1　地球物理反演理论和方法的发展与评价

1.1　引　言

有关地球深部的信息，绝大多数来源于对地表地球物理观测资料的解释。因此，任何一种地球物理观测，最终都要求解反演问题，即根据各种位场、地震波地球物理观测资料，来推测地球内部的结构形态和物质成分，定量计算各种地球物理参数（朱介寿， 1985）。

解反演问题是一门广泛应用于各个科学技术领域的基本研究方法，凡涉及到观测与演绎的科学技术，都会碰到这类问题。例如，天体物理学中根据光谱的测定研究宇宙的物质组成及星体的演化；天体力学中根据人造卫星瞬时位置及速度的观测数据计算其轨道要素；气象学中根据气象观测数据研究大气层运动规律；大地测量学中根据大地测量数据推算近代应力场；地球物理学中根据地表的地球物理观测数据推算地下物质的结构、形态及成分等等，这些都是反演问题的例子。可见，相当多的科学技术问题，都以不同的方式、在不同程度上涉及反演问题。

解反演问题的古典方法是最小二乘法及统计学中的回归、参数估计等。近 40年来，由于在计算中广泛应用了信息论、线性及非线性规划、广义逆理论及最优化方法等一些数学工具，使得地球物理反演问题在理论和方法上都有了重大进展。

任何反演问题的计算都是建立在正演问题的基础之上。对于某个地球物理问题，只有得到了它的数学解答（解析的或数值的），即能计算其正演问题时，才有可能解反演问题。对于那些成因或机制尚不清楚的地球物理问题，是谈不上解反演问题的。所谓正演问题，是由理论模型参数计算观测数据的理论值。而反演问题，是由实际观测数据推算所有可能的模型参数。

反演问题是地球物理学研究中一个重要而又困难的问题，最主要的困难是反演结果存在着多解性（不唯一性）。这种不唯一性不是由于反演方法或技巧上的缺陷引起的，而是地球物理问题本身存在的固有困难。困难之一是观测资料不完备。由于地球物理观测只能在地球表面或其上空进行，无法直接获得反映地球内部物质的充足的信息。这就使得反演结果有多种可能性，而每一种可能性都能满足不充分的观测数据。地球物理观测信息的不足是无法用数学技巧来弥补的。困难之二是任何地球物理观测都存在干扰和误差。这种受畸变的观测数据使得反演计算不稳定，即观测数据中少量的错误可以导致反演结果的很大变动。因此，在解地球物理反演问题时，要根据已有资料，尽可能多地引入先验信息、附加约束条件，以减少解的不唯一性。

在不同的科学技术领域中，对反演问题有不同的提法。如，根据数学的观点，反演是在给定的子空间中寻找与函数空间中的某点最靠近的点；根据统计学的观点，反演是回归和参数估计问题；根据信息论的观点，反演又可看成滤波或过程鉴别问题。尽管术语和提法不同，但它们都表示同一内容，即从观测数据计算地球内部物质的各种参数。

需要全本请访问 www.ertongbook.com

下面对地球物理反演理论和方法的发展做一简要回顾和归纳，并对现有的（传统的）一般的反演方法做一定性的评价，其中不涉及为解决专门问题而设计的特殊反演方法。

1.2 地球物理反演理论和方法的发展

在地球物理学发展的初期，由于观测资料比较粗糙，一般只对观测数据做定性的或粗略的定量解释。从 20 世纪初到 50 年代，随着地球物理方法的发展，相继提出了各种解反演问题的定量计算方法。但这些方法大多是孤立地为解决某一特定问题而提出来的。例如，重、磁位场解释中常用到的特征点法、积分法就是为解释某一孤立位场异常而提出的一些简单反演计算方法。又如，地震学中根据地球表面体波走时观测来反演速度随深度变化的 Herglotz—Wiechert 公式，也是在假设速度是深度的单调函数时的一种积分运算。这些方法的特点是，仅能提供有关平均化的参量，如异常物体的质量或埋藏深度等。

近 40 年来，随着科学技术的迅速发展，地球物理资料无论在空间分布范围上（从大陆、大气层扩展到海洋和太空）还是数据的可靠性及丰富程度上，都较过去有很大改进。数学方法特别是电子计算机的广泛应用，促进了地球物理一般反演理论的建立和数据处理自动化的发展。

1967～1970 年间，Backus 和 Gilbert 发表了一系列重要文章。这些文章奠定了近代地球物理线性反演理论的基础，具有重要意义。这个比较完整、比较严格的反演理论被称之为 Backus-Gilbert（或 BG）理论。在他们的原始文献中，一般研究整个地球内部物理参数的连续变化，故反演问题以连续型函数表达。 1972 年，Wiggins 用矩阵将 Backus—Gilbert 理论表达为离散形式。同年， Jackson 用广义逆理论详细地讨论了线性反演问题在各种情况下的解答。另外，BG 理论中引入了分辨矩阵、信息矩阵和协方差矩阵等重要概念，这对于理解反演问题的实质及评价反演效果有重要作用。

1987 年，Tarantola 从概率的角度提出广义反演方法，该方法把表示先验信息的模型协方差矩阵和表示数据误差的数据协方差矩阵等先验信息直接引入到反演公式中，有效地减小了解的非唯一性（A. Tarantola，1987）。

同时，解反演问题的数值计算方法也得到了迅速发展，其中发展比较快的有广义逆理论和最优化方法。

1.2.1 广义逆理论

广义逆理论是矩阵论近年来的新成就之一。早在 1920 年，Moore 就提出了奇异矩阵的逆的问题，从而建立了广义逆矩阵的概念。 1935 年，他又对广义逆概念进行了重新讨论。 1955 年， Penrose 也独立提出了一个广义逆矩阵的概念。他的这个概念实际上与 Moore 的概念等价，现在称为 Morre-Penrose 广义逆，常用 A^{+} 表示。同年，Rao 提出了一个更一般的广义逆概念，通常称为 g 逆，用 A^{-} 表示。此后，广义逆理论得到了迅速发展，并广泛应用于许多科学技术领域。

1.2.2 最优化方法

最优化方法是一个有悠久历史的方法，至少可以追溯到 Newton 发明微积分的时代。另外，早在 1847 年，Cauchy 就提出了最速下降法，这也许就是最早的求解最优化问题的方法。但这些方法对参数较多的问题常常不适用。这是因为，最速下降法收敛得比较慢，而且，我

