

文章编号:1671-8585(2009)04-0265-05

基于 MPI 并行实数编码混合遗传算法的 波阻抗反演

白俊雨,赵俊省,潘凌飞,郭嵩魏

(成都理工大学地球探测与信息技术教育部重点实验室,四川成都 610059)

摘要:地球物理反演的局部线性方法易使解陷入局部极值,并严重依赖初始模型,而传统的遗传算法在优化应用中存在局部搜索能力弱、早熟收敛等问题,为此,提出了一种解决地球物理反演问题的并行实数编码混合遗传算法(MRCGA)。该方法采用拟网格法初始种群、综合交叉策略和线性算子,实现了并行实码混合遗传算法。理论模型试算证明了该算法反演地震波阻抗的有效性。

关键词: MPI;遗传算法;实数编码;初始种群;综合交叉策略;线性算子;波阻抗反演

中图分类号: P631.4

文献标识码: A

波阻抗反演是储层预测的有效手段之一,波阻抗与含油气储层有很好的对应性^[1]。波阻抗反演方法多种多样^[2~5],其中最常用的是广义线性反演,虽然该方法有较高的反演精度,但是在反演过程中,高频噪声对解病态方程组的影响很大,即使加了正则化因子,也无法消除高频噪声的影响。

地震资料的波阻抗反演是一个非线性问题^[6]。反演中的目标函数是一个非凸性的复杂多峰函数,如果使用对初始模型有很大依赖性的线性优化方法进行波阻抗反演,则容易陷入局部极值而难以得到全局解。近年来,出现了具有全局优化特性且通用性强的随机搜索算法,如遗传算法、模拟退火算法、蚁群算法和人工神经网络法等,其中遗传算法、模拟退火算法和人工神经网络等算法^[7,8]已广泛应用于地震反演中。限于计算机技术的发展,人们研究的传统遗传算法都是串行的,但是遗传算法自诞生起就是并行的,串行遗传算法虽然在某种程度上能够克服局部反演法的缺点,但是存在二进制编码反演参数少且效率低的不足,不能充分发挥遗传算法全局寻优的特性^[9~11]。

本文对传统遗传算法进行了改进,将其与线性优化方法结合,使用消息传递接口 MPI(Message Passing Interface)实现了粗粒度并行实码混合遗传算法(MRCGA)。理论模型试算和分析证明了该算法的有效性。

1 方法原理

1.1 MPI 介绍^[12]

消息传递接口 MPI(Message Passing Inter-

face)是一个消息传递函数库的标准说明,它吸取了众多消息传递系统的优点,是目前国际上最流行的并行编程环境之一,尤其是分布式存储的可缩放并行计算机和 workstation 网络以及机群的一种编程范例。在基于的 MPI 编程模型中,计算是由一个或多个彼此通过调用库函数进行消息收、发通信的进程所组成。在绝大部分 MPI 实现中,一组固定的进程在程序初始化时生成。一般情况下,一个处理器只生成一个进程。这些进程可以执行相同或不同的程序,相应地称为单程序多数据(SMPD)或多程序多数据(MPMD)模式。进程间的通信可以是点到点的,也可以是集合的。

MPI 只是为程序员提供一个并行环境库,程序员通过调用 MPI 的库来达到程序员所要达到的并行目的。

1.2 串行遗传算法实现过程

1.2.1 初始种群生成

针对已有的研究成果及所存在问题^[13~15],本文采用拟网格法,即把各待优化参数的搜索范围均匀分割,在各小区间端点序列中随机生成一个初始个体。如:在 $[0, d]$ 中产生一个随机整数 k_i , d 为第 i 个参数所在区间被分割的份数, k_i 即为个体的元素。把待优化参数对应的 k_i 依次连在一起构成问题解的编码形式 (k_1, k_2, \dots, k_n) 。以三维空间为例,设参数 $X_1 \in [X_{1\min}, X_{1\max}]$, $X_2 \in [X_{2\min}, X_{2\max}]$, $X_3 \in [X_{3\min}, X_{3\max}]$; 将区间 $[X_{1\min}, X_{1\max}]$, $[X_{2\min}, X_{2\max}]$, $[X_{3\min}, X_{3\max}]$ 均匀

