重磁反演的群体优生遗传算法

张小路
(桂林工学院·桂林·541004)

重磁反演常具有众多变量,通常的遗传算法编码方式和搜索模式不能适用。采用0—1编码群体优生遗传算法,实现了把遗传算法运用于众多变量的重磁反演,使用进化变异能使运算后期顺利收敛,同时更加接近全局极值,取得了良好的实用效果。

关键词 遗传算法 群体优生 进化变异 重磁反演

1 引言

重磁拟合法自动反演属于非线性最优化问题,具有多变量、目标函数多极值、反演多解性等特点。

遗传算法(Genetic algorithm)是全局最优化随机搜索方法中的一种,是由 Holland 于 1975 年提出来的,是一种以概率论为基础的求解多极值非线性问题的最优化方法。它模仿生物界自然选择和遗传规律,以适者生存、优胜劣汰为原则,在模型参数空间进行完全搜索,逼近全局极值。

通常使用的遗传算法,都是用二进制编码来表示模型参数,表示一个参数就需要很多位,要求的分辨率越高,编码就越长。如果为单个参数或参数不是太多时(几至十几),使用起来还算方便,但是当参数众多时(几十至上百),使用起来就很不方变,以致不能适应众多变量的反演计算,至今尚未发现把遗传算法应用于太多变量的反演。为了解决多变量反演中节省计算机内存和方便译码的问题,周辉等[1]提出了0-1编码方式(1997),这种编码方式使用方便,能适应众多变量的反演问题。

2 方法介绍

2.1 目标函数

遗传算法重磁反演属于非线性最优化问题,可由下式表示:

$$\| f_i(r) - g_i(r,p) \| < \varepsilon \qquad (1)$$

$$s.t.\ p_l \leqslant p \leqslant p_u$$

式中 $f_i(r)$为实测函数,$g_i(r,p)$为理论函数,ε 为任一小量,r 为测点空间坐标,p 为参数变量,(p_l, p_u)为约束范围,问题的解是求异常残差的极小值。

目标函数可取异常残差 l_1范数 $\psi_1(p)$或 l_2范数 $\psi_2(p)$,

$$\psi_1(p) = \frac{1}{N}\sum_{i=1}^{N} | f_i(r) - g_i(r,p) | \qquad (2)$$

$$\psi_2(p) = \left\{ \frac{1}{N}\sum_{i=1}^{N} [f_i(r) - g_i(r,p)]^2 \right\}^{1/2} \qquad (3)$$

实测数据都存在各种干扰,因为 l_2范数含有二次项,抗干扰的能力较差,所以取异常残差的 l_1范数作为反演的目标函数。

2.2 遗传算法

2.2.1 参数编码和初始模型参数群体的产生

设模型由 M 个参数组成,表示为 $P_i(i = 1,2...M)$,约束范围为(P_{li}, P_{ui}),要产生 L 个模型群体作为搜索空间,参数编码采用 0—1 编码。所谓 0—1 编码,就是把每个参数的编码用[0,1]之间的一个数来表示,对应每一个参数,都在区间[0,1]之间产生 L 个均匀分布的随机数,表示为 $r_i(i = 1,2...L)$,就是这个参数的 L 个编码,L 的取值根据模型参数所期望的分辨率决定。M 个参数的 L 组模型共有 $M \times L$ 个编码。

在约束条件下把表示参数的 $M \times L$ 个编码译成参数值,译码规则为

$$P_{i,j} = P_{li,j}^{*} + r_{i,j}(P_{ui,j} - P_{li,j}) \qquad (4)$$

$$(i = 1,2...M),(j = 1,2...L)$$

按照遗传算法,参数群体中的一个编码就代表一个基因,表示一个模型的 M 个基因构成一条染色体,群体中共有 L 条染色体。每条染色体译码后才能计算出这个模型的目标函数。首先要计算初始群体的 L 个目标函数,以目标函数作为判据来进行搜索。一般来说,遗传算法的繁殖搜索过程是在遗传编码中进行的,每次迭代都要先译码再计算目标函数。这对于二进制编码的算法是必要的,因为一个参数的编码有多位,也就是说由多个基因组成,遗传时一个参数不一定改变全部基因,因此每次繁殖都要先译码才能计算目标函数。但对于 0-1 编码来