们经常遇到的问题都是非线性的。后来，计算机的使用和实际需要的增长，使这些古老的方法取得了实质性的发展。如 Daviden 提出的变度量法（即拟 Newton 法），是最优化计算方法中最好的方法之一。Broyden，Powell，Fletcher 等人在这方面也做了大量工作，他们提出的 powell 直接方法和共轭梯度法在最优化方法中也占有重要位置。 Marquardt 针对平方和这种特殊形式的目标函数也做了很好的工作，提出了阻尼最小二乘法。总之，出现了很多新的最优化计算方法，而且这些方法在解决实际地球物理反演问题中都得到了广泛的应用。

1.3 地球物理反演理论的一般数学原理

设：U 为地球物理参数值构成的集合，称为模型空间；E 为映射地球物理参数与地球物理观测场之间的算子；D 是地球物理观测值构成的集合，称为数据空间。

如果模型空间 U 和数据空间 D 都是线性空间，且映射算子 E 也是线性的，则这样的反演问题为线性反演问题。但是，大部分地球物理反演问题都不是线性的。为此，必须使用两种方法使其线性化。

(1)将映射算子 E 用线性算子 E_0 代替。常用的方法是将其在某一初始模型 W_0 附近展开，使线性函数 $E_0(W)$ 与非线性函数 $E(W)$ 充分地接近。

(2) 通过约束条件使模型空间 U 与数据空间 D 都成为线性空间。

假定模型空间 U 为希尔伯特空间 $L_2(\Omega)$，其中 Ω 为实空间中的有限域。通过 m 个映射算子的核 $G_i(r) \in L_2(\Omega)$，可以把模型空间 U 映射到数据空间 D，即

$$\int G_i(r)W(r)\mathrm{d}r = g_i(W) \quad (i=1,2,\cdots,m) \tag{1.1}$$

其中，$W(r)$ 为连续型地球模型参数（r 可看成是地球半径），$g_i(W)$ 等于或近似等于观测数据 d_i。这样表达的问题是连续型的。

若将 Ω 分成 n 个子域 Ω_1，Ω_2，…，Ω_n（例如将地球内部划分成 n 层），使 $\Sigma\Omega_i=\Omega$，则上述连续型函数就被离散化。模型参数 W 被离散为 W_k $(k=1,2,\cdots,n)$，积分方程 (1.1) 可写成矩阵方程形式

$$(G\times W)_i = \Sigma G_{ik}\times W_k = g_i(W) \quad (i=1,2,\cdots,m) \tag{1.2}$$

其中，G 为 $m\times n$ 矩阵，模型参数 W 为 n 维向量，观测数据 g 为 m 维向量。由 (1.2) 式表达的问题是离散的。

所谓线性反演问题，就是从 m 个观测数据 d_1，d_2，…，d_m 中获得所要求的模型参数 W（或 n 个离散值 W_1、W_2，…，W_n）的信息。Backus-Gilbert 理论是对连续型问题（1.1)给出平均解的一种途经，而广义逆理论和最优化方法是对离散型问题（1.2）求解的基本方法。

1.4 连续型线性反演理论

前面已经指出，地球物理反演问题实质上是寻找一个物理模型，它与观测数据拟合得最好。但这个物理模型往往是不唯一的。 Backus 和 Gilbert 的主要贡献在于，从全部可能的希尔伯特空间，特别是从那些满足观测数据的模型空间中减少了不唯一性，因为可以根据模型的性质对它进行约束。如模型的非负性质，反演计算模型与初始模型在模型空间中最短距离性

质等，都可用来对模型做出约束，使反演结果的不唯一性得到控制。

地球物理反演问题的一般提法是，根据一组观测值 d_j，$j=1, 2, \cdots, m$，找出适应于该观测数据的地球物理模型 W。一般来说，观测数据 d_j 是模型参数 W 的非线性函数：

$$d_j = E_j(W), \quad j = 1,2,\cdots,m \tag{1.3}$$

对于真实地球物理模型 W_0，其相应的观测值为

$$d_j^0 = E_j(W_0), \quad j = 1,2,\cdots,m \tag{1.4}$$

引入希尔伯特空间的目的之一，是使我们能定义并计算观测数据相对于模型参数的导数。考虑两个模型 W 和 W^1，它们之间仅相差一个微量 δm：

$$W^1 = W + \delta m \tag{1.5}$$

若它们的函数值之间有关系

$$E_j(W + \delta m) = E_j(W) + (F_j, \delta m) + \varepsilon(\delta m) \tag{1.6}$$

其中，每个 F_j 都是希尔伯特空间中的一个元素，$(F_j, \delta m)$ 为 F_j 与 δm 的内积，且 $\varepsilon(\delta m)$ 比范数 $\|\delta m\|$ 更快地趋于零。如（1.6）式成立，则称此函数是 Frechet 可微的，F_j 则称为观测数据 d_j 的 Frechet 核函数（或算子）。通常，F_j 是 d_j 对 W 的导数。所有有地球物理意义的数据都可以看成 Frechet 可微的。显然，(1.6) 式实质上是将问题线性化的一个重要步骤。

设（1.5）式中的 W 是初始模型，而 W^1 是真实模型，或至少是满足观测数据的模型。由于初始模型 W 是给定的，而 W^1 是待求的，所以我们希望得到 δm。但 δm 是无穷维向量空间的一个元素，而仅有 m 个观测数据，这就使得解是不唯一的。Backus 和 Gilbert 提出的一个约束条件是，寻找一个与初始模型 W 最接近的 W^1，或在希尔伯特空间中与 W 距离最短的 W^1 作为解。这种约束条件意味着寻找一个 δm 范数为最小，即 $\|\delta m\| = \min$ 的解，或满足：

$$\int[\delta m(r)]^2 \mathrm{d}r = \min \tag{1.7}$$

的解。

若 W 与 W^1 是最靠近的，则（1.6）式中的 ε 项可以忽略。用变分法可得：

$$\delta m = \sum_{j=1}^{n} a_j F_j \tag{1.8}$$

式中 a_j 可通过下面 n 个线性方程求得：

$$\Sigma(F_j, F_k)a_k = E_j(W^1) - E_j(W), \quad j = 1,2,\cdots,n \tag{1.9}$$

其中，(F_j, F_k) 是 Frechet 导数的内积。计算时忽略 ε 项，说明这种方法是近似的，即由此得到的 δm 是不精确的。但可以应用迭代法，每次迭代用上一次的结果（$W+\delta m$）为起点，逐步得到与实际观测值符合最好的模型。