收稿日期:2009-02-26;改回日期:2009-03-20。

第一作者简介:白俊雨(1984—),男,在读硕士研究生,主要研究方向为地震资料叠前反演及叠前偏移成像。

分割为 N_1, N_2, N_3 份(图 1)。在 $[0, N_1], [0, N_2], [0, N_3]$ 上分别随机产生整数 k_1, k_2, k_3 , 则 (k_1, k_2, k_3) 即为个体。在计算适应值时将 k_i 做如下映射:

$$X_i = X_{\min} + k_i * \Delta_i \quad (1)$$

式中: $\Delta_i = \frac{X_{\max} - X_{\min}}{N_i}$; $X_i \in [X_{\min}, X_{\max}]$, $i = 1, 2, \dots, N$ 。这样生成的初始个体将很大程度地均匀分布在解的空间上,并能保证随机产生的个体间有明显的差距,增大了得到全局最优解的可能。这样做可以大大加快算法的收敛速度。

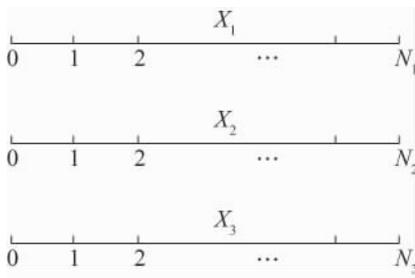


图 1 拟网格法示意

1.2.2 选择操作

本文采用精英选择策略。如果当前群体的最佳个体适应值大于下一代群体最佳个体的适应值,则将当前群体最佳个体或适应值大于下一代最佳个体适应值的多个个体直接复制到下一代,随机替换适应值或者替换适应值最差者。

1.2.3 交叉操作

本文采用综合的交叉算子,其全局寻优能力较常规的交叉算子显著提高。具体实现方法如下:

1) 交叉算子 1。

设 X_i^t 和 X_j^t 分别为第 t 代要进行交叉的两个个体,则交叉后产生的子代个体为

$$\begin{cases} X_i^{t+1} = X_i^t + \tau_1 (X_i^t - X_j^t) \\ X_j^{t+1} = X_j^t + \tau_2 (X_i^t - X_j^t) \end{cases} \quad (2)$$

式中: τ_1 和 τ_2 为 $[0, 1]$ 上均匀分布的随机数。观察式(2)的交叉操作,可以发现这种交叉策略不允许后代个体处于区间 $[X_j^t, X_i^t]$ (假设 $X_j^t < X_i^t$) 之外。后代个体位置的这种限制没有最大化改善后代个体的基因模式,限制了整个群体的搜索速度,并有未成熟收敛于局部最优的可能,后代要跳出 $[X_j^t, X_i^t]$ 之界限,只能借助于算法中的变异操作,而变异概率很小。为此,本文引入交叉算子 2 和交叉算子 3。

2) 交叉算子 2。

改变式(2)中 τ_1 和 τ_2 为 $[-1, 0]$ 上均匀分布的随机数。

3) 交叉算子 3。

改变式(2)中 τ_1 和 τ_2 为 $[1, 2]$ 上均匀分布的随机数。

综合使用上述 3 个交叉算子可保证搜索区域覆盖整个搜索区间,也有可能处于搜索区间之外,从而克服了只用交叉算子 1 的缺陷。

4) 交叉算子 4^[16]。

引进潜在搜索方向 d , 设交叉对为 X_{low}^t 和 X_{high}^t , 对应的适应度为 f_{high} 和 f_{low} , $f_{\text{high}} > f_{\text{low}}$, 有

$$d = \frac{f_{\text{high}} - f_{\text{low}}}{f_{\text{high}}} \quad (3)$$

则交叉后产生的子代个体为

$$\begin{cases} X_1^{t+1} = X_{\text{low}}^t + \alpha_1 (X_{\text{high}}^t - X_{\text{low}}^t) d \\ X_2^{t+1} = X_{\text{low}}^t + \alpha_2 (X_{\text{high}}^t - X_{\text{low}}^t) d \end{cases} \quad (4)$$

