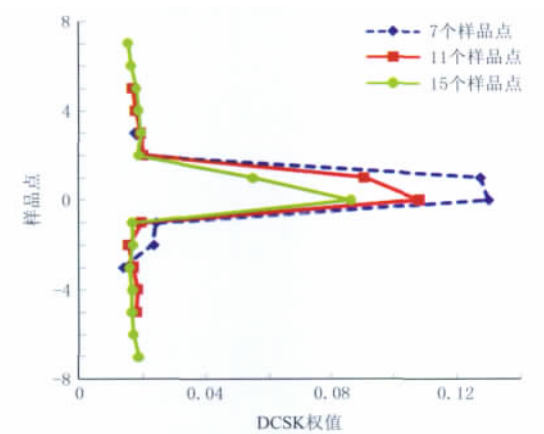


图 4 不同条件数据个数时的权值

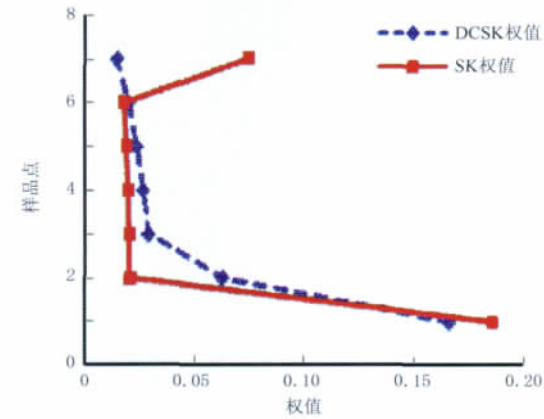


图 5a 待估点位于(1,20)处时估计的权值

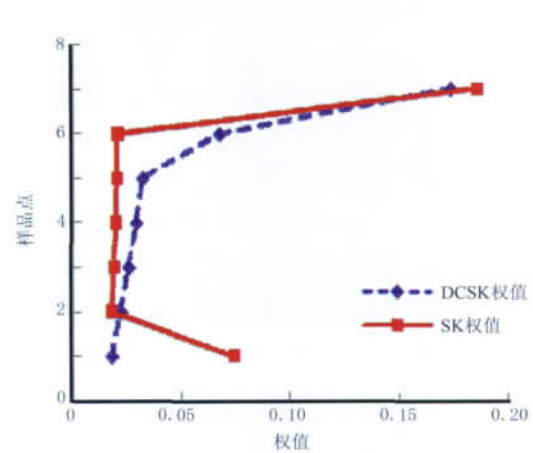


图 5b 待估点位于(7,20)处时估计的权值

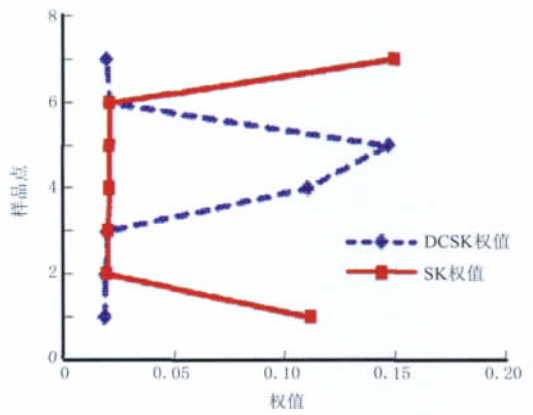


图 5c 待估点位于(5,20)处时估计的权值

6 数值模拟

假定有一组呈条带状分布的数据(图 6),分别用 DCSK 和 SK 法对其做数值模拟,所得结果如图 7所示。估值时用基台值为 1、变程为 40 的球状变差函数模型;网格划分为 15×15 。比较图 7a 和

图 7b 可看出:用 DCSK 法模拟时处于区域靠中间处的数值比用 SK 法模拟的值偏大,这是因为采用 SK 法进行插值时,因数据呈条带分布,使得两端低值条件数据点占据了偏大的权值,进而导致位于中间部分的估值偏小。图 7c 是图 7a 和图 7b 模拟结果的差异图,从图中同样可看出:靠近条件数据和两端的区域,这两种方法的模拟结果差异较小;而位于中上部和中下部的区域,DCSK 法模拟结果数值比 SK 法的大,这是缘于待估位置离条件数据越近条带效应就越小^[12]。因此,采用 DCSK 法可对克里金插值中的条带效应进行有效校正。

