

混合加速遗传算法在流域模型参数优化中的应用

杨晓华¹, 陆桂华², 郇建强³

(1. 河海大学数学物理系, 江苏 南京 210098; 2. 河海大学水文水资源及环境学院, 江苏 南京 210098;

3. 水利部水利水电规划设计总院, 北京 100011)

摘要: 在实编码遗传算法中加入单纯形搜索算子和加速搜索算子, 提出了混合加速遗传算法。通过实例对该法与其它一些遗传算法进行了比较。并在大坳流域模型的参数优选中得到成功的应用。结果表明, 混合加速遗传算法具有直观、简便、快速及适用性强等特点, 是一种既可以较大概率搜索全局最优解, 又能进行局部细致搜索的优秀非线性优化方法。

关 键 词: 实编码遗传算法; 单纯形法; 优秀个体; 流域模型

中图分类号: P 332.2 **文献标识码:** A **文章编号:** 1001-6791(2002)03-340-05

求解不可微非线性函数优化问题的主要方法有单纯形法、Rosenbrock 法及模式搜索法等。目前, 应用单纯形法、Rosenbrock 法进行参数优选需给定初值, 这就要求计算者具有相当丰富的经验, 才能得到“最优解”。通常只能得到局部极值解而非全局解, 通用性差。

遗传算法(Genetic Algorithm, 简称 GA)正是处理这类问题的通用方法, 该算法不限制模型结构、优化准则、参数数目、约束条件等, 自 J. H. Holland 于 1975 年首次提出后, 已引起人们广泛应用和研究^[1~3]。在应用 GA 时, 笔者发现二进制编码有时不便于反映所求问题的结构特征, 对于一些连续函数的优化问题等, GA 的局部搜索能力也较差。另外, 使用二进制编码, 需频繁的编码和解码, 计算量大。对那些具有众多参数、结构复杂的流域水文模型的参数优选, 为了提高流域水文模型参数的计算精度、计算速度和节约内存等, 本文采用实数(浮点数)编码。在实编码遗传算法中加入单纯形搜索算子及加速搜索算子, 逐步调整参数变化区间的方法, 形成了本文混合加速遗传算法(Hybrid Accelerating Genetic Algorithm, 简称 HAGA), 并在大坳流域模型的参数优选中得到了成功的应用。

1 HAGA 的计算技术

设非线性函数的优化问题为如下最小化问题:

$$\begin{aligned} \min & f(x) \\ \text{s. t. } & a(j) \leq x(j) \leq b(j), \quad (j = 1, 2, \dots, P) \end{aligned} \quad (1)$$

式中 $x = \{x(j)\}$ 为优化变量集; $[a(j), b(j)]$ 为 $x(j)$ 的变化区间; P 为优化变量数目; f 为目

收稿日期: 2001-06-14; 修订日期: 2001-08-30

基金项目: 教育部优秀骨干教师基金资助项目 (教技司 [2000] 65—14); 水利部科技创新基金项目 (SCX2000—53)

作者简介: 杨晓华(1966—), 女, 福建漳平人, 河海大学副教授, 博士研究生, 主要从事计算数学、非线性规划和工程系统优化研究。

标函数。HAGA 包括 8 步: (1) 变量取值范围的实数编码; (2) 初始父代个体的生成。随机生成 n 个个体作为初始群体 $P(0)$, 设置进化代数计数器 $t \leftarrow 0$, 设置最大进化代数 T ; (3) 个体评价。计算群体 $P(t)$ 中各个个体的适应度, 第 i 个个体的适应度 $F(i) = 1/(f(i) \times f(i) + 0.001)$, 目标函数值 $f(i)$ 越小, 表示该个体的适应度 $F(i)$ 越高; (4) 父代个体的选择运算。按与个体的适应度 $F(i)$ 成正比的概率 $p(i)$ 选择第 i 个个体。共选择 $n - 5$ 个个体。为增加 HAGA 进行持续全局优化搜索能力, 这里把最优秀的 5 个父代个体直接加进子代群体中, 进行移民操作后, 得到 n 个子代个体 $P_1(t)$; (5) 父代个体的杂交运算。按概率随机选择一对父代个体作为双亲, 并进行随机线性组合, 产生 n 个子代个体 $P_2(t)$; (6) 父代个体的变异, 在 HAGA 中, 任意父代个体 $P(t)$, 若其适应度数值 $F(i)$ 越小, 其选择概率越小, 则对该个体进行变异的概率应越大, 因此 HAGA 的变异操作是采用 $1 - p(i)$ 的概率对个体 $P(t)$ 进行变异, 从而得到子代个体 $P_3(t)$; (7) 演化迭代。由 (4) ~ (6) 得到的 $3n$ 个子代个体, 按其适应度值从大到小进行排序, 并引进单纯形搜索算子, 用上面所产生的最优秀的第一个个体作为初值, 利用单纯形法搜索得到局部最优解, 如果所得解满足精度要求, 则停止, 否则所得解替代上面所产生的第二个优秀个体, 取排在最前面的 n 个子代个体作为新的父代群体, 算法转入 (3); (8) 引进加速搜索算子^[1], 用上面第一、二次演化迭代所产生的前 s 个优秀个体, 这一子群体所对应的变量变化区间, 作为变量新的初始变化区间, HAGA 算法转入 (1), 如此加速循环, 优秀个体的变化区间将逐步调整和收缩, 与最优点的距离将越来越接近, 直到最优个体的目标函数值小于某一设定值或算法运行达到预定加速(循环)次数, 结束整个算法的运行, 并把当前群体中最优个体指定为 HAGA 的结果。

