

文章编号:1673-5005(2008)03-0045-05

叠前弹性波反演非线性优化方法

殷文^{1,2}

(1. 东营职业学院, 山东 东营 257091; 2. 中石化胜利油田物探研究院, 山东 东营 257002)

摘要:针对传统遗传算法自身存在的早熟收敛、搜索空间小以及计算效率低的问题,在保证算法收敛和最大限度地搜索模型空间的基础上,对遗传算子采取相应策略进行了改进,并通过界约束以增加解的稳定性。为了提高计算效率,采用粗粒度并行遗传算法,将并行计算机的高速并行性和遗传算法固有的并行性相结合,进行多种群并行搜索。选择合适的迁移拓扑结构和迁移策略,构建了并行模型,并给出了改进后并行遗传算法的设计流程图及详细算法描述。采用该算法进行了叠前弹性波反演的实际计算,取得了良好的效果。

关键词:非线性多参数优化;叠前弹性波反演;并行遗传算法

中图分类号:P 631.443 **文献标识码:**A

Nonlinear optimization method on prestack elastic-wave inversion

YIN Wen^{1,2}

(1. Dongying Vocation College, Dongying 257091, Shandong Province, China;

2. Geophysical Research Institute of Shengli Oilfield, SINOPEC, Dongying 257002, Shandong Province, China)

Abstract: Due to the problems of premature convergence, searching space and computational efficiency in routine genetic algorithm, some relevant improved strategies were adopted for genetic operators in genetic algorithm ensuring convergence and the effective search of the model space. Additionally, search boundaries were set up to stabilize the solutions. In order to enhance computational efficiency, coarse-grained parallel genetic algorithm (PGA) which combines high-speed concurrency of parallel computer with inherent one of genetic algorithm was adopted to perform parallel search, select appropriate migrating topological architecture and migrating strategy, and establish parallel model. The designed flow chart and the detailed algorithm description of modified PGA were given. The calculated results show that the proposed algorithm is effective.

Key words: nonlinear multiparameter optimization; prestack elastic-wave inversion; parallel genetic algorithm

随着油气勘探开发工作不断深入,在构造油气藏研究发展到隐蔽油气藏研究的过程中,从地震数据中确定性参数显得更加重要。波形反演可利用同相轴的振幅及旅行时信息提取物性参数,并且已成功地应用于地震资料解释、储层描述和四维地震处理等方面。由于使用全角度多次叠加后的地震资料,叠后反演在某种程度上削弱了反映储层特征的敏感性,而叠前弹性波反演技术,克服了叠后反演的不足,不但适合薄储集层物性反演,还可进行含油气性反演。叠前弹性波反演属于多参数、非线性优化问题,采用经典的优化方法难以解决,而遗传算法是一种自适应启发

式及概率性迭代式全局搜索算法,在解决这类问题方面有着独到的优势。但遗传算法自身还存在一些不足,其计算效率也是制约叠前反演技术发展的重要因素。鉴于此,笔者采用多种遗传算子技术及并行策略对传统算法进行改进,并结合叠前弹性波反演实例对改进后算法的有效性进行检验。

1 叠前弹性波反演方法的改进策略

1.1 编码策略

传统的遗传算法采用二进制编码把搜索空间的解映射到遗传空间,对于高维连续变量的优化问题,

收稿日期:2007-10-18

基金项目:中石化重点科研攻关项目(P05027)

