

文章编号: 1009-6248(2005) 03-0100-05

基于十进制遗传算法的含水层参数的反演

张娟娟¹, 万伟锋¹, 韩淑敏²

(1. 长安大学, 陕西 西安 710054; 2. 中国科学院石家庄农业现代化研究所, 河北 石家庄 050021)

摘要: 含水层参数的反演是一个复杂的非线性优化问题, 针对传统二进制遗传算法收敛性能差的缺陷, 提出了反演含水层参数的十进制遗传算法。以直线隔水边界附近的井流模型为例, 讨论了十进制遗传算法在含水层参数反演中的应用, 并与二进制遗传算法的进行比较。结果表明, 该方法在含水层参数的反演中不仅是可行的, 而且具有较好的确定性和较高的精度; 与二进制遗传算法相比, 十进制遗传算法的收敛性较好, 省时高效, 且表示较为自然, 容易引入相关领域知识。同时, 结合实例的分析结果得出种群的规模对算法的收敛性没有明显的影响。

关键词: 十进制遗传算法; 二进制遗传算法; 反演; 含水层参数; 参数识别

中图分类号: P641.2 **文献标识码:** A

1 引言

含水层参数的可靠性估计对地下水资源的评价和管理有至关重要的意义。目前, 主要是利用野外抽水试验来确定含水层参数, 而非稳定流抽水试验应用更为广泛。用非稳定流抽水试验资料确定含水层参数的方法很多, 主要有配线法、直线图解法、单纯形法、非线性优化技术(如最速下降法、共轭梯度法、变尺度法和叠代算法)^[1,2]等。配线法和直线图解法是目前最常用的两种方法, 但在求参过程中需要查表和点绘关系图, 人为的主观任意性较大, 计算结果因人而异, 且不易程序化, 往往产生较大的误差; 单纯形法易收敛于局部最优解; 最速下降法、共轭梯度法和变尺度法, 都要利用目标函数的导数值等其他一些辅助信息才能确定搜索方向, 因此限制了它们的使用; 叠代算法虽然可程序化, 运算速度较快, 但其收敛性对初值的依赖性较强, 初始值确定的不合理可能导致计算结果不唯一, 所以, 必须合理地选取初值, 以保证算法收敛和解的唯一性。

因为这些方法在实际应用中都有其局限性。所以, 本文将遗传算法应用于抽水试验资料的分析中, 来实现含水层参数的识别。

遗传算法是近年来发展的基于自然选择规律的一种优化方法, 算法能成功的解决无功变量中的离散问题, 避免常规数学优化方法的局部最优现象, 其主要特点是直接对结构进行操作, 不存在求导和函数连续性的限制, 克服了上述方法的不足。传统的遗传算法一般采用二进制编码, 但在求解最优化问题时, 二进制编码存在着优化精度差的缺陷^[3,4], 并降低了算法的收敛效率。为了克服这些缺点, 对于函数优化问题, 可以直接采用实数编码, 采用实数编码不仅可以克服二进制编码算法优化精度差、收敛效率低的缺陷, 而且便于引入与问题领域相关的启发信息以增加遗传算法的搜索能力。

因此, 针对二进制遗传算法优化精度低、收敛性能弱的缺陷, 笔者采用了十进制编码方法, 并以分析抽水试验资料估计含水层参数为例, 讨论了十进制遗传算法在含水层参数识别中的应用方法和步

收稿日期: 2005-04-04; 修回日期: 2005-06-17

基金项目: 中国科学院知识创新工程重要方向性项目(KZCX3-SW-428)

作者简介: 张娟娟(1981-), 女, 陕西武功人, 硕士研究生, 主要研究智能优化算法在水科学中的应用。

骤, 结合时间函数给出了十进制遗传算法与二进制遗传算法的计算时间, 并进行了比较。结果表明, 该方法不仅求解精度高, 而且计算速度较二进制遗传算法的计算速度快。

2 遗传算法的思想和步骤

遗传算法^[5,6] (genetic algorithms, 简称 GA) 是 J. Holland 于 1975 年受生物进化论的启发而提出的。GA 是基于“适者生存”的一种高度并行、随机和自适应的优化算法, 它将问题的求解表示成“染色体”的适者生存过程, 通过“染色体”群的一代代不断进化, 包括复制、交叉和变异操作, 最终收敛到“最适应环境”的个体, 从而求得问题的最优解或满意解。

遗传算法是一类随机优化算法, 但它不是简单的随机比较搜索, 而是通过对染色体的评价和对染色体中基因的作用, 有效地利用已有信息来指导搜索有希望改善优化质量的状态, 标准遗传算法的主要步骤可描述如下 (图 1):