1.5 离散型反演问题

任何一个地球物理反演问题，如根据地球重、磁异常计算矿体或构造的空间位置及形态，根据地震体波或面波研究地球结构，根据地热流或电磁场推算地球内部物质状态，根据人造卫星轨道计算地球外表形状及内部物质分布，根据地磁场资料推算海底扩张及大陆漂移史等，都可以用下述离散模型表示。

由实际观测得到的 m 个离散的地球物理数据，以 m 维向量表示

$$\boldsymbol{Y}^{o}=(y_1^{o},y_2^{o},\cdots,y_m^{o})^{T} \tag{1.10}$$

它称为观测值向量。

一个具体的地球物理问题，可以用 n 个离散的（具有相同或不同量纲的）物理参数来表示反演所要求的模型，写为 n 维向量

$$\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^{T} \tag{1.11}$$

它称为模型参数向量。

由初始的模型参数，根据具体的地球物理问题，可以计算出一组理论值，它可用 m 维向量表示

$$\boldsymbol{Y}^{c}(\boldsymbol{X})=(y_1^{c}(\boldsymbol{X}),y_2^{c}(\boldsymbol{X}),\cdots,y_m^{c}(\boldsymbol{X})) \tag{1.12}$$

它称为模型的理论值向量。

地球物理反演的关键是，用一定的数学方法寻找一个模型，使它尽量符合真实地球物理模型。即寻找参数向量 $\boldsymbol{X}$，使其计算的理论值向量 $\boldsymbol{Y}^{c}$ 与观测值向量 $\boldsymbol{Y}^{o}$ 尽可能符合。为此，可以利用最小二乘原理，构造一个 m 维残差向量：

$$\boldsymbol{\varepsilon}(\boldsymbol{X})=(\varepsilon_1(\boldsymbol{X}),\varepsilon_2(\boldsymbol{X}),\cdots,\varepsilon_m(\boldsymbol{X}))^{T} \tag{1.13}$$

其中，

$$\begin{aligned}\varepsilon_1(\boldsymbol{X})&=y_1^{o}-y_1^{c}(\boldsymbol{X})\\ \varepsilon_2(\boldsymbol{X})&=y_2^{o}-y_2^{c}(\boldsymbol{X})\\ &\cdots\cdots\cdots\\ \varepsilon_m(\boldsymbol{X})&=y_m^{o}-y_m^{c}(\boldsymbol{X})\end{aligned}$$

令残差向量范数平方为极小：

$$\|\boldsymbol{\varepsilon}(\boldsymbol{X})\|^{2}=\Sigma\varepsilon_i^{2}(\boldsymbol{X})=\min \tag{1.14}$$

经过一定的数学变换，可由上式解出 $\boldsymbol{X}$，从而使问题得解。

1.6 解离散型反演问题的数值计算方法

用于解离散型地球物理反演问题的数值计算方法常用的有广义逆方法和最优化方法。

1.6.1 广义逆方法

广义逆方法的基本步骤是把正问题线性化，由目标函数列出线性方程组，解这个线性方程组，得模型参数。具体方法是：由给定的模型参数，计算相应的理论值。若（1.12）式中的 Y^{c}（$\dot{X}$ 为线性函数，则其各分量可以写成

$$\begin{aligned}y_1^{c}(\boldsymbol{X})&=a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n\\ y_2^{c}(\boldsymbol{X})&=a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n\\ &\cdots\cdots\\ y_m^{c}(\boldsymbol{X})&=a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n\end{aligned} \tag{1.15}$$

其中 a_{ij}为常系数。上式也可以直接写成矩阵形式

$$\boldsymbol{Y}^{c}(\boldsymbol{X})=\boldsymbol{A}\times\boldsymbol{X} \tag{1.16}$$

其中 $\boldsymbol{Y}^{c}$ 为 m 维向量，$\boldsymbol{X}$ 为 n 维向量，$\boldsymbol{A}$ 为 $m\times n$ 矩阵（系数矩阵）。

令残差向量为观测值与理论值之差

$$\boldsymbol{\varepsilon}(\boldsymbol{X})=\boldsymbol{Y}^{o}-\boldsymbol{Y}^{c}(\boldsymbol{X})=\boldsymbol{Y}^{o}-\boldsymbol{A}\times\boldsymbol{X} \tag{1.17}$$

需要全本请访问 www.ertongbook.com

于是有关系

$$A \times X = Y^{o} - \varepsilon(X) \tag{1.18}$$

一般来说，地球物理问题所反映的函数关系都是非线性关系，所以要对正问题做线性化处理。在模型参数的某个初始值 X_0 点进行 Taylor 展开，略去二阶以上高次项，有

$$\begin{aligned} y_1^c(X) &= y_1^c(X_0) + \Sigma(y_1^c/x_j)_{x0}\Delta x_j \\ y_2^c(X) &= y_2^c(X_0) + \Sigma(y_{c2}/x_j)_{x0}\Delta x_j \\ &\cdots \quad \cdots \quad \cdots \\ y_m^c(X) &= y_m^c(X_0) + \Sigma(y_m^c/x_j)_{x0}\Delta x_j \end{aligned} \tag{1.19}$$

写成矩阵形式为

$$Y^c(X) = Y^c(X_0) + A \times \Delta X \tag{1.20}$$

其中 A 为微分系数矩阵（即 Jacobi 矩阵），而 ΔX 为模型参数 X 的改正量。

由（1.20）式，残差向量可以写成

$$\varepsilon(X) = Y^{o} - Y^{c}(X) = Y^{o} - Y^{c}(X_0) - A \times \Delta X \tag{1.21}$$

令向量 $b = Y^{o} - Y^{c}(X_0)$，表示观测值与初始模型理论计算值之差，则（1.21）式可以写成

$$A \times \Delta X = b - \varepsilon(X) \tag{1.22}$$

可见，(1.22) 式与 (1.18) 式在形式上是相似的，即一个非线性问题可以转换成线性问题进行计算。所以，在广义逆方法中一般只考虑线性方程组

$$A \times X = b \tag{1.23}$$

的解法。对于线性反演问题，可直接把 X 做为模型参数，b 做为观测值。对于非线性反演问题，X 做为模型参数的改正量 ΔX，b 做为观测值与理论值之差。在 (1.23) 式中，均略去了 (1.18) 和 (1.22) 式中的残差项 $\varepsilon(X)$，这说明这种计算是近似的。但用迭代法可以使这种近似计算逐步逼近真值。