2 算法理论分析

(1) HAGA 的收敛性分析 定理: HAGA 依概率 1 收敛^[1]。证明方法与二进制编码类似详见文献 [1]。

(2) HAGA 的精确全局优化性能分析 HAGA 是先利用单纯形法搜索得到局部最优解, 再利用得到局部最优解作为优秀个体加入实编码遗传算法所产生的优秀个体中, 从而得到这一新优秀个体群来调整搜索范围的。HAGA 的全局优化不仅是精确的, 也是稳健的。在 p 个优化变量问题和 HAGA 加速循环 q 次的情况下, 优秀个体包围最优点的概率 P_{op} 为 $(1 - 0.5^{2s})^{pq}$ 。例如 s, p, q 分别取 10、10、10; 10、15、10; 10、20、10。 P_{op} 分别为 0.999 905; 0.999 857; 0.999 809。可见 HAGA 是依较大概率收敛于全局最优解的。

(3) HAGA 测试 本文提供一组数值模拟, 以展示所提 HAGA 法的有效性。将 HAGA 法与目前较好的三个遗传算法比较, 它们分别是 Tsutsui 等提出的 Forking 遗传算法^[2]、Michalewicz 提出的实编码非一致变异遗传算法^[3]及文献 [4] 的遗传算法。下面例 1、例 2 中的函数是数学规划问题中的经典函数, 一般算法很难找到它的全局最优解, 表 1 是 4 种算法在相同精度设置下所花费的 CPU 时间。本文用 HAGA 法求解例 1、例 2 时执行的参数设置如下: 最大加速次数为 6, 最大演化代数 $T = 2$, 最大种群规模 $N = 300$, 算法规模 $K = T * N = 600$, 杂交概率 $P_c = 0.8$, 优秀个体数目 $s = 10$ 。

$$\text{例 1 目标函数: } \max f(x) = 0.5 - \frac{(\sin(x_1^2 + x_2^2))^{1/2} - 0.5}{(1.0 + 0.001(x_1^2 + x_2^2))^2}; \quad -128 \leq x_i \leq 128$$

例2 目标函数: $\min f(x) = \sum_{i=1}^4 |x_i| + \prod_{i=1}^4 |x_i|$; $-5.01 \leq x_i \leq 5.01$

图1~2给出了测试问题 $\max f$ 用HAGA 法搜索过程中个体分布图。从表1及图1~2可见, HAGA 无论是收敛速度还是精度都较好。

表1 例1、例2 各方法比较

问 题		精 度	CPU 时间/s			
			非一致变异 GA	Forking GA	文献 [4] GA	本文算法
例1		10^{-8}	78.82	84.73	30.83	1.2
例2		10^{-6}	48.65	48.65	30.12	1.2