式中: α_1 是 $[0, 1]$ 上均匀分布的随机数, α_2 是 $[1, 2]$ 上均匀分布的随机数, 这样 X_1^{t+1} 将处于 $[X_{\text{low}}^t, X_{\text{high}}^t]$ 界内, X_2^{t+1} 有可能处于 $[X_{\text{low}}^t, X_{\text{high}}^t]$ 界外也可能处于整个搜索区间之外。

从加快收敛速度和全局搜索性能两方面考虑,受自然界中家庭内兄弟间竞争现象的启发,在算法中加入小范围竞争、择优操作。方法是:对将要进行交叉的父母对 A, B 进行 n ($n=3 \sim 6$) 次交叉,生成 $2n$ 个个体,选出最优的两个个体进入子代。该方法的实质是在相同父母的情况下,预先加入兄弟间的小范围的竞争择优机制。

本文的交叉操作依次运行上述 4 个算子,且依次对选中交叉的父母对都进行 6 次交叉,在产生的 48 个子代和对应的父母对中进行优选,选出两个最优的个体进入子代。

1.2.4 变异操作

按照一定的变异概率对选中的个体进行高斯变异操作^[17]。

1.2.5 线性算子的引入

针对非线性方法反演速度慢的问题,将采用实数编码遗传算法与线性反演方法相结合的方法予以解决。把进化一定代数后种群中的个体作为线性反演方法的初始模型,将线性反演方法迭代一定次数后的结果返回种群中,更新种群。本文的线性算子采用地震道的广义线性反演方法^[5]。

2 波阻抗反演的目标函数^[18]

本文从褶积模型出发,采用地震道非线性反演,目标函数设为

$$f(i\Delta t, Z) = \left\| D(i\Delta t) - W(t) * \sum_{i=1}^N \frac{Z_{i+1} - Z_i}{Z_{i+1} + Z_i} \delta(t - i\Delta t) \right\| \rightarrow \min \quad (5)$$

式中: $D(i\Delta t)$ 为实际地震记录; $W(t)$ 为地震子波; Z 为波阻抗; N 为反射系数长度; Δt 为采样间隔。

3 并行遗传算法^[19,20]设计

本文采用独立型-粗粒度模型(coarse-grained model),将种群分成若干个子种群并分配给各自对应的处理器,每个处理器不仅独立计算适应度,而且独立进行选择、重组交叉及变异操作,还要定期地相互传送一定数目的适应度优良的个体(迁移策略),从而加快满足终止条件的要求。为协调种群多样性和算法收敛速度的辩证关系,采用一传一的迁移策略,并用较小的迁移率和较大的迁移间隔。

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 决定处理器之间通讯的频率,如 send_rate=3 时表示当进化代数(Generation)是 3 的倍数时(即 $Generation \% send_rate = 0$),各处理器之间相互传个体;二是 send_best 决定每次传送最佳个体的数目,如 send_best=5 时表示每个处理器把最佳的前 5 个个体传给各自的相邻处理器。