从以上讨论可见，解上述反演问题实际上归结为求解线性方程组 (1.23) 的问题，也就是求系数矩阵 A 的逆矩阵问题。根据逆矩阵的定义，只有当系数矩阵 A 为满秩（非奇异）时，它的逆矩阵才有意义。这时，(1.23) 式有唯一解，并可以写成简单而又紧凑的形式

$$X = A^{-1}b \tag{1.24}$$

不过，非奇异矩阵只是矩阵的一种特殊形式，常见的是奇异矩阵或长方阵。对于这些一般的矩阵，显然不存在通常意义下的逆矩阵，并且这时 (1.23) 式也没有 (1.24) 表示的那种唯一解。

广义逆理论就是解决这种一般矩阵的求逆问题。对于任意的 $m \times n$ 矩阵，一般 $m \neq n$，且 A 可以有任意秩，在广义逆的概念下，它们存在着逆矩阵。并当 (1.23) 式有解时，其解也可以表示成类似于 (1.24) 式的紧凑形式，即

$$X = G \times b$$

其中，G 是 $m \times n$ 矩阵 A 的广义逆矩阵。这样，广义逆理论就把解一般方程组的理论概括统一起来。其中，常用的计算方法有正交分解法和奇异值分解法。

1.6.2 最优化方法

最优化方法是直接求目标函数 $\varphi(X)$ 的极小值。其基本思路是用数值方法产生一个模型参数的点列

$$x_0, x_1, \cdots, x_k, \cdots$$

使其满足

$$\varphi(x_0) > \varphi(x_1) > \cdots > \varphi(x_k) > \cdots$$

直到 $\varphi(x_n)$ 小于给定精度为止。

最优化方法的一般计算步骤为：

①给定实际观测数据与测点位置信息；

②选择进行解释计算用的地球物理模型；

③给定理论模型各个参数的初值；

④用当前模型参数正演计算模型体的理论观测值，并将它与实际观测值对比；

⑤评定理论观测值与实际观测值的符合程度，即它们之间差异的大小，判断是否需要修改理论模型的参数重新计算对比；

⑥如果不满足要求，自动修改理论模型的参数，使理论观测值与实际观测值之间的差异不断减小；

⑦如此反复迭代，直到满足要求为止；

⑧最终得到的理论模型的参数，作为实际模型体的解释结果。

设实际观测值用 ΔT_k 表示（k 为观测点号），$k=1, 2, \cdots, m$；m 为观测点数。模型体的理论观测值用 f_k 表示，它是观测点坐标和模型体参数的函数，即

$$f_k = f(x_k, y_k, z_k, x_1, x_2, \cdots, x_n) \tag{1.25}$$

其中，x_k，y_k，z_k 为观测点坐标；x_1，x_2，…，x_n 模型体位置、大小、产状和物性参数等参量；n 为参量的个数。

通常，实测异常与理论异常之间差异的大小，是以各测点上两者之间偏差的平方和表示，即

$$\varphi = \Sigma(\Delta T_k - f_k)^2 \tag{1.26}$$

(1.25) 式中，x_k，y_k，z_k 在计算迭代过程中是不变的，需要求的是地质体参数 $\boldsymbol{X} = (x_1, x_2, x_n)^{\mathrm{T}}$。所以，$\varphi$ 可视为 $\boldsymbol{X}$ 的函数，即 $\varphi = \varphi(\boldsymbol{X})$。在上述的基本步骤中，关键问题是如何实现自动修改模型体的参数，即如何计算出这些参数的改正量。就是说，如何求一组参数 $\boldsymbol{X}$，它使 $\varphi(\boldsymbol{X})$ 为极小。

设给定的模型体的初始参数为 $x_1^0, x_2^0, \cdots, x_n^0$，我们的目的在于求出一组关于模型体参数的改正量 $\delta_1, \delta_2, \cdots, \delta_n$，来修改给定的初始参数，即有

$$\begin{aligned} x_1 &= x_1^0 + \delta_1 \\ x_2 &= x_2^0 + \delta_2 \\ &\cdots \\ x_n &= x_n^0 + \delta_n \end{aligned} \tag{1.27}$$

写成向量的形式为：

$$\boldsymbol{X} = \boldsymbol{X}^{(0)} + \boldsymbol{\delta}$$

这样便求出关于模型体参数的一组新值。如果将这组新参数代入 (1.26) 式能使目标函数 φ 减小，这就是成功的计算。如果能够确定这组新参数在一定精度范围内使 φ 取极小，就取 $x_1, x_2, \cdots, x_n$ 做为模型体的最终解释结果。

就如何自动修改理论模型参数，最优化方法又分为只利用目标函数值的直接法和利用目标函数的一阶导数或二阶导数的间接法。直接法不用计算目标函数的导数，通过对目标函数

值的比较，确定最佳模型参数。常用的有单纯形法、共轭方向法和 Powell 法。间接法需要计算目标函数的一阶或二阶导数，由此构造线性方程组。在迭代过程中，通过解线性方程组，得到模型参数的改正量，从而逐步得到最佳模型参数。常用的有最速下降法、拟 Newton 法、Marquardt 法、共轭梯度法和广义最小二乘法。

1.7 现有反演计算方法的评价

根据以上回顾可以看到，现有反演计算方法存在如下问题：

(1) 有些属于线性方法。对于非线性的反演问题，首先要把正问题线性化，由此构造目标函数及线性方程组。对于实际的地球物理非线性问题而言，这是不精确的。

(2) 有些属于局部方法。因为它们利用的是局部信息，如目标函数的梯度等。这种性质使它们容易陷入局部极值。

(3) 基本上都涉及解线性方程组。这对于模型参数较多的反演问题，会出现方程组的阶数很高的情况。在这种情况下，方程组的条件数很大，使得解的误差增加。

(4) 局部方法的特点是，它们都强烈地依赖于初始模型参数的选取。但是，要对初始模型选择一个好的初值往往是很困难的，它要求对问题的解有较多的了解。然而，如果没有好的初值，往往容易陷入局部极值。

(5) 对于需要计算导数的方法，一般来说比不计算导数的方法效率要高一些。但如果目标函数不存在导数或导数不易计算时，这些方法不便使用。

(6) 对于不需要计算导数的方法，它们的效率往往很低，收敛速度很慢。

但是，尽管现有方法有上述局限性，但它们并不是没有用的。目前，这些方法仍然是在解地球物理反演问题中用得较多的方法。这些方法以及它们的各种组合，已成功地应用于地球物理学的很多领域，解决了很多实际地球物理问题。

然而，随着科学技术对解决非线性、多参数问题的需求，上述方法在某些方面已不能满足要求。因此，必须研究新的反演方法。

小 结

本章中简要回顾了地球物理反演理论和方法的发展，并对现有地球物理反演方法做了定性评价。广义逆理论和最优化方法是当前解地球物理反演问题时应用得最多的两大类方法，但它们也具有局限性。为满足科学技术中对解决非线性、多参数问题的需要，有必要研究新的非线性的反演方法。