作者简介:殷文(1978-),男(汉族),甘肃武威人,讲师,博士,在胜利油田博士后流动站从事地震资料处理研究。

在遗传空间占用二进制位数很多,搜索速度特别慢,而且二进制编码存在着连续函数离散化时的映射误差,同时不便于反映所求问题的特定知识。

目前,也有学者采用十进制浮点数编码的方法^[1],这样做的优点是避免了编码解码的繁杂运算。本文中采用五进制编码^[2]。在编码位数相同的情况下,五进制搜索空间比二进制的搜索空间大得多;十进制编码的每一次遗传操作进行交叉和变异后,数值的改变很可能会突然增大或减小,跳跃性大,局部收敛比较缓慢,五进制的收敛速度比十进制快;五进制能够达到的精度为 1×10^{-5} ,而十进制和二进制只能达到 1×10^{-4} 。可以证明,当编码的进制为2~10的整数时,四到六进制的编码多数情况下能提高遗传算法的计算精度,同时减少循环次数^[3]。

1.2 逐步缩小搜索范围

遗传算法能够满足给定精度的全局极值,但其代价是大量的时间消耗,从参数搜索精度的角度,可采用逐步缩小参数搜索范围^[4]的方法对其进行改进。遗传算法对模型空间的搜索速度及搜索精度与模型空间的初始划分精度有关,而模型空间的初始划分精度是参数空间中参数的取值间隔,它与各参数的初始搜索范围(x_{\min}^i, x_{\max}^i)及用户所要求的参数取值精度 d^i ($d^i = (x_{\max}^i - x_{\min}^i) / (2^n - 1)$,其中 2^n 为参数范围的划分间隔数)有关,这样,各模型参数的可能值可以表示为

$$x^i = x_{\min}^i + jd^i, j=0, 1, \dots, N^i; i=1, 2, \dots, m.$$

式中, N^i 为各模型参数可能取值的个数,对每个模型参数是不同的; m 为模型参数个数。随着演化进程的发展,模型群体在其搜索范围内逐步向其真值附近集中,最后,绝大多数模型分布在真值附近的一个局部范围内。因此,可以在模型群体的演化过程中逐步缩小搜索范围,这样做有如下优点:①可以逐步圈定出真值所处的大致范围;②在保持搜索范围原有分段数不变的情况下,相应地提高了搜索精度,而提高搜索精度并不影响搜索范围变化后的搜索速度;③由于缩小搜索范围后不必再去搜索远离真值的区域,可相应地提高搜索效率。

石琳珂等^[4]已证明缩小范围后,真值仍然保留再缩小后的搜索范围内,即当前最优值 X 进入真值的领域,可实施搜索范围的压缩,即

$$\begin{cases} X'_{\min} = X_{\min} + (X - X_{\min})/D, \\ X'_{\max} = X_{\max} - (X_{\max} - X)/D. \end{cases} \quad (1)$$

其中, $D \geq 2$,按式(1)得到的新的搜索范围(X'_{\min} ,

X'_{\max})一定包含真值和当前最优值在内。安全起见,可取 $D=3$ 或4。

1.3 基因交叉和变异

基因交叉一般采用一点交叉的方式,但当模型参数较多时,一点交叉方式的搜索范围是很有限的,会降低遗传算法的搜索能力,势必会对收敛速度产生负面影响。同理,单点变异也是如此。

为了更好地进行基因重组以产生新的模式,可采用多点交叉和多点变异算子。其中多点交叉基本思想是,给每个模型参数赋予均等的获得改变的机会,以模型参数为单位,由随机数决定是否对该模型参数做基因交叉,如果要做交叉,则在该模型参数中选一基因交叉点,进行交换。多点变异的思想与此相同。需要注意的是,多点交叉和多点变异的目的是扩大模型空间的搜索范围和保持个体的多样性。但如果选择的交叉和变异点过多,将会破坏已有的一些好的模式,因此应根据模型参数的规模适当选择合适的交叉和变异点数。

1.4 适应度函数的设计

1.4.1 从目标函数到适应度函数

在许多问题中,求解目标更自然地表示成某个代价函数 $f(x)$ 的极小化,而不是某个利益函数 $g(x)$ 的极大化^[5]。即使问题被表示成极大化形式,但不能确保利益函数 $g(x)$ 对所有的 x 都是非负的。因此,常常需要通过一次或多次变换把目标函数转换到适应度函数 $F(x)$ 。