本文并行遗传算法的拓扑结构及流程见图 2 和图 3。

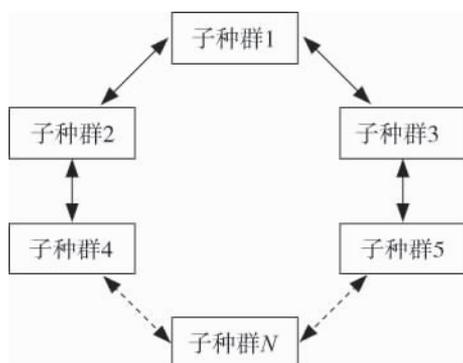


图 2 环状拓扑

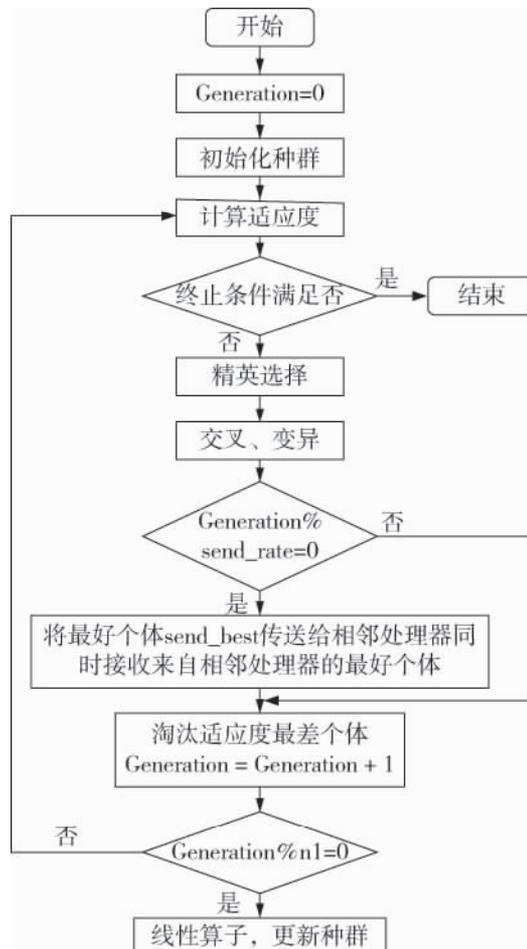


图 3 MRCGA 程序流程

4 算例

为检验该算法的正确性与可行性,设计了理论模型如图 4。该模型共有 12 层,其中有 2 个薄层,密度取 $1g/cm^3$ 。子波采用主频为 35 Hz 的雷克子波,采样率为 2 ms。理论合成记录见图 5。

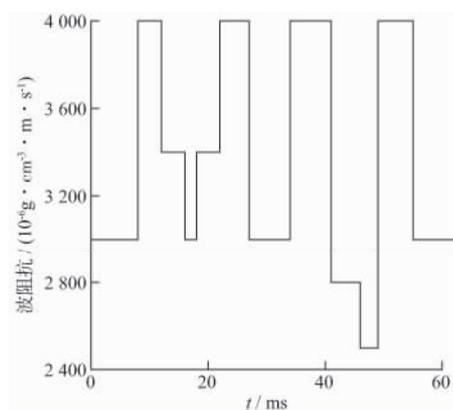


图 4 理论波阻抗模型

反演过程中采用的基本参数及种群大小和进化代数的选择原则是:种群越大,进化代数越多,则

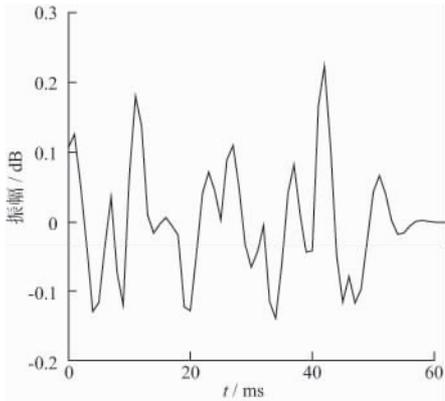


图 5 理论模型合成记录

算法全局寻优能力越强,但是计算效率低。综合计算效率和全局寻优性能两方面考虑,本次选取子种群大小 N_i_Pop 为 300,子种群最大进化代数 N_i_Gen 为 1 000;每进化 n 代调用一次线性算子来更新种群,本次取 $n=30$;处理器之间通讯频率 $send_rate$ 取为 20,表示当进化代数是 20 的倍数时,各处理器之间相互传送个体;子种群向相邻子种群(即相邻处理器)每次传送最佳个体的数目 $send_best$ 取为 3。