2 遗传算法(GA)

2.1 引 言

在第 1 章中回顾了地球物理反演理论和方法的发展及现状，认识到，现有的反演计算方法存在着局限性。但是，这一领域仍然是一个发展着的、十分活跃的领域。随着科学技术的发展，各种学术思想、方法和技术在相互借鉴、相互结合、相互渗透。地球物理反演理论和方法也是这样。

例如，利用对人的大脑工作原理的研究成果，出现了人工神经网络（Artificial Neural Networks，ANN）方法，这种方法已广泛应用于模式识别和分类问题中（杨行峻等， 1992）。利用物理学中晶体生成的方法，出现了模拟退火法（Simulated Annealing，SA） 。这种方法可用于函数优化，曾解决过著名的“旅行商”问题（Traveling Salesman Problem，TSP）。利用自然界生物进化原理，出现了遗传算法（Genetic Algorithms，GA） 。这种方法在国外发展得比较快，已广泛应用于人工智能、工业自动控制、图像识别、通讯等方面。国内在自动控制、决策管理等方面也得到了应用。

Ga 作为一种优化方法，其本身就是一种反演方法，近几年来也应用于地球物理反演问题中（Stoffa，Sen，1991）。本章研究 GA 的基本原理和方法。作为一种比较健全的真正非线性反演方法，它与现有的反演方法在解决问题的思路方面存在着显著的差别。本章通过一个简单的例子来说明 GA 的工作过程，利用图式（Schema）的概念来说明其能力，并给出它的实现方法。

2.2 GA 简介

GA（David，1989）是以自然选择和自然遗传机制为基础的优化方法。它通过在模型群体中做有结构的、随机的信息交换，开发和利用群体中最适应环境的成分，以形成具有某些智能特点的搜索能力。在每一代遗传演化中，用上一代最适应环境的那些成分来构造新的模型参数，以形成下一代新的模型群体。

GA 在搜索过程中不用计算目标函数的导数，只需要计算目标函数值。所以，它具有直接反演方法的全部优点。同时它控制着一群模型。随着算法的执行，它有效地处理包含在群体中的有关目标函数形状的信息，这些信息在模型群体中迅速地交换，使每个模型都能从群体中获得有关的信息，这使得模型群体能迅速确定出接近最优的解。同时，它充分利用全局方面的信息，所以不致于陷入局部极值。由于 GA 综合了全局优化和局部优化方法两方面的优点，可以认为它是基于导数方法的效率和完全非线性彻底搜索方法的健全性之间的结合。

GA 是由美国密西根大学的 John Holland 教授和他的同事们发展起来的。他们研究 GA 已有 40 余年的历史，主要研究目标有两个：①提取并精确地解释自然系统的适应过程；②设

计具有自然系统机制的人工系统软件。GA 研究的中心主题是健全性（robustness），即在不同的环境中生存时系统所必须具备的能力。人们常常感到，凡是要求健全性的地方，生物系统是做得最好的。人工系统——软件和硬件（无论工程系统、计算机系统或商业系统）——的设计者只能对生物系统的健全性、效率及灵活性感到望尘莫及，而自我完善、自我引导和再生能力是生物系统中的普遍特征。因此，探讨适应性和生存奥秘的最好方式，是仔细研究生物中的例子。理论和实践都证明，GA 能够在复杂空间中做健全的搜索。专门论述这方面问题的重要专著是 Holland（1975）的《自然系统和人工系统中的适应性》（Adaptation in Natural and Artificial System），这本书以及其它一些有关的论文及专著奠定了 GA 在函数优化及控制应用方面的理论基础。作为一种有效的优化方法，GA 目前已在商业界、科学界及工程界得到了更为广泛的应用。它之所以得到广泛应用，其原因很简单，即这种方法计算上简单，但搜索能力强大。另外，它对搜索空间没有苛刻的假设（如连续性、存在导数、单峰等等）。

2.3 GA 与现有方法的差别

GA 在健全性方面超过现有方法，是因为它在一些非常基本的方面与传统方法有所不同。即：

GA 对模型参数的编码进行操作，而不是对参数本身进行操作；

GA 同时用一个模型群体对模型空间进行搜索，而不是只用一个模型点；

GA 只利用目标函数值的信息，而不使用目标函数的导数或其它特殊的辅助信息；

GA 使用随机的（概率的）转移规则，而不是使用确定性的转移规则。

下面具体讨论：

(1)GA 要求把所要优化的问题的模型参数用一些有限的符号编成有限长度的数字串。如求某个函数 $f(x)$ 在某个区间内的极小值，现有方法是直接用参数 X 计算它在该区间内的每个函数值，直到得到最小函数值。而用 GA，第一步是把参数 X 转换成一个有限长度的数值编码。

可以用很多方法对参数编码，现以图 2.1 中的开关设置问题为例介绍一种简单的编码方法。这个开关设置问题是，某设备有 5 个输入开关，对应这 5 个开关的每一种设置，有一个输出信号 $f=f(s)$。其中，s 是 5 个开关的某个特定设置。该问题的目的是设置这些开关，以得到 f 的最大可能输出值。用一个 5 位的数字串来表示这 5 个开关，如果开关是“ON”，用 1 表示；如果开关是“OFF”，用 0 表示，这样可以得到一种简单的参数编码。对于这种简单的编码，数字串 11110 表示：前 4 个开关是“ON”，第 5 个开关是“OFF”。后面介绍的编码方案不会是这样明显，这里主要是强调 GA 是对参数的编码进行操作，而不是直接对参数本身。

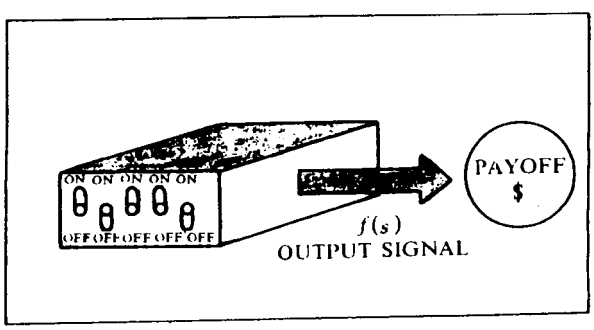


图 2.1 开关设置问题示意图

(2) 在很多其它优化方法中，对参数空间的搜索是根据某种确定性的转移规则在参数空间中从一点移动到下一点。这

种点到点的搜索方法很不安全，因为在多峰的搜索空间中它容易陷入局部极值。相反， GA 同时对一群点（模型）进行操作，并行地对很多极值做搜索，因此，它陷入局部极值的概率比点到点的方法小得多。