当目标函数是利益函数时,可以直接得到适应度函数。如果出现了负利益函数 $g(x)$ 的情形,可以利用下面的变换来克服:

$$F(x) = \begin{cases} g(x) + C_{\min}, & g(x) + C_{\min} > 0; \\ 0, & g(x) + C_{\min} \leq 0. \end{cases}$$

其中,参数 C_{\min} 可以取为输入参数,取当前代和最近几代中 $g(x)$ 的最小值的绝对值。

1.4.2 适应度函数的尺度变换

在遗传算法中,某个体被选中进行遗传操作的概率是由该个体的适应度决定的,适应度函数的构造对收敛结果产生非常大的影响。在算法初期阶段降低适应度较高的个体与其他个体适应度之间的差异,从而限制其复制数量,以维护群体的多样性;在最后阶段需扩大最佳个体的适应度与其他个体适应度之间的差异程度,以提高个体之间的竞争力。在遗传算法运行的不同阶段应对个体的适应度进行适当的扩大和缩小,即进行适应度尺度变换。可采用线性尺度变换方法,即

$$F' = F'_{\min} + \frac{F'_{\max} - F'_{\min}}{F_{\max} - F_{\min}}(F - F_{\min}).$$

其中

$$F'_{\max} = cF_{\text{ave}},$$

$$F'_{\min} = F_{\text{ave}} - \frac{F'_{\max} - F'_{\min}}{F_{\max} - F_{\min}}(F_{\text{ave}} - F_{\min}),$$

$$F_{\text{ave}} = \frac{1}{M} \sum_{i=1}^M F_i.$$

式中, F 和 F' 分别为旧的和新的适应度函数; c 为 1.2 到 2 之间的常数; M 为群体规模。

1.5 约束条件

由于叠前反演问题具有多解性, 如果没有合适的约束条件进行引导搜索, 就很容易陷入局部解估

计, 并且有可能存在明显的不稳定振荡, 而恰当的约束会降低问题的非惟一性, 并缩短计算时间。为此, 在算法中可加入以下约束: ①先验信息。通过一定渠道获得一些初始的信息, 将这些约束条件放到目标函数中, 以引导计算结果向有物理意义的趋势发展。②考虑模型和测井数据的约束^[6], 对每一参数设定搜索范围。

以 Ostrander 模型^[7]为例, 采用改进后的遗传算法对其进行叠前反演计算, 反演结果见图 1(图中实线代表反演的地层模型, 虚线代表真实的地层参数), 两种方法反演所用时间如表 1 所示。

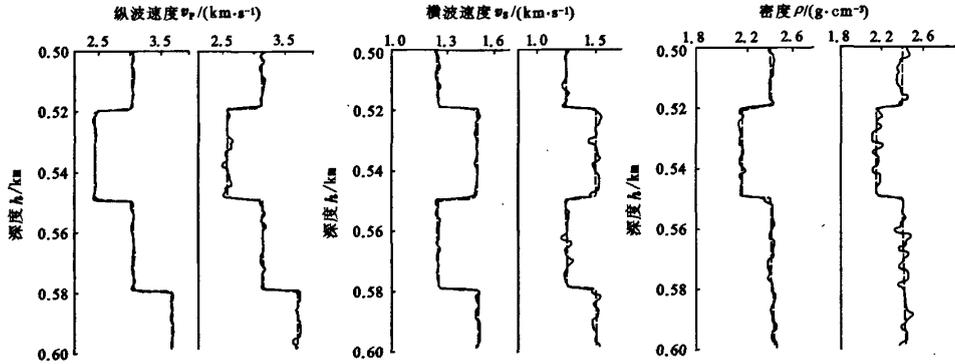


图 1 传统的(右图)和改进后的(左图)遗传算法叠前弹性波反演结果对比

表 1 采用传统的和改进后的遗传算法进行反演所用的时间

群体规模 M	遗传代数	时间 /s	
		传统的算法	改进后的算法
10	20	150.536	138.791
20	30	448.1508	420.317
30	50	1149.812	1116.585
40	80	2515.252	2214.664
50	100	3928.408	3228.800