本文 1998 年 6 月收到,王延忠编辑。

说，一个参数的编码只有一个数，就是只有一个基因组成，参数和编码的作用实质上是一样的，因此，在初始模型产生并译码后，遗传算法的繁殖过程就可以直接在参数群体中进行，每个参数变量就构成一个遗传基因，每个参数模型就构成一条染色体，遗传过程中如果要产生新的编码，则随即译成参数，加入群体。这样，遗传算法的实质不变，又节省了译码的运算量。

2.2.2 繁殖(Reproduction)过程

遗传算法的繁殖过程即是优化过程，每完成一代新的繁殖，都应该出现优于上一代的参数组合。繁殖是通过选择(Selection)、交叉(Crossover)和变异(Mutation)3个过程来完成，每个过程的作用不同。

1)优良选择

在 L 个染色体中选出较优秀的作为交叉过程的父本。目标函数最小者视为最佳，选择过程首先对 L 个染色体从小到大排序，按一定概率选择排在前面的染色体，需要交叉多少对，就选择多少个父本，然后由选择好的父本在群体中随机选择配对。这样选择是为了既能选出群体中优秀的遗传基因，又能在模型空间中进行广泛的搜索，涉取好的基因。选择多少个父本为好，应该考虑整个算法过程的效率。因为每产生一对新的后代，都要经过正演计算产生两个新的目标函数，而计算目标函数是反演过程的主要计算量，如果要尽量减少计算目标函数的次数，就要尽可能使新生的一代被优化的概率提高。如果每次父本选择过多，排在后面的就不够优秀，产生的后代被优化的概率降低，优化的程度也可能降低，增加了无益目标函数的计算量，收敛速度变慢。如果每次父本选择过少，遗传基因不够丰富，不能保证广泛的搜索，收敛速度也会降低。所以，用一个合适的概率选择父本是必要的。试验表明，选用染色体的5%为父本比较合适，即父本数量 = 取整($L \times 0.05$)，每次交叉可产生约10%的新生成员。

2)交叉与比较

交叉的过程就是交换基因过程，对每一对需要交叉的染色体都随机选择一个位置，把位置右边的基因全部互换，而在被选择的位置上随机产生一个新的基因，置换原来的基因。对新产生的一代染色体计算目标函数并进行比较，好的保留，差的淘汰。

通常所用的比较方法，是把新产生的一对染色体与其父本自身比较，4个中保留两个目标函数较小的，收敛过程中以群体中目标函数最小者为判别标准，最终选择目标函数最小的模型作为反演结果。这样的做法实际上是在群体模型空间中只进行部分模型搜索。在参数较少的情况下，在整个模型空间中进行部分搜索能找到全局最优解，计算的次数也较少。但在参数较多的情况下，由于多解性的增加，必须在整个模型空间中进行完全搜索，才能逼近最优解，这就需要增加遗传次数，直至群体模型的参数趋于一致。采用部分搜索方式进行完全搜索，势必要增加繁殖迭代的次数，使收敛速度变慢。如果取模型参数的平均值作为反演结果，在收敛过程中还会产生跳跃。

如果要加快收敛过程，就应该直接采用完全搜索方式。基于群体优生的设想，对于新产生的染色体，每次都与群体中最差的染色体相比较，每次都淘汰最差的染色体，使群体逐步达到优化。目标函数最大者是最差的染色体，在搜索过程中与目标函数最大者相比较，选择目标函数最大者作为判别标准，收敛到小于 ε 值计算结束，取群体模型的平均值作为反演结果。这样做是在模型空间中进行完全搜索，收敛过程中不会发生跳跃，能够加快群体收敛的速度，提高了优化效率。取模型的平均值作为反演结果更能接近全局最优模型。群体优生方式更适用于众多变量的反演计算。

3)进化变异

变异是遗传算法中的一个重要环节，是通过让某些基因发生变异而产生进化。如果没有变异过程，就没有足够的新的基因产生，在繁殖过程中会造成还未接近全局极值就不再收敛的后果。