- (1) 随机产生一组初始个体构成初始种群, 并评价每一个体的适应值 (fitness value)。
- (2) 判断算法收敛准则是否满足。若满足则输出搜索结果; 否则执行以下步骤。
- (3) 根据适应值大小以一定方式执行复制操作。
- (4) 按交叉概率 P_c 执行交叉操作。
- (5) 按变异概率 P_m 执行变异操作。
- (6) 返回步骤 (2)。

上述算法中, 适应值是对染色体 (个体) 进行评价的一种指标, 是 GA 进行优化所用的主要信息, 它与个体的目标函数值存在一种对应关系; 复制操作通常采用比例复制, 即复制概率正比于个体的适应值, 如此意味着适应值高的个体在下一代中复制自身的概率大, 从而提高了种群的平均适应值; 交叉操作通过交换两父代个体的部分信息构成后代个体, 使得后代继承父代的有效模式, 从而有助于产生优良个体; 变异操作通过随机改变个体中某些基因而产生新个体, 有助于增加种群的多样性, 避免早熟收敛。

3 适应度函数的建立

在分析水平无限大承压含水层中的抽水试验资料, 含水层中任一点的井流问题解析解可描述为:

$$s = \frac{1}{4\pi T} \sum_{i=1}^n Q_i W(u_i) \quad (i=1, 2, \dots, n) \quad (1)$$

$$u_i = \frac{r_i^2 S}{4Tt} \quad (i=1, 2, \dots, n) \quad (2)$$

$$\text{式中 } W(u_i) = \int_0^{\frac{1}{u_i}} \frac{e^{-y}}{y} dy = -0.5772 - \ln u_i + u_i - \frac{u_i^2}{2 \cdot 2!} + \frac{u_i^3}{3 \cdot 3!} - \frac{u_i^4}{4 \cdot 4!} + \dots \quad (i=1, \dots, n)$$

在进行具体计算时, 可采用以下近似式^[8]:

$$\text{当 } u_i < 1 \text{ 时, } W(u_i) = -\ln u_i + a_0 + a_1 u_i + a_2 u_i^2 + a_3 u_i^3 + a_4 u_i^4 + a_5 u_i^5 \quad (i=1, \dots, n)$$

$$\text{当 } u_i \geq 1 \text{ 时, } W(u_i) = \frac{1}{u_i e^{u_i}} \frac{b_0 + b_1 u_i + b_2 u_i^2 + b_3 u_i^3 + u_i^4}{u_i e^{u_i} c_0 + c_1 u_i + c_2 u_i^2 + c_3 u_i^3 + u_i^4} \quad (i=1, \dots, n)$$

其中: $a_0 = -0.57722$ 、 $a_1 = 0.99999$ 、 $a_2 = -0.24991$ 、 $a_3 = 0.05519$ 、 $a_4 = -0.00976$ 、 $a_5 = 0.00108$ 、 $b_0 = 0.26777$ 、 $b_1 = 8.63476$ 、 $b_2 = 18.05902$ 、 $b_3 = 8.57333$ 、 $c_0 = 3.95850$ 、 $c_1 = 21.09965$ 、 $c_2 = 25.63296$ 、 $c_3 = 9.57332$ 。

上述式中 S ——含水层中任一点的水位降深, [L]; Q_i ——第 i 口抽水井的定流量抽水量, [$L^3 T^{-1}$]; T ——含水层导水系数, [$L^2 T^{-1}$]; $W(u_i)$ ——泰斯井函数; r_i ——观测孔距第 i 口抽水井的距离, [L]; S ——含水层储水系数, 无量纲; t ——抽水时间, [T]; u_i ——无因次。

在应用十进制遗传算法分析时间-降深抽水试验资料, 反演含水层参数时, 要求欲估计的参数值使下式达到最小。

$$\Phi(\theta) = \frac{1}{n} \sum_{k=1}^n (S_k^o - S_k^c)^2 \Rightarrow \min \quad (3)$$

式中 S_k^o —— k 时刻观测到的实际降深值, [L]; S_k^c —— k 时刻利用 (1) 式计算的降深值, [L]; θ ——待估计参数向量。其中 $a < \theta < b$, a 、 b 分别为参数 θ 的上下限。式 (3) 的意义为选取适当的参数值, 使得降深计算值与观测值的离差平方和的均值达到极小。此时对应的参数值, 即为问题所求。对于问题 (3), 设含水层的渗透系数 T 为 θ_1 , 储水系数 S 为 θ_2 , 观测井到映射井的距离 r_2 为 θ_3 。