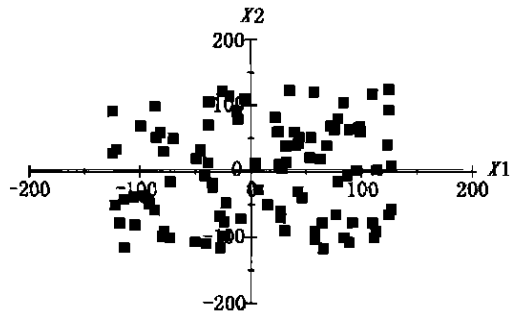


图1 初始群体分布

Fig.1 Distribution of the initial population

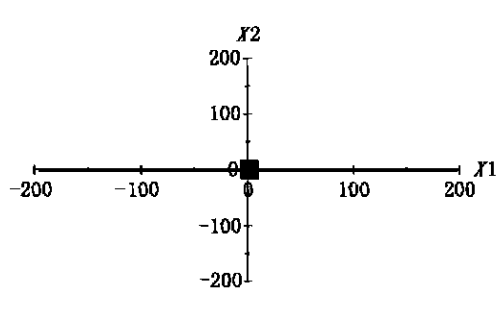


图2 第四代群体分布

Fig.2 Distribution of the fourth offspring

3 HAGA 在大垵流域新安江模型中的应用

上述的混合加速遗传算法可应用于任何概念性流域水文模型的参数优化或其他水文优化问题。本文仅以大垵流域新安江模型为例, 说明其应用情况。

关于大垵流域新安江模型的详细概念, 在此不再赘述。表2为参数的变化范围(值域)。其中, WM 为蓄水容量; X 为上层蓄水容量系数; Y 为下层蓄水容量系数; KE 为蒸散发系数; B 为蓄水容量曲线系数; SM 为自由水蓄水容量; EX 为自由水容量分布指数; CI 为壤中流出流系数; CG 为地下径流出流系数; IMP 为不透水面积比例; C 为深层蒸散发系数; KI 为壤中流消退系数; KG 为地下径流消退系数; N 、 NK 为纳希模型参数。

表2 参数值域

参数	WM	X	Y	KE	B	SM	EX	CI	CG	IMP	C	KI	KG	N	NK
最小值	30.0	0.05	0.1	0.5	0.1	0.0	0.05	0.05	0.01	0.01	0.05	0.5	0.95	1.0	1.0
最大值	220.0	0.5	0.9	2.0	2.0	50.0	2.0	1.0	1.0	0.5	0.3	0.9	1.0	10.0	20.0

(1) 目标函数 在模型参数的优化识别中, 目标函数的选取直接影响优化计算的结果和速度。本文选用的目标函数考虑了流量量级(门槛值 Q_r) 和不同目标函数权重系数 α , 即

$$f_1 = \sum_{i=1}^{n_1} [Q_0(i) - Q_c(i)]^2 \left[1 + \frac{\overline{Q}_0 - \overline{Q}_c}{\overline{Q}_0} \right] \quad Q_0 > Q_r \quad (2)$$

$$f_2 = \sum_{j=1}^{n_2} [Q_0(j) - Q_c(j)]^2 \left[1 + \frac{\bar{Q}_0 - \bar{Q}_c}{\bar{Q}_0} \right] \quad Q_0 \leq Q_r \quad (3)$$

总目标函数为

$$f = \min\{\alpha f_1 + (1 - \alpha)f_2\} \quad (4)$$

式中 Q_0 为观测流量; Q_c 为模拟流量; $\bar{Q}_0$ 为观测流量平均值; $\bar{Q}_c$ 为人为设定的很小的流量, 目的是较好地模拟洪峰。 n_1 为 $Q_0 > Q_r$ 的资料数; n_2 为 $Q_0 \leq Q_r$ 的资料数, $n = n_1 + n_2$ 为率定资料长度。从式(4)可看出, 对于不同的应用要求, 可取不同的 Q_r 和 α 值^[6]。

(2) 计算结果 本文用 HAGA 法来优选大坳流域新安江模型参数时, 参数设置同例 1。混沌法见文献 [5]。单纯形法、Rosenbrock 法见文献 [6], 当参数初始值分别取区间的中点和下面的值: $M = 50$, $X = 0.1$, $Y = 0.5$, $KE = 1.0$, $B = 1.0$, $SM = 25.0$, $EX = 1.0$, $CI = 0.5$, $CG = 0.5$, $IMP = 0.25$, $C = 0.2$, $KI = 0.6$, $KG = 0.96$, $N = 5.0$, $NK = 10.0$ 时, 分别记为 S 法(1)、R 法(1)和 S 法(2)、R 法(2)。

表 3 给出了 $Q_r = 0$, $\alpha = 1$ 时大坳流域新安江模型率定期(1989–1996 年)、检验期(1997–1998 年)的平均观测流量与平均计算流量比值 Q_0/Q_c 及模型效率系数用 HAGA 法、R 法(1)、R 法(2)、S 法(1)、S 法(2) 及 Chaos 法的计算结果。

表 3 大坳流域新安江模型各方法比较

方 法	率定期		检验期	
	比值 Q_0/Q_c	模型效率	比值 Q_0/Q_c	模型效率
HAGA 法	1.000	91.156	1.044	92.346
R 法(1)	0.995	89.267	1.017	91.172
R 法(2)	0.999	90.543	1.030	90.953
S 法(1)	1.000	85.702	1.030	92.715
S 法(2)	1.004	80.617	1.050	88.028
Chaos 法	1.139	87.087	1.146	91.671