对于上述开关设置问题，用其它方法，通常是以某种开关设置开始，然后利用某种转移规则产生新的设置。而 GA 是从随机选择的某个群体（表示这些开关的不同初始设置）出发，相继产生下一代的群体。例如，对这个开关设置问题，用连续投掷硬币 20次的方法（正面＝1，反面＝0）可得到有 4 个成员的随机初始群体：

成员 1：0 1 1 0 1
成员 2：1 1 0 0 0
成员 3：0 1 0 0 0
成员 4：1 0 0 1 1

随后，GA 产生下一代群体。用适应多种情况的群体而不是用单个点对参数空间做搜索，这种做法体现出群体中有安全感 （There is safety in numbers）。后面将会看到，这种用多个模型并行搜索的方法怎样提高了 GA 的健全性。

（3）为了顺利地搜索，现有搜索方法要求有很多辅助信息。例如，梯度方法要求有导数存在（解析的或数值的），其它局部搜索方法要求目标函数连续等。相反， GA 不需要所有上述辅助信息。为了执行有效的搜索， GA 只需要与每个数字串有关的目标函数值，而这种信息是绝大多数传统搜索方法都要用到的最一般的信息。这个特点使得 GA 比其它搜索方法更具普遍性。因为，不同的搜索问题有截然不同的辅助信息，只有那些不用特殊的辅助信息的方法，才是适用于最普遍情况的搜索方法。

（4）与很多其它搜索方法不同， GA 用概率的转移规则来引导搜索。但是，概率的使用并不意味着 GA 是简单的随机搜索。它是用随机选择作为工具，引导搜索过程向着模型空间中最有可能使搜索过程得到改善的区域移动。

总之，这 4个差别——用参数的编码、用群体搜索、不用特殊的辅助信息和非确定性的转移规则——提高了 GA 的健全性，使它优于现有反演方法。

2.4 GA 运算

GA 运算的原理相当简单，只涉及参数编码的复制和部分编码的交换。至于这种简单的运算为什么会导致成功的搜索，这里将给出有说服力的解释。简单的运算和较高的效率是 GA 的两个主要特征。

简单 GA 通常由三步运算组成：

1 选择 （Selection）和再生 （Reproduction）

选择和再生是根据群体成员的目标函数值 f（生物学家称这个函数为适应度函数）对其编码进行复制的过程。直观上，可以把 f 想象为要求其极大值的某种利润或效率等等。按照成员的适应度函数值对其编码做复制，意味着具有高适应度函数值的成员对后代有较高的贡献概率。显然，这步运算是达尔文“适者生存”自然选择规律的体现。在自然群体中，适应度由生物克服天敌、瘟疫以及克服其它一些妨碍其生长及再生障碍的生存能力所决定。而在人工系统中，目标函数值是决定群体成员生存与淘汰的最关键因素。

需要全本请访问 www.ertongbook.com

再生运算可以用很多方法来执行，最容易的方法是建立一个有偏轮盘赌。当前群体中的每一个成员的编码在该轮盘赌中占有一段区间，其大小与该成员的适应度函数值成正比。假定上述开关设置问题中的 4 个成员的适应度函数值等数据如表 2.1 所示，与这一代相应的有偏轮盘赌如图 2.2 所示。为了执行再生操作，只需简单地转动这个轮盘赌 4 次，就可以决定哪个群体成员获得再生。例如，对于上述给定问题，第 1 个成员的适应度函数值为 169，它在所有适应度函数值中占 14.4%，因此它在轮盘赌中占 14.4%的区间。每次转动轮盘赌，第 1 个成员有 14.4%的选中概率。用这种方法，适应度函数值高的成员被选中的机会多，可以在下一代中有较多的后代。一旦某个成员被选中，则复制它的编码，为产生下一代新的群体做准备。

表 2.1 开关问题的群体成员及其适应度函数值

成员编号	编码	适应度函数值	百分比/%
①	01101	169	14.4
②	11000	576	49.2
③	01000	64	5.5
④	10011	361	30.9

2）基因交换(Crossover)

再生运算之后，可用两个步骤做基因交换运算。第 1 步，对选中的群体成员做随机匹配。第 2 步，每对成员的编码按下述方法做交换基因；在编码中随机选一交换位置，该位置之前的编码保持不变，之后的编码相互交换，由此产生两个后代。例如，对于某一对成员 A_1 和 A_2：

$$A_1 = 0\ 1\ 1\ 0|1$$
$$A_2 = 1\ 1\ 0\ 0|0$$

交换点

随机选择的交换位置如竖线所示，则基因交换所产生的下一代两个新成员为 A'_1 和 A'_2：

$$A'_1 = 0\ 1\ 1\ 0\ 0$$
$$A'_2 = 1\ 1\ 0\ 0\ 1$$

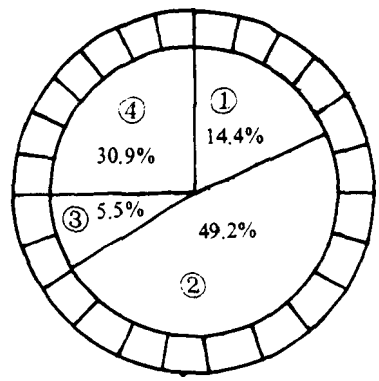


图 2.2 开关设置问题的有偏轮盘赌

再生和基因交换的原理非常简单，只有随机数的产生、编码的复制和部分编码的交换。然而，这种再生和基因交换相结合的随机过程产生了结构式的信息交换的效果，它赋予 GA 以很强的搜索能力。这是因为，在一个有 n 个成员的群体（如上述开关设置问题中的 4 个数字串的群体）中，每个成员不仅表示对执行一个具体的任务所必须的完整的概念（在这个例子中，每个数字串代表一个完整的开关设置概念），而且还包含着与该任务有关的各种因素。因此群体中包含的不仅是 n 个概念，而且包含着大量的为执行该任务所必须的其它因素以及这些因素的组合。GA 以如下两种方式充分地利用这些丰富的信息：①根据群体成员的适应度函数值使高质量的因素得以再生，②把这些因素与其它成员中高质量的因素做信息交换。因此，再生和基因交换运算就可以根据过去试验中所得到的好的结果，来构造出更好的结果。这样，