因此, 可设适应度函数为

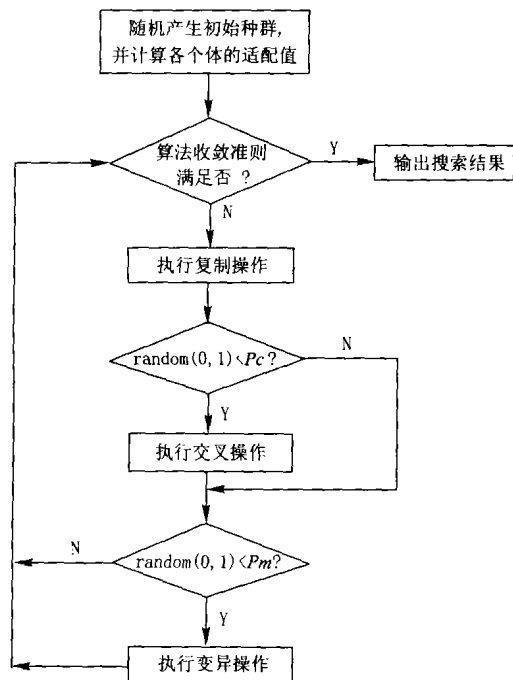


图1 标准 GA 的程序流程图

Fig. 1 The flow chart of standard Genetic Algorithm

$$fit(\theta_1, \theta_2, \theta_3) = \frac{1}{\varphi(\theta)} = \frac{1}{n} \sum_{k=1}^n (S_k^o - S_k^c)^2 \quad (4)$$

4 算例

4.1 原始数据与参数设置

本文算例采用文献[9]中的资料, 单井抽水试验在具有直线隔水边界的承压含水层中进行, 已知主井作定流量抽水, 抽水时间长达960 min。抽水量 $Q = 4.581 \text{ m}^3/\text{min}$, 观测孔距抽水井的距离 $r_1 = 30.48 \text{ m}$ 。应用遗传算法估计含水层参数时, 收敛条件设置为 $\epsilon < 0.00001$; 状态参数约束为 $0.3 < T < 300$, $0.0001 < S < 1$, $10 < r_2 < 1000$; 遗传操作的关键参数分别设置为: 种群大小 $M = 100$ 、交叉率 $P_c = 0.5$ 、变异率 $P_m = 0.07$ 。

在利用十进制遗传算法反演参数时, 只采用抽水试验过程中的部分观测资料(如表1所示), 而把其他时刻降深的观测值作为检验数, 不参与运算。利用作者依据图1编制的C语言程序, 代入定流量抽水量, 观测孔距抽水井的距离和文献中的部分观测时间 $t(\text{min})$ 和降深 $s(\text{m})$ 的数据, 在计算机上运行此程序, 即可在很短的时间内求出参数值。

表1 t_i-s_i 实测数据Tab. 1 True drawdown data of t_i-s_i

$t(\text{min})$	5	10	20	30	50	100	120	180
$s(\text{m})$	0.024	0.067	0.125	0.171	0.235	0.347	0.381	0.460

4.2 结果分析

4.2.1 算法可行性的检验

根据上述资料, 采用本文的十进制遗传算法计算出的导水系数、储水系数、观测孔至映射井的距离分别为 $T = 3.0119 \text{ m}^2/\text{min}$ 、 $S = 0.0665$ 、 $r_2 = 105.3037 \text{ m}$ 。

为了检验十进制遗传算法求参的可靠性, 文中根据十进制遗传算法所求得的参数值, 反求整个抽水过程中各时刻的降深值, 并与降深实际观测值作一比较(图2)。从图2可以看出, 实际观测值与计算值两条曲线基本上完全重合, 这表明从所求得的参数反推的各时刻的降深值与实际观测的降深值基本上是一致的, 也就是说, 采用文中的方法反演出的水文地质参数符合实际含水层的水力特性。由此可以知道, 十进制遗传算法在分析抽水试验资料, 反演水文地质参数和识别隔水边界的函数优化问题中是可行的、可靠的, 从而为计算水文地质参数提供了一条可靠的途径。

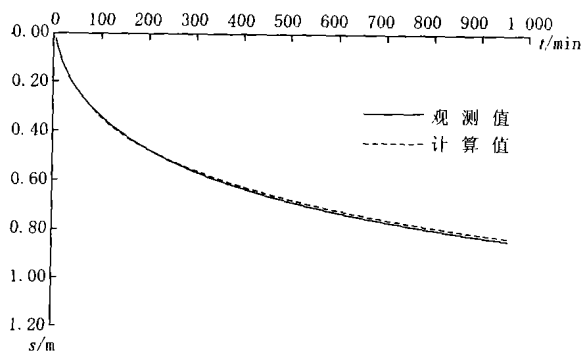


图2 降深观测值与计算值比较曲线图

Fig. 2 The comparison curve graph of true and calculated drawdown data

4.2.2 十进制遗传算法与二进制遗传算法的比较

在应用遗传算法进行参数计算时发现, 采用传统的二进制编码时, 不论种群数为多少, 反演的参数所模拟出的降深计算值与其实观测值相差较大。在采用十进制编码时, 模拟出的降深的计算值与观测值的相对误差均小于2%。