由图 1 不难看出, 采用改进后的遗传算法的反演结果是比较理想的, 且改进后的遗传算法的精度更高。从表 1 中还可以看出, 改进后的算法比传统算法所用时间要短, 而且群体规模越大, 优势越明显。在相同精度条件下, 改进算法迭代次数更少, 收敛速度更快。

2 遗传算法的并行化

串行遗传算法在解决叠前反演问题时, 由于它一般具有较大的群体规模, 需要对较多的个体进行大量的遗传和进化操作, 特别是在反演过程中要多

次进行正演计算, 从而使得算法的进化运算过程进展缓慢, 难以达到计算速度的要求, 因而在反演过程中引入了并行处理技术。并行遗传算法(PGA)将并行计算机的高速并行性和遗传算法固有的并行性相结合, 不但加速了遗传算法的搜索过程, 而且由于种群规模的扩大和各子种群的隔离, 使得种群的多样性得以丰富和保持, 减少了未成熟收敛的可能性, 提高了求解质量, 极大地促进了遗传算法的研究和应用。

目前, 粗粒度模型最为流行, 是适应性最强和应用最广的遗传算法并行化模型。粗粒度模型是将随机生成的初始群体依处理器个数分割成若干个子群体, 各个子群体在不同的处理器上相互独立地并发执行进化操作, 每经过一定的进化代, 子群体间会交换若干个个体以引入其他子群体的优秀基因, 丰富各子群体的多样性, 防止未成熟收敛的发生。另外, 粗粒度模型的通信开销较小, 非常适合运行在通信带宽较低的集群系统上。在迁移策略方面, 考虑到染色体的多样性, 每个处理器都将自己最好的个体仅传给与之相邻的一个处理器, 同时增加两个参数:

一是 send_rate 决定处理器之间通信的频率,如 send_rate = 3 时表示当遗传代数是 3 的倍数时,各处理器之间相互传个体;二是 send_best 决定每次传送给最好个体的数目。这里给出有关改进后的并行遗传算法设计流程,如图 2 所示。

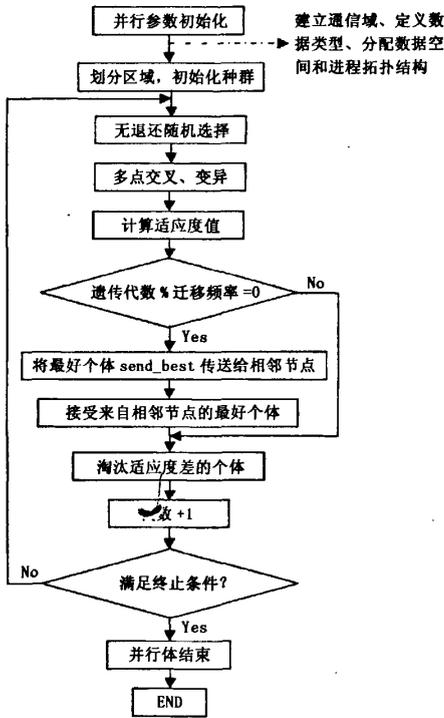


图 2 并行遗传算法流程图

3 应用实例

采用改进后的并行遗传算法来对实际资料进行叠前弹性波反演,利用正演合成记录来进行判别,即

$$u(t, r) = \frac{1}{\pi \sqrt{2r}} \frac{\partial^2 s(t)}{\partial t^2} * \lambda(t) * \int_0^\infty \sqrt{p} \bar{r}_{pp}(t - t_0,$$