两个简单的运算会产生有效的、快速的而且是健全性的搜索效果就不难理解了。

这种方式与人类根据以往的有效的方法产生新的更有效的方法有相同的地方。根据人类的经验，新的方法是过去比较有效的方法的交流与结合。GA 中的再生和基因交换也是这种方法的交流与结合，好的方法在搜索过程中被反复地利用和交换，最后得到新的更好的方法。

3）基因变异(Mutation)

如果再生和基因交换给了 GA 以强大的信息处理能力，那么基因变异的目的是在搜索过程中不断地引入新的信息量。“变异是需要的，因为尽管再生和基因交换可以有效地进行搜索，但偶尔它们也会丧失某些潜在有用的遗传物质”(Goldberg，1989)。在人工遗传系统中，变异运算可以防止这种不可恢复的有用遗传物质的损失。在上述简单 GA 中，变异是以很小的概率随机地改变编码中某个位置上的值，即只是简单地把 1 变成 0，或把 0 变成 1。就其本身来讲，变异是在编码空间中随机执行。但当把它与再生和基因交换同时使用时，可以防止某些重要因素的过早丢失。

上述三种运算在 GA 中起着不同的作用：

(1) 选择和再生体现适者生存的原理。在地球物理反演过程中，相当于根据目标函数衡量模型参数接近其真值的程度。

(2) 基因交换控制着模型之间的混合程度及分享信息的程度。在地球物理反演过程中，这是对参数空间进行全局搜索的一种方式。

(3) 基因变异的目的是在群体中保持一定程度的差异及随机性，这种差异和随机性会由于前面的两个步骤而逐渐减弱。由于基因变异的存在，会不时地引入新的遗传物质，这有利于利用包含在群体中的信息对模型空间中新的区域做进一步的开发。它在地球物理反演过程中是对参数空间做局部搜索的一种方式。

2.5 GA 的实现方法

1）确定目标函数

有两种方法：

①以计算值与观测值拟合程度最高为标准，即求模型参数 m，使目标函数 $F(m)$为最大：

$$F(m)=\frac{f_0+f_c}{F_0^2+f_c^2} \tag{2.1}$$

其中 ：+为相关运算。

②以计算值与观测值之差的平方和最小为标准，即求模型参数 m，使目标函数 $F(m)$

$$F(m)=\left(\frac{\Sigma(f_0-f_c)^2}{n}\right)^{1/2} \tag{2.2}$$

为最小。其中： n 为观测值的个数。

2）指定搜索范围

给定每个模型参数 m 的搜索范围，并按精度要求分段，其分段数为 2 的幂次方：

模型参数 m_i 的变化范围：$(X_{\min}, X_{\max})$

分段间隔数：2^k

每段长度（即计算精度）：$\Delta=(X_{\min}-X_{\max})/2^k$

3）模型参数编码

从上述讨论可知，GA 运算是对参数的编码进行操作。但当正演时，又要用到参数的十进制数。所以，要考虑各种数制的相互转换问题。

对于十进制与二进制的相互转换问题，我们已比较熟悉， 不再赘述。但 GA 中有时用到一种 Gray 码，它与二进制的相互转换关系为：

a 二进制转换到 Gray 码：

$$G_i = \begin{cases} B_i, & i = 1 \\ B_{i-1} + B_i & i > 1 \end{cases} \tag{2.3}$$

其中：+为二进制加法。

b）Gray 码转换到二进制：

$$b_i = \sum_{j=1}^{i} G_j \tag{2.4}$$

其中的求和运算用二进制加法。

c）术语：

参数编码中的每一位称为基因，模型参数可以由若干位基因组成。所有模型参数的编码串连起来组成一个数字串，称为一条染色体，代表一个模型。尽管基因和染色体与模型参数和模型是不同学科中的术语，但在 GA 中它们表示同一种事物，所以，在以后的讨论中，有时是把它们混在一起讨论的。

d）三种参数编码方案：

GA 同时控制一组模型（Q 个），称为模型群体。初始模型群体由随机方式产生，然后把它们转换成编码。如两个 3 参数的模型：

（18，28，6）和（16，30，3）

它们可由如下方法进行编码：

• 规则编码：即通常的二进制编码

（18，28，6）→（10010 11100 00110）

（16，30，3）→（10000 11110 00011）

• 量化编码：把上面每个参数二进制编码的第一位放在一起，第二位放在一起等等，这样组成一种新的编码形式：

（18，28，6）→（110 010 011 101 000）

（16，30，3）→（110 010 010 011 001）

量化编码的意义是，前 3 个元素（110）表示，前两个参数（18，28）的第 5 位大于 24，后一个参数（6）的第 5 位小于 24，等等。

• Gray 编码：二进制转换成 Gray 码：公式见（2.3）式。

如上述参数的二进制编码转换成 Gray 码为：

11011 10010 00101

11000 10001 00010

Gray 码转换成二进制：公式见（2.4）式。

如上述参数的 Gray 码转换成二进制为：

10010 11100 00110

10000 11110 00011

Gray 编码的性质是，任何两个相邻的十进制数用 Gray 码表示时，它们的编码中只有 1 位数字不同。若用二进制表示，只在局部范围才有这种性质。

二进制编码与 Gray 编码的比较

十进制	0	1	2	3	4	5	6	7	8	9
二进制	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001
Gray 码	0000	0001	0011	0010	0110	0111	0101	0100	1100	1101

4）随机产生初始群体

a）在上述给定的模型参数搜索范围及分段的基础上，由一随机数 a（$0 \leqslant a \leqslant 1$）与某模型参数的总分段数相乘，其结果再与该参数的最小值相加，即得该模型参数的随机选择值。因为分段是均匀的，所以在指定范围内，该模型参数的各个可能值有相等的选中机会。