通常意义下的变异，是在群体中以一个较小的概率，随机改变染色体中的一个基因，变异后的染色体与原本比较，如果较好则保留，说明产生了进化，否则保留原本。但是这种方法产生的变异进化的概率是较小的，往往达不到预期的目的，群体距全局极值较远。为了改变这种状态，人们设想了一些办法，有人引入灾变过程，就是在某时引入突发性的大规模变异，使多个染色体随机变异多个基因，促使较多的变异基因产生进化。这样虽然能使群体状态得到改善，但代价过高，每个新产生的染色体都要重新计算目标函数，重新判别，增加了总体计算量，而且，所有的变异基因都是随机产生的，与初始模型类似，被优化的概率仍然不高。

如果要使每一代繁殖都能保持较好的状态，就必须提高变异进化的概率。如果每次变异都能使染色体产生进化，每代繁殖就只需要一次正常的变异，就能使群体进化繁殖顺利地进行下去，本文所采用

的进化变异过程就是为了达到这个目的。

为了在整个空间进行更完全、更广泛的搜索,采用的方法是,在变异过程中,让变异染色体的每一个基因都发生改变,增加多种新的基因,为了让新产生的变异基因尽可能的优化,不是任意随机产生,而是取群体中多个基因的平均值。做法是,在群体中随机选取多个染色体,(本文选取 M 个,与变量数相同),把每条染色体对应的基因相加平均,取平均值作为新的基因,这样就产生了一条全新的染色体。在比较时仍然采用群体优生的原则,把变异染色体的目标函数与群体中目标函数极大者相比较,保留目标函数较小的。若变异染色体得到了保留,就是得到了进化。采用这种方法变异产生的新染色体,几乎 100% 是进化变异,每次繁殖只用 1% 至 2% 的小概率产生变异染色体就足够了,减小了总体计算量。本文采用的变异数 = 取整$(L\times0.01)+1$。

变异过的基因成为群体中新的遗传基因,不断的更新使群体繁殖一直保持良好状态,在多代的繁殖搜索中顺利逼近全局极小值。进化变异的过程恰好模拟了生物演化中的基因转变过程,完成了预期的基因转化,就可获得新一代的优良品种。

采用了优良选择、群体优生和进化变异,不但使众多变量的遗传算法得以实现,而且使算法变得简单易行,加快了总体收敛速度,切实成为一种有实用意义的算法。

2.2.3　群体大小和 ε 值

群体中共有 L 条染色体,每条染色体有 M 个基因。L 的大小直接影响反演计算的收敛过程。把 L 取的足够大,参数分辨率高,遗传基因丰富,有利于收敛过程。但是 L 并不需要取的太大,因为在约束范围内,L 个染色体是随机产生的,在繁殖过程中还会不断的产生新参数,L 值不必过大,也能达到模型参数的期望分辨率,L 过大反会造成搜索过程延长,收敛减慢,还会使后期收敛状态变坏。L 过小,不能做到完全搜索,难以逼近全局极小点。因此需要取一个适当的 L 值。要保证 L 大于 M,M 增大,L 也要随之增大,但不需要成倍增加,因为 M 大基因数多,交叉和变异的组合形式也多,所以,M 越小倍数越大。试验表明,L 的取值范围约在 1.1 M ~ 2.4 M,参考数据列在表 1。

迭代过程结束的判别标准 ε 值没有固定的选择方法,要根据数据的整体大小和噪声水平来定。数据大,拟合差的绝对值就大,ε 就应该选大些,否则难于收敛至 ε;噪声水平越高,干扰就越大,ε 值也应该选大些,与噪声水平相当,如果 ε 小于噪声水平,数据会向干扰拟合,使反演结果变差。但一般没有试算之前,很难一次选定 ε 值。在计算时,ε 值可逐步选择,先给出一个较大的数,收敛到小于 ε 后再进一步给出一个小一些的数,直至收敛较慢,迭代结束。给出 ε 减小的步长越到后期应该越小。