$p) dp,$
其中

$$\lambda(t) = H(t) / \sqrt{t}.$$

式中, k 为水平波数; $S(\omega)$ 为源函数; p 为水平慢度; r 为偏移距; $H(t)$ 为阶跃函数。

按照上述并行遗传算法分析与设计方案,对目的层位于 950 ~ 1 200 ms, 4 ms 采样, 50 次覆盖, 群体规模为 30 的实例, 在集群系统上, 采用标准的消息传递模型 MPI^[8-10] 编程进行叠前弹性波反演。反演结果如图 3 所示, 其中实线是反演的曲线, 虚线是实际的测井曲线。图 4 为实际叠前记录及合成记

录, 其中的主要层位匹配得很好。可见反演结果与实际记录具有很好的相似性。因此, 改进的并行遗传算法具有对叠前数据进行全弹性波反演的能力。

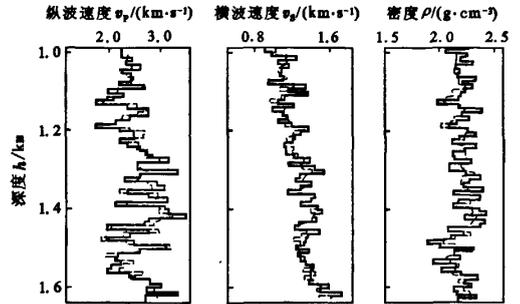


图 3 反演的纵波速度、横波速度及密度曲线

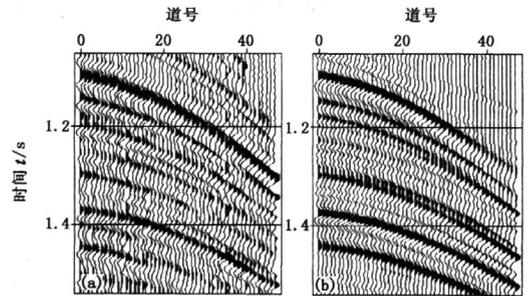


图 4 实际叠前记录(a)及合成记录(b)

另外, 还采用改进的并行算法利用不同节点对不同问题规模进行模拟, 图 5 和图 6 中分别给出了所用的时间和并行效率。

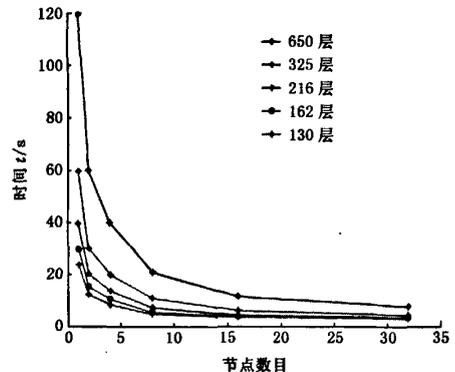


图 5 不同问题规模采用不同节点的并行时间

由图 5 和图 6 不难看出:

- (1) 随着节点数目的增多, 并行时间呈下降趋势, 但增加到一定的程度, 节点数目增加的同时也增加了任务之间的通信开销, 并行时间呈增长趋势, 效率呈下降趋势。
- (2) 随着问题规模的增加, 效率呈上升趋势。