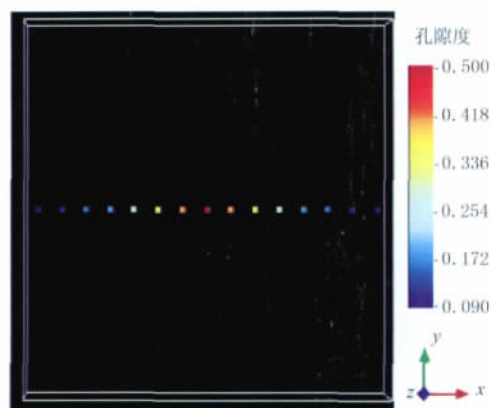


图 6 条带数据实例

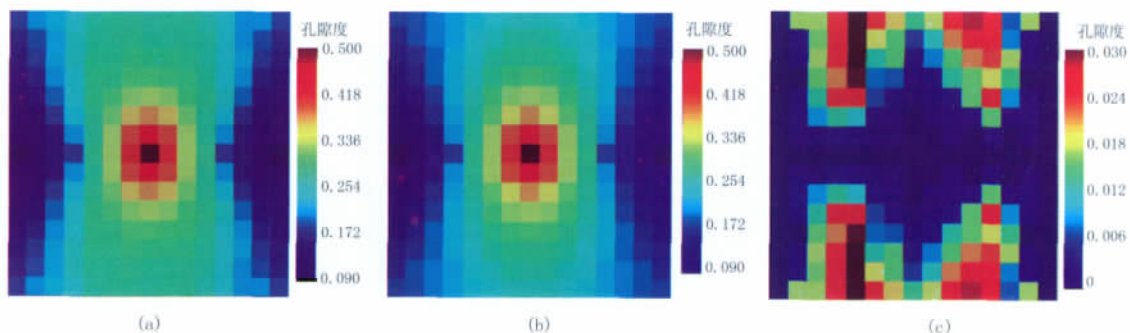


图 7 数值模拟结果

(a)DCSK 法模拟结果;(b)SK 法模拟结果;(c)DCSK 法与 SK 法模拟结果的差异图

7 结论

当数据呈有限条带状分布时,利用简单克里金插值法易产生条带效应,导致条件数据中两个端点获得明显偏大的权值。基于此,本文采用距离约束下的克里金插值法。通过待估点与条件数据之间的距离对克里金得到的权值进行校正,使得距离越远的点权值越小,从而消除了条带效应。距离约束克里金法是一种无偏估值,相对于传统克里金方法,其估值方差基本不变或略有增加。克里金插值中的条带效应随着变差函数模型中块金值的增大而减弱,不同变差函数理论模型之间差异较小。距离约束下的克里金插值法具有较好的稳定性,在不同条件数据个数时 also 具有很好的权值修正效果。

参考文献

[1] 侯景儒. 中国地质统计学(空间信息统计学)发展的回顾与前景. 地质与勘探, 1997, 33(1): 53~58
Hou Jingru. Review and prospect of geo-statistics (Spatial-Information Statistics) development in Chi-

na. *Geology and Prospecting*, 1997, 33(1): 53~58
[2] 肖斌, 赵鹏大, 侯景儒. 现代地质统计学的新进展. 世界地质, 1999, 18(3): 81~91
Xiao Bin, Zhao Pengda, Hou Jingru. New development of modern geo-statistics. *Global Geology*, 1999, 18(3): 81~91
[3] 杜德文, 马淑珍, 陈永良. 地质统计学方法综述. 世界地质, 1995, 14(4): 79~84
Du Dewen, Ma Shuzhen, Chen Yongliang. Over-view of geo-statistical method. *Global Geology*, 1995, 14(4): 79~84
[4] 李庆谋. 多维分形克里格方法. 地球科学进展, 2005, 20(2): 248~256
Li Qingmou. Multi-fractal Kriging interpolation method. *Advances in Earth Science*, 2005, 20(2): 248~256
[5] 肖克炎, 张晓华, 王全明等. 应用改进的克里格法分离重力区域异常与局部异常. 物探化探计算技术, 1995, 17(2): 19~25
Xiao Keyan, Zhang Xiaohua, Wang Quanming et al. Separation of regional field and local field by improved Kriging method. *Computing Techniques for Geophysical and Geochemical Exploration*, 1995, 17(2): 19~25
[6] 高美娟, 朱庆忠, 张淑华. 利用贝叶斯—克里金估计技术进行储层参数预测. 石油地球物理勘探, 1999, 34(4): 390~397
Gao Meijuan, Zhu Qingzhong, Zhang Shuhua. Reser-

- voir parameter prediction using Bayes-Kriging estimation technique. *OGP*, 1999, 34(4): 390~397
- [7] 李君, 李少华, 毛平等. Kriging 插值中条件数据点个数的选择. *断块油气田*, 2010, 17(3): 277~279
Li Jun, Li Shaohua, Mao Ping et al. Choosing the number of conditional data in Kriging interpolation. *Fault-Block Oil & Gas Field*, 2010, 17(3): 277~279
- [8] 汪保, 孙秦. 改进的 Kriging 模型的可靠度计算. *计算机仿真*, 2011, 28(2): 113~116
Wang Bao, Sun Qin. Structural reliability computation based on Kriging model. *Computer Simulation*, 2011, 28(2): 113~116
- [9] 牛文杰, 朱大培, 陈其明. 滑动邻域克里金插值法的改进. *计算机辅助设计与图形学学报*, 2001, 13(8): 752~756
Niu Wenjie, Zhu Dapei, Chen Qiming. Improvement of moving neighborhood Kriging method. *Journal of Computer-Aided Design & Compute Graphics*, 2001, 13(8): 752~756
- [10] 王仁铎, 胡道光. 线性地质统计学. 北京: 地质出版社, 1988: 123~132
- [11] Deutsch. Kriging with Strings of Data. *Mathematical Geology*, 1994, 26(5): 623~638
- [12] Olena Babak. *Theoretical and Practical Improvements to Geostatistical Estimation and Simulation* [D]. Canada: University of Alberta, 2008
- [13] 璞定国, 王华, 田蔚文. 带非精确线搜索的调整搜索方向 DFP 算法(英文). *运筹学学报*, 2004, 8(3): 1~9
Pu Dingguo, Wang Hua, Tian Weiwen. The revised DFP algorithm without exact line search. *Operations Research Transactions*, 2004, 8(3): 1~9
- [14] 徐大川. 无约束优化 DFP 算法的全局收敛性. *烟台师范学院学报(自然科学版)*, 1997, 13(2): 106~108
Xu Dachuan. Global convergence of DFP algorithm for unconstrained optimization. *Yantai Teachers University Journal (Natural Science)*, 1997, 13(2): 106~108
- [15] 濮定国. 修改的 DFP 算法. *应用数学学报*, 1990, 13(1): 118~121
Pu Dingguo. Modified DFP algorithm. *Acta Mathematicae Applicatae Sinica*, 1990, 13(1): 118~121

(本文编辑: 朱汉东)