表 1　群体大小的选择试验数据

M	13	27	43	64	117	247
L	30	60	70	90	150	280

3　反演计算

3.1　理论模型反演计算

对二度体及三度体重磁模型做了多例试验,模型有二度体矩形组合棱柱、二度和似二度体多边形组合、三度体矩形组合棱柱及三度体单体棱柱组合等。不同理论模型的反演试验取得了相似的效果,在此选取其中一例,来说明反演方法的使用及效果。

选取由 3 个多边形组成的似二度体复杂模型,模拟脉状矿体和似层状矿体,进行剖面△T 磁反演。参数包括角点横坐标、纵坐标、走向长度坐标、磁化强度、磁化倾角和磁化偏角,共 43 个变量。

把理论模型作正演计算(图 1(a)),然后再加上 ±20 nT 随机干扰,模拟实测曲线(图 1(b))进行反演计算。正演模型的参数、约束范围和反演结果列于表 2(反演结果与图 1(b)对应),约束范围给的较宽,大部分参数不在中心。

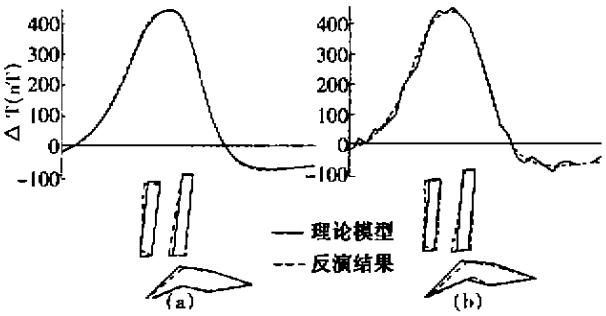

图 1　完全搜索方式反演结果

(a)—不加干扰(目标函数平均值 4.63);

(b)—加干扰(加 ±20nT 随机干扰,目标函数平均值 12.14)

反演试验采用两种方法,一是采用群体优生算法,交叉和变异后与目标函数极大者相比较,以群体中目标函数最大者为判别标准,也就是采用完全搜索方式,取群体的平均值作为反演结果(图 1)。二是采用自身优生算法,交叉后与其自身相比较,变异后随机与一条染色体相比较,以群体中目标函数最小者为判别标准,采用的是部分搜索方式,取群体中目标函数最小者作为反演结果(图 2)。反演中参数

M 为 43 个，群体染色体总数 L 都取 70 个，每代繁殖交叉 3 对，变异 1 条染色体。

表 2 理论模型参数、约束范围和反演结果

变量	真值	约束范围	结果	变量	真值	约束范围	结果
$x11$	220	170 ~ 260	211	$y22$	350	300 ~ 400	346
$x12$	250	210 ~ 300	254	$J2$	1000	500 ~ 1400	905
$x13$	230	200 ~ 270	232	$I2$	50	40 ~ 60	50
$x14$	200	180 ~ 320	201	$D2$	20	10 ~ 30	19
$z11$	100	70 ~ 140	105	$x31$	300	250 ~ 380	314
$z12$	100	70 ~ 140	108	$x32$	380	300 ~ 460	382
$z13$	300	250 ~ 360	301	$x33$	500	430 ~ 550	487
$z14$	300	240 ~ 350	288	$x34$	380	310 ~ 430	373
$y11$	350	300 ~ 400	352	$x35$	310	270 ~ 370	330
$y12$	350	300 ~ 400	348	$x36$	200	170 ~ 250	209
$J1$	1000	500 ~ 1400	913	$z31$	330	300 ~ 370	334
$I1$	50	40 ~ 60	51	$z32$	340	300 ~ 400	344
$D1$	20	10 ~ 30	19	$z33$	380	320 ~ 430	381
$x21$	310	250 ~ 390	317	$z34$	400	330 ~ 460	399
$x22$	340	270 ~ 400	340	$z35$	380	320 ~ 450	372
$x23$	320	250 ~ 400	326	$z36$	430	400 ~ 470	432
$x24$	290	220 ~ 350	275	$y31$	350	300 ~ 400	347
$z21$	80	40 ~ 110	76	$y32$	350	300 ~ 400	354
$z22$	80	40 ~ 110	78	$J3$	800	400 ~ 1100	758
$z23$	300	210 ~ 400	297	$I3$	50	40 ~ 60	51
$z24$	300	220 ~ 400	306	$D3$	20	10 ~ 30	17
$y21$	350	300 ~ 400	343				