以上参数设计的原则是保证当前子种群进化到一定优良状态时再向相邻处理器传送当前子种群中最好的个体,为保证种群多样性,每次传送最好个体数目不要太多。

图 6 为两种反演算法进化 1 000 代,目标函数值随进化代数的变化。可以看出并行实码混合遗传算法在进化 200 代的时候目标函数值已经接近于 0,而常规串行遗传算法却在种群进化一定程度后没有发生明显的进化,可见并行实码混合遗传算法的全局寻优能力明显高于常规串行遗传算法。

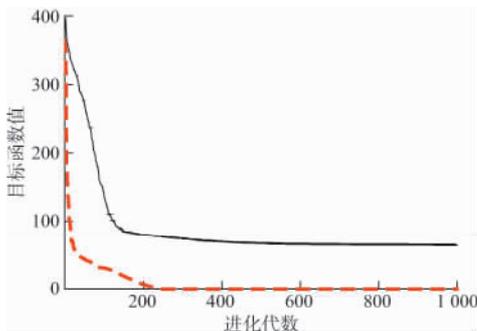


图 6 并行实码混合遗传算法(虚线)与串行遗传算法(实线)反演的目标函数值随进化代数变化对比

图 7 和图 8 为两种算法进化 1 000 代后的最优个体的合成记录,并行实码混合遗传算法反演结果的合成记录与模型合成记录的相关系数大于常规串行遗传算法反演结果合成记录与模型合成记

录的相关系数。图 9 为两种反演结果的对比,并行实码混合遗传算法反演结果与模型整体形态已基本一致,而常规串行遗传算法的反演结果却远远不如并行实码混合遗传算法的反演结果。