图 3 给出了大坳流域 1998 年 6 月 1~30 日用 HAGA 法计算的流量值与实测值, 两者较吻合。

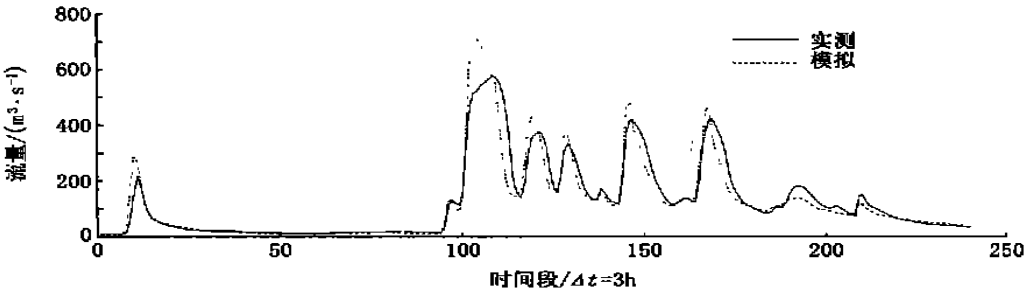


图 3 大坳流域 1998 年 6 月 HAGA 法流量实测值与模拟值比较

Fig. 3 Comparison of predicted by using of HAGA and measured of runoff in Daao basin

4 结 语

在大坳流域新安江模型中采用 HAGA 法优化求解, 结果表明, HAGA 法可用来求解该优化问题, 在实现过程中我们发现怎样按某种过程产生新的搜索区间是个值得探索的问题。产生新的

搜索区间越好,则搜索的解越准确,搜索的速度越快。为保证在全局中较快地寻求最优解,本文产生的新的搜索区间采用实编码遗传算法及单纯形法共同搜索的多个优秀个体所生成的区间。应用实例表明,HAGA 法收敛性及收敛速度都令人满意。表 4 中 R 法(2)、s 法(2)分别是对 R 法(1)、s 法(1)的参数初始值进行相同扰动而形成的方法,但结果却大相径庭。由此可见, simplex 法、Rosenbrock 法明显依赖于初值,混沌方法对初值也具有一定的依赖性,而本文方法却具有不依赖于初值的优点。寻优结果用本文方法率定期较好,检验期好于或近似于传统方法。由于本文方法的参数初始值域是经验设定的,因此该结果也是合理的。对于多参数的概念性流域水文模型的参数优选本文方法具有一定的参考价值。它对于在参数率定时所主要关心的问题,如优选出的参数符合物理概念、在大区域范围内使目标函数达到全局最优等,都能圆满地解决。作为一种新的并行优化方法,其将在水科学研究中具有广阔的应用前景。

参考文献:

- [1] 金菊良,丁 晶. 遗传算法及其在水科学中的应用[M]. 成都:四川大学出版社,2000.
- [2] Tsutsui S, Fujimoto Y, Ghosh A. Forking genetic algorithms: GAs with search space division schemes[J]. *Evolutionary computation*, 1997, 5(1): 61–80.
- [3] Michalewicz. Genetic algorithm+ data structure= evolutionary programs[M]. 3rd ed. New York: Springer Verlag, 1996.
- [4] 孙建永, 申建中, 徐宗本. 一类自适应遗传算法的理论分析与数值模拟[J]. *西安交通大学学报*, 34(12): 84–87.
- [5] 李 兵, 蒋慰孙. 混沌优化方法及其应用[J]. *控制理论与应用*, 1997, 14(4): 613–615.
- [6] 水利部水利信息中心三队. 模型参数优化方法及其应用[J]. *水文*, 1999(增刊): 61–65.

Applications of the hybrid accelerating generic algorithm to parameter optimization of basin model

YANG Xiao-hua¹, LU Guo-hua², LI Jiar-qiang³

(1. Dept. of Mathematics and Physics, Hohai Univ., Nanjing 210098, China;

2. College of Water Resources and Environment, Hohai Univ., Nanjing 210098, China;

3. Water Resources and Hydropower planning and Design General Institute, MWR, Beijing 100011, China)

Abstract: A hybrid accelerating genetic algorithm (HAGA) is presented by adding simplex searching operator and accelerate searching operator in real coding genetic algorithm. The HAGA is compared with some other genetic algorithms through some examples, and applied to the parameter optimization of the Daao basin model. The results show that this method has the features of direct, convenient, fast and good application. It is a superior non-linear optimal method which could locally search the global solution for greater probability

Key words: real coding genetic algorithm; simplex algorithm; superior individual; basin model