图 1(a)是不加干扰的理论模型群体优生算法反演计算结果。ε 取 10，共繁殖 27 代，目标函数极大值为 9.68，平均值为 4.63，结果误差较小。

图 1(b)是加干扰的理论模型群体优生算法反演计算结果。如果 ε 取 20，则需要繁殖 21 代，目标函数极大值为 18.6，平均值为 12.9，结果偏离实际模型不大。如果接着继续反演，ε 取 15，需要再繁殖 6 代，目标函数极大值为 14.9，平均值为 12.1(表 2，图 1(b))，结果变化不大，稍有改善。如果再接着反演，ε 取 12，需要再繁殖 24 代才能完成，收敛已经很慢了，目标函数极大值为 11.9，平均值为 11.4，结果不但不再改进，反而偏差更大些。所以，该例在干扰情况下 ε 取 20 或 15 是恰当的，只需繁殖二十几代就足够了。

图 2 是加干扰的理论模型自身优生算法反演计算结果。ε 先取 15，繁殖 21 代，目标函数极小值为 14.4，平均值为 20.6，取极小值，结果见图 2(a)。继续反演，进一步搜索，ε 取 13，需要再繁殖 21 代，目标函数极小值才能达到 12.8，平均值为 21.1，取极小值，结果见图 2(b)，有所改进，但共需要繁殖四十几代，结果偏离实际模型比图 1(b)要大的多，说明众多变量反演采用群体优生方式为好。

3.2 实例反演计算

广西大厂为世界著名的超大型矿床，矿体赋存在隐伏花岗岩体顶部或边部。采用重力资料，反演隐伏岩体的顶面形态。异常为重力低异常，去掉背景后的异常见图 3。采用群体优生遗传算法，分两步进行反演。

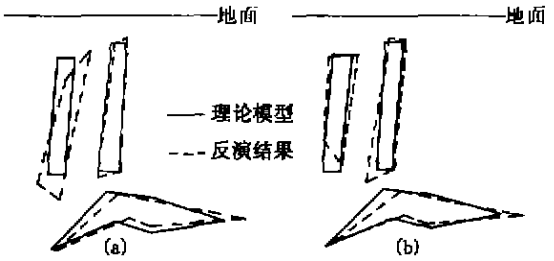

图 2 加干扰部分搜索方式反演结果

(加 ± 20nT 随机干扰)

(a)—目标函数最小值 14.4；(b)—目标函数最小值 12.88

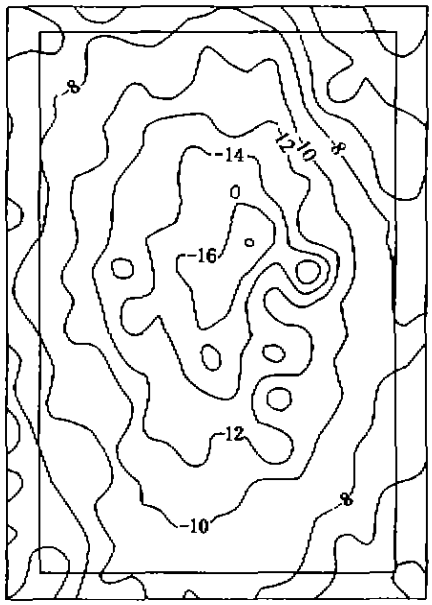

图 3 大厂矿区重力异常及反演岩体模型范围

首先选用 9 个立方柱作为模型进行初步反演，给出较宽的约束范围，顶深 0 km ~ 3 km，底深 2 km ~ 6 km，柱的宽、长和角点坐标位置都给出较大的约束范围，密度差作为已知给出较小的范围 ~ 0.16 g/cm^3 ~ ~ 0.14 g/cm^3。反演计算得出了岩体的分布范围、大致的顶深和底深(图 3)。岩体长约 19 km，宽约 13 km，范围约为 250 km^2，底深为 3.3 km ~ 4.3 km，平均取 4 km，密度差取 ~ 0.15 g/cm^3。