从表 1 可以看出，不论是进化代数，还是计算时间，十进制都明显的小于二进制。从进化代数来看，无论种群数为多少，十进制算法的进化代数都小于二进制的进化代数，这无疑减少了计算的叠代次数，改善了算法的收敛性。从计算时间来看，十进制在本例中所花的时间仅仅用了 1 s 左右，较二进制算法的时间小得多，这一点在一些高维的复杂的函数优化问题中很值得关注。由此可以看出，采用十进制编码方法，可以明显地改善算法的收敛性。

并且能得到满意的水文地质参数结果。

另外，从表 1 还可以看出，当种群数为 10，即种群规模过小时，算法在较少进化代数条件下达不到收敛的条件。但在一定的收敛情况下，随着种群规模的变化，算法的进化代数及计算时间也没有呈明显的增加或减小的趋势。因此，在算法设计时，在保证种群规模不至于过小，在算法收敛的条件下，种群规模的大小对算法的收敛性没有明显的影响。

表 2 十进制编码与二进制编码算法的比较

Tab.2 The comparison of decimal strings and binary strings genetic algorithm

种群规模		10	20	30	50	100	300	500	700	1000
进化代数	二进制	—	505	250	334	441	28	56	26	136
	十进制	—	119	57	49	44	9	8	6	8
时间 (s)	二进制	—	4. 01	3. 49	7. 49	16. 26	3. 47	11. 31	7. 87	54. 77
	十进制	—	0. 47	0. 34	0. 40	0. 66	0. 48	0. 74	0. 83	1. 6

5 结束语

文中将十进制遗传算法应用于直线隔水边界附近承压含水层参数反演的函数优化问题, 结果表明:

- (1) 十进制遗传算法在含水层参数反演中是完全可行的, 且具有求解精度高、收敛于全局最优解、无需求导和初值鲁棒性的优点。
- (2) 与二进制遗传算法相比，十进制编码的计算和编程均较为简单，同时能较好地解决收敛问题和效率问题。
- (3) 在保证收敛的条件下，种群的规模对算法的收敛性能没有明显的影响。

参考文献:

[1] 齐学斌. 非稳定流抽水试验参数计算的叠代算法及计算机模型 [J] . 水利学报, 1995. 67-71, 77.
[2] Manoj P. Samuel and Madan k. Jha, M. ASCE.

Estimation of Aquifer Parameters from Pumping Test Data by Genetic Algorithm Optimization Technique [J] . Irrigation and Drainage Engineering, 2003, 129 (5): 348-349.
[3] 刑文训. 现代优化计算方法 [M] . 北京: 清华大学出版社, 1999.
[4] 潘正君. 演化计算 [M] . 北京: 清华大学出版社, 1998. 64-65.
[5] 王凌. 智能优化算法及其应用 [M] . 北京: 清华大学出版社, 施普林格出版社, 2001. 36-61.
[6] 张颖. 软计算方法 [M] . 北京: 科学出版社, 2002. 69-108.
[7] 郭东屏. 地下水动力学 [M] . 西安: 陕西科学技术出版社, 1994. 170-175.
[8] Rajesh Srivastava. Implication of Using Approximate Expression For Well Function[J] . Journal of Irrigation and Drainage Engineering. 2003, 121 (6): 459-462.
[9] Sushil K Singh. Identifying Impervious Boundary and Aquifer Parameters from Pump-Test Data [J] . Journal of Hydraulic Engineering. 2001, 127 (4): 280-285.

Identifying aquifer parameters based on the decimal strings genetic algorithm

ZHANG Juan-juan¹, WAN Wei-feng¹, HAN Shu-min

(1. *Chang'an University, Xi'an 710054, China*; 2. *Institute of Agricultural Resources Research, Chinese Academy of Science, Shijiazhuang 050021, Hebei, China*)

Abstract: The problem of aquifer parameter identification is actually a complex nonlinear program. With the limit of the ability of convergence of the binary strings genetic algorithm, the decimal strings genetic algorithm is proposed for identifying the aquifer parameters. An application of decimal strings genetic algorithm for aquifer parameter identification is discussed as an example of the well model near the linear impervious boundary, and compared with the binary strings algorithm. Application of the method to the published data shows the proposed method is not only applicable, but has higher accuracy and better determinacy. It is also shown from this paper that compared to the binary strings genetic algorithm, the decimal strings genetic algorithm can be natural to represent the problem to be studied, easy to introduce relative area knowledge, and has high efficiency, especially for multi-dimensional optimization problem.

Key words: decimal strings genetic algorithm; binary strings genetic algorithm; inverse problem; aquifer parameter; parameter identification