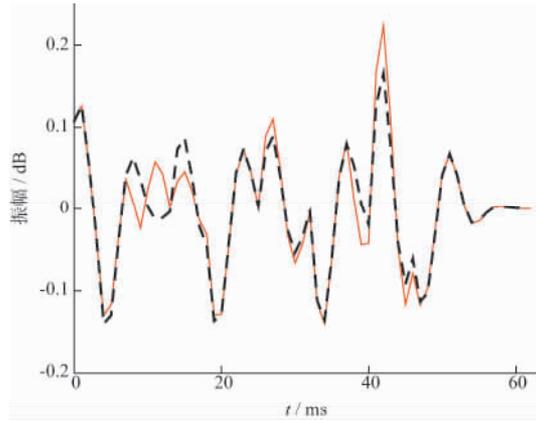


图 7 常规遗传算法反演结果的合成地震记录(虚线)与模型合成地震记录(实线)对比

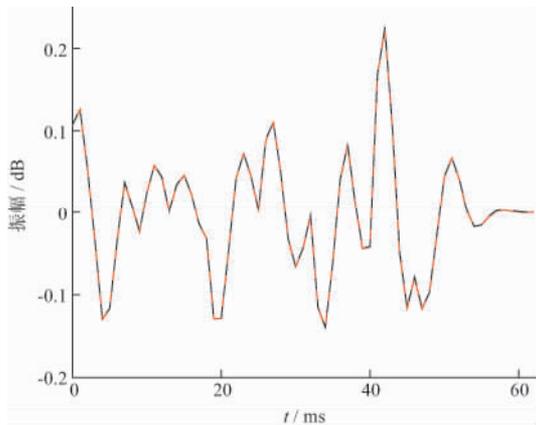


图 8 并行实码混合遗传算法反演结果的合成地震记录(虚线)与模型合成地震记录(实线)对比

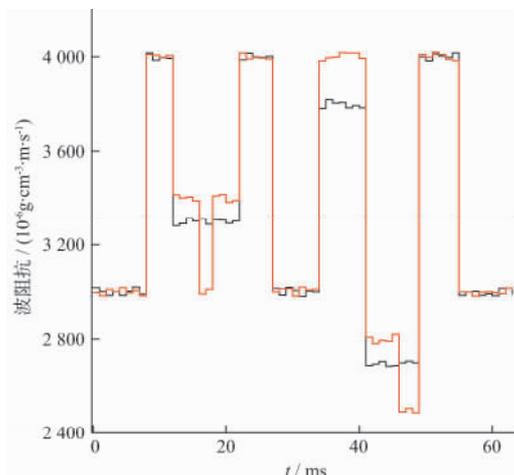


图 9 并行实码混合遗传算法反演结果(实线)与串行遗传算法反演结果(虚线)对比

5 结论

本文通过对理论模型试验结果的分析,证明了对于非线性、多参数和多极值的组合优化问题,常规的遗传算法在种群进化后期存在早熟收敛以及对模型参数的自拷贝现象(因为在进化后期,种群中个体都很相似,普通交叉操作已很难产生更优良个体,更优良个体的产生则主要依靠变异操作,而变异概率通常很小),需进行大量正演计算,其效果并不一定很好。而本文的并行实码混合遗传算法可以很好地解决进化后期参数的自拷贝问题,快速地求出全局最优解。经过模型地震道序列数据的反演计算,证实了该方法用于地震资料波阻抗反演的高效性。

参 考 文 献

- 1 王西文,石兰亭,雍学善,等.地震波阻抗反演方法研究[J].岩性油气藏,2007,19(3):80~88
- 2 李庆忠.走向精确勘探的道路——高分辨率地震勘探系统工程剖析[M].北京:石油工业出版社,1993.1~196
- 3 杨文采.地震道的非线性混沌反演(一)[J].地球物理学报,1993,36(2):223~231
- 4 杨文采.地震道的非线性混沌反演(二)[J].地球物理学报,1993,36(2):376~387
- 5 邹振桓,杨文采.地震道的广义线性反演[J].石油地球物理勘探,1987,22(4):363~374
- 6 杨文采.地球物理反演的理论与方法[M].北京:地质出版社,1997.11~35
- 7 张永刚.地震波阻抗反演技术的现状和发展[J].石油物探,2002,41(4):385~390
- 8 姚姚.地球物理反演基本理论与应用方法[M].武汉:中国地质大学出版社,2002.72~82
- 9 廖平.归一化实数编码的多维并行遗传算法[J].计算机仿真,2005,22(10):122~124
- 10 陈辉,张家树,张超.实数编码混沌量子遗传算法[J].控制与决策,2005,20(11):1300~1303
- 11 邹琳,夏巨湛,胡国安.基于实数编码的多种群并行遗传算法研究[J].小型微型计算机系统,2004,25(6):982~986
- 12 都志辉.高性能计算并行编程技术——MPI并行程序设计[M].北京:清华大学出版社,2001.1~336
- 13 Zalza A M, Fleming P J. Genetic Algorithms in Engineering Systems[M]. London: The Institution of Electrical Engineers, 1997. 1~263
- 14 李志俊,程家兴.基于数论佳点集的遗传算法初始种群均匀设计[J].电脑与信息技术,2007,15(4):29~32
- 15 蒋泰,陈洛均,黄源.带有约束优化的遗传算法求解TSP[J].计算机应用研究,2008,25(5):1323~1325
- 16 陈小平,于盛林.实数遗传算法交叉策略的改进[J].电子学报,2003,31(1):71~74
- 17 李鹏,董聪.基于实数编码的广义遗传算法及其在优化问题中的应用[J].控制与决策,2002,17(4):487~490
- 18 渥·伊尔马滋.地震资料分析——地震资料处理、反演和解释[M].刘怀山,王克斌,童思友,等译.北京:石油工业出版社,2006.126~128
- 19 郭绚,石晓虹.并行遗传算法的性能分析[J].航空计算技术,1998,28(3):86~89
- 20 戴晓明,陈昌领,邵惠鹤,等.粗粒度并行遗传算法收敛性分析及优化运算[J].上海交通大学学报,2003,37(4):500~502

(编辑:任 鹏)