在初步反演的基础上，把岩体的坐标、分布范围、底深和密度差固定，只取顶深为变量。把岩体分成 117 个柱体(13 × 9)，每个柱的长和宽分别为 1.46 km 和 1.45 km，异常中心部位根据已知先验信息给出较窄的约束范围，其余给出较宽的范围。采用压缩质线正演公式计算理论曲线进行反演(图 4)

由图可见，地质构造的分布与岩体的起伏对应很好，主岩体最高处对应三角形断裂，两条隆起带也

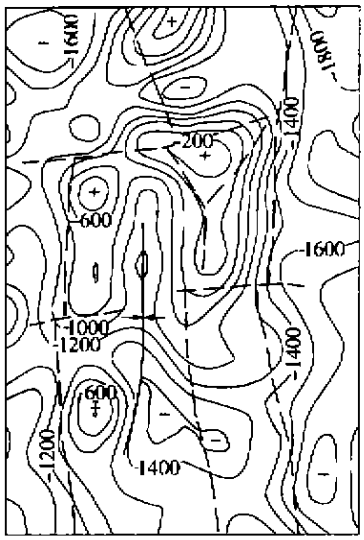

图4　根据每个反演柱体顶深数据绘制的隐伏花岗岩体等深线及地质构造图

对应断裂和背斜轴部,说明岩体沿断裂侵入,侵入过程中又产生次级断裂。岩体中部的凹陷岩沟正好对应向斜轴部,说明岩体侵入时受阻。反演结果与地质构造的对应,证明了反演的有效性,说明该算法具有实用意义和实用效果。

群体优生遗传算法具有反演变量多,收敛速度快的特点,不仅适用于重磁反演,也适用于其它多参数的地球物理反演。

参考文献

1　周辉,等.0-1编码遗传算法.石油物探,1997(36),1

2　Stoffa, Sen. Nonlinear multiparameter optimization using genetic algorithms: Inversion of plane - wave seismograms. Geophysics, 1991(56), 11: 1794 - 1810.

GENETIC ALGORITHM OF POPULATION EUGENESIS OF GRAVITY AND MAHNETIC INVERSION

Zhang Xiaolu

The gravity and magnetic inversion has multitudinous variations. The common binary code mode and the search mode of genetic algorithm are not suitable to it. The 0 - 1 code mode and the genetic algorithm of population optimization mode have applied to the gravity and magnetic inversion of multitudinous variations. The genetic evolution and variation can result in convergence of late stage work, approaching the global limit and good practical results.

Key words　genetic algorithm, population eugenesis, evolution and variation, gravity and magnetic inversion

第一作者简介:

张小路　女,1950年生。1982年毕业于桂林冶金地质学院应用地球物理专业,1987年在长春地质学院获硕士学位。现任桂林工学院副研究员,主要从事应用地球物理学找矿研究工作。

通讯地址:广西壮族自治区桂林市七星区建干路12号　桂林工学院隐伏矿研究所　邮政编码:541004

(上接第30页)

GEOCHEMICAL ANOMALY OF THE ZHEYAOSHAN MARINE VOLCANIC COPPER DEPOSIT

Liu Chongmin

Study on the geochemistry of the Zheyaoshan copper deposit, the indicator elements are characterized by assemblage of Cu, Ag, As, Sb, Hg, Pb, Zn, Au, Cd, and Ba. Based on the anomalies and vertical zoning characteristics, the geochemical anomaly pattern is established.

Key words　Zheyaoshan copper deposit, geochemistry, anomaly pattern

第一作者简介:

刘崇民　男,1955年生。1979年毕业于北京大学地质地理系地球化学专业,1993年在中国地质大学地化系获硕士学位。现任国土资源部物化探所化探高级工程师,主要从事矿产勘查地球化学研究工作。

通讯地址:河北省廊坊市金光道84号　国土资源部地球物理地球化学勘查探研究所　邮编:102849