由于规模的增加,问题的计算时间占总的运行时间的比重增加,而任务之间的通信时间占总的运行时间的比重相对就会减少。

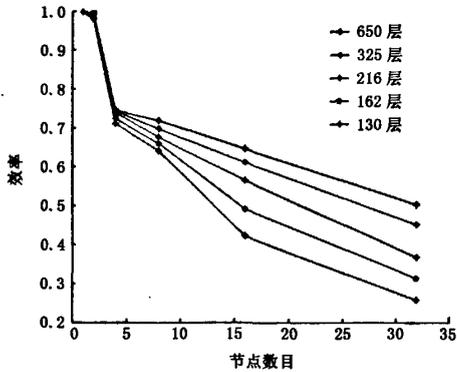


图6 不同问题规模采用不同节点的并行效率

4 结束语

对于非线性、多参数、多极值的叠前弹性波问题,常规的遗传算法或者一些简单的遗传算法其效果并不理想,本文中对遗传算法中的遗传操作采取针对性的改进策略,并通过加入界约束以降低解的非唯一性。通过实际应用,在克服早熟收敛、扩大搜索范围、提高精度等方面,收到了显著的效果。另外,在并行遗传算法设计过程中,选择合适的迁移拓扑结构和迁移策略,按照并行设计方法进行算法设计,尽量减少通信开销和任务创建代价,选择合适的通信模式以免造成通信瓶颈。研究多角度综合的进化搜索策略对于推动地震反演技术的研究与发展是有意义的。

参考文献:

- [1] 王宝珍, 杨文采. 用改进的遗传算法进行地震波阻抗反演研究[J]. 石油地球物理勘探, 1998, 33(2):258-264.
WANG Bao-zhen, YANG Wen-cai. Seismic wave impedance inversion using improved hereditary algorithm[J]. Oil Geophysical Prospecting, 1998, 33(2):258-264.
- [2] 李晶, 陈裕明, 王振国. 用改进的遗传算法进行地震波阻抗反演[J]. 新疆石油地质, 2004, 25(2):150-152.
LI Jing, CHEN Yu-ming, WANG Zhen-guo. Improve genetic algorithm for wave impedance inversion[J]. Xinjiang Petroleum Geology, 2004, 25(2):150-152.
- [3] 邹强, 周熙襄, 钟本善. 地球物理反演中遗传算法编码进制的讨论[J]. 新疆石油地质, 2004, 25(3):337-339.
ZOU Qiang, ZHOU Xi-xiang, ZHONG Ben-shan. Discussion on genetic algorithm encoding scale in geophysical inversion[J]. Xinjiang Petroleum Geology, 2004, 25(3):337-339.
- [4] 石琳珂, 孙铭心, 王广国, 等. 地球物理遗传反演方法[M]. 北京:地震出版社, 2000.
- [5] ZHANG Fanchang, YIN Xingyao. Spherical prestack seismicogram forward modeling and genetic algorithm inversion[R]. SEG/CPS International Exposition and Conference, Beijing, 2004-04-03.
- [6] 张繁昌, 印兴耀. 用多层前馈网络进行三维储层参数反演的方法[J]. 石油大学学报:自然科学版, 2004, 28(1):31-35.
ZHANG Fan-chang, YIN Xing-yao. Three-dimensional seismic inversion with a fast multi-layer feed-forward neural network[J]. Journal of the University of Petroleum, China (Edition of Natural Science), 2004, 28(1):31-35.
- [7] OOSTRANDER W J. Plane-wave reflection coefficients for gas sands at non-normal incidence[J]. Geophysics, 1984, 49(10):1637-1648.
- [8] KUMAR V, GRAMA A, GUPTAA, et al. Introduction to parallel computing[M]. New York: Benjamin Cummings Press, 1994.
- [9] SNIR M, OTTO S, HUSS-LEDERMAN S, et al. The complete reference[M]. Cambridge: MIT Press, 1996.
- [10] 都志辉. 高性能计算之并行编程技术——MPI并行程序设计[M]. 北京:清华大学出版社, 2001.

(编辑 修荣荣)

石油百科知识:综合录井(一)

钻时录井 即记录钻探每米进尺所用的时间。钻时能表明钻头破碎岩石的快慢。地下岩石是各种各样的,有砂岩、砾岩、泥岩、页岩以及石灰岩、火山岩、变质岩等不同的岩石,它们本身的坚硬程度不同,钻穿它们所耗时间也就不一样。这样,就可根据钻头破碎不同类型岩石的时间来推断可能的岩石类型。通常泥岩要比砂岩难钻一点,而石灰岩、火山岩一般比较坚硬,所用时间就长。