用公式表示：

$$a = \text{random}\quad ()$$

$$k = a \times 2^k$$

$$m_j = X_{\min} + k \times \Delta$$

b）把这个随机选择的模型参数按指定的编码方案进行编码。

c）对每个模型参数重复步骤 1 和 2，得到各个模型参数的编码。把它们串连起来形成一个染色体，代表一个模型。

d）重复步骤 1）～3）到指定次数（如 Q 次），得到 Q 个模型，做为初始群体，为下面的 GA 运算做准备。

上述过程的示意图见图 2.3。

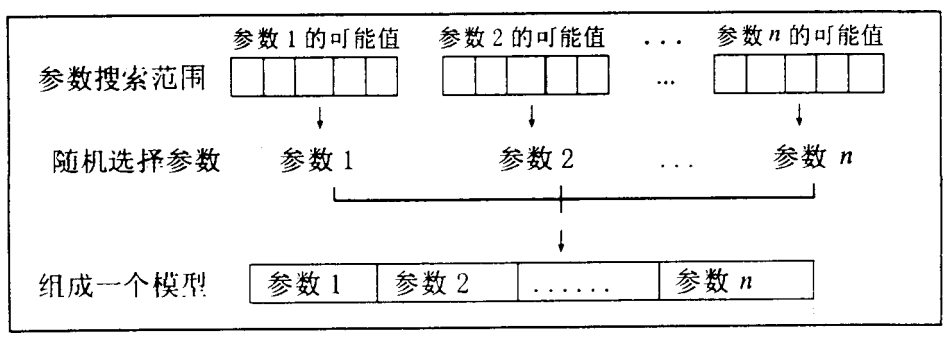


图 2.3 模型产生示意图

5）计算目标函数

$$F_j(m) = (\Sigma(f^c - f^o)^2/n)^{1/2} \qquad j = 1, Q$$

其中，f^c 和 f^o 分别为模型参数的理论计算值和实际观测值，n 为观测点数，Q 为群体大小。

在上述群体中，求出所有目标函数 F_j 的最小、最大、平均值和总和，并根据目标函数的最小值保存最好的模型。

6）遗传运算

a）目标函数的转换：

用途：把目标函数转换成适应度函数 Fitness(i)，为模型的再生选择做准备。对于求目标函数极小的问题，常用(2.5)式或(2.6)式。对于求目标函数极大的问题，可选用（2.7)式。

线性转换：

$$\text{Fitness}(m_k) = \frac{\Phi_{\max} - \Phi(m_k)}{Q \times (\Phi_{\max} - \Phi_{\text{mean}})} \tag{2.5}$$

指数转换：

$$\text{Fitness}(m_k) = \frac{\exp[-\Phi(m_k)/\sigma]}{\Sigma\exp[-\Phi_j/\sigma]} \tag{2.6}$$

其中，σ 是所有目标函数的标准偏差。可见， 目标函数越小，其适应度函数值越大。

不转换

$$\text{Fitness}(i) = F(i) \tag{2.7}$$

b）选择和再生：

适应度函数归一化

$$\text{Fitness}(i) = \text{Fitness}(i)/\text{sum} \qquad i = 1,2,\cdots,Q$$

其中，sum 为 Q 个适应度函数值的总和。

在归一化的基础上确定各模型（在轮盘赌中）所处的范围：

$$S(i) = \sum_{j=1}^{i} \text{Fitness}(j) \tag{2.8}$$

由随机数 a （$0 \leqslant a \leqslant 1$）决定各模型的再生机会：

若

$$S(i) \leqslant a \leqslant S(j) \tag{2.9}$$

则第 j 个模型获得再生的机会。

期望的后代数

$$\text{Offspring} = \text{Fitness}(i) \times Q \tag{2.10}$$

上述方法按各模型的适应度函数值决定其再生的机会和在下一代中的后代数，体现优胜劣汰、适者生存的进化原理。

c 基因交换：

在上述获得再生的群体中，用随机选择的方法使群体中的染色体两两配对，组成 $Q/2$ 对双亲，并按概率 P_c 进行基因交换。交换方法为，对于某对双亲，如果随机数大于 P_c，则不做基因交换，直接进入下一代。如随机数小于 P_c，则在表示该对双亲的数字串中随机地选一交换点。交换点之前的基因保持不变，交换点之后的基因相互交换，由此产生后代。

d 基因变异：

按概率 P_m 对群体中的所有基因做随机改变，即由 0 变为 1，或由 1 变为 0。对于某个基因，如随机数大于 P_m，则不做基因变异。如随机数小于 P_m，则按上述方法对该基因做变异运算。

2.6 GA 的模拟计算

现在，把上述 GA 运算一步一步地应用到一个实际的优化问题中。考察求函数

$$f(x)=1-\mathrm{e}^{-0.001|x|}\cos(|x|^{1/2})$$

的极值问题，其中参数的编码采用二进制编码。由于正数的二进制编码较为简单，所以限制 x 在 [0，500] 之间变化，见图 2.4。

GA 参数：①参数编码：采用二进制编码；

②目标函数转换方式：指数形式；

③群体大小：$Q=4$；

④交换概率：$P_c=0.9$；

⑤变异概率：$P_m=0.07$。

另外，参数空间大小：$n=2^9=512$，搜索精度为 $\Delta=1$，随机数种子：seed=90。

对于这样一个函数，它在区间 [0，500] 之间有 5 个局部极值，其中全局极值为 $y=f(0)=0$。可以想象，如果用现有搜索方法，除非对这个函数有比较多的先验信息，可以给出较好的初值，否则，对于任何一种搜索方法，这都将是一个困难的问题。

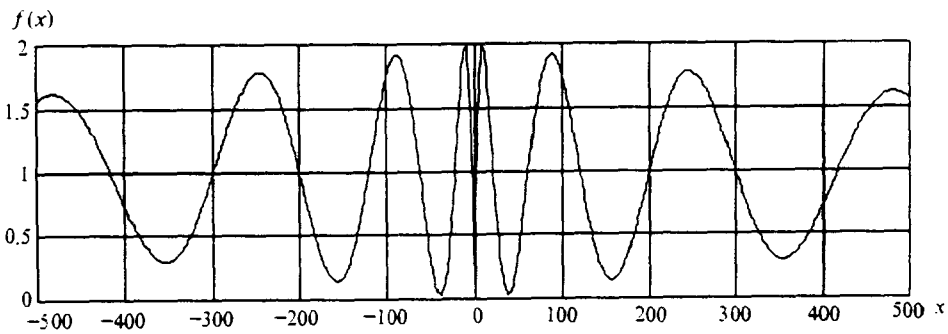


图 2.4　函数 $f(x)$的曲线示意图

现在，用 GA 来解这个问题。为了应用 GA，首先必须把问题中的参数 x 转换成二进制编码。由于 x 的最大值为 500，可以用 9 位二进制数来表示。利用已定义的适应度函数和编码方案，现在可以用再生、基因交换和基因变异运算来模拟 GA 的一代遗传演化。

首先用随机数发生器选择一个有四个成员的随机初始群体，见表 2.2 的左边，其中也有群体成员(x)的编码、相应的目标函数值 $f(x)$和适应度函数值 $F(x)$。

GA 的一代遗传演化由选择和再生开始。旋转轮盘赌 4 次，选出作为下一代父本的成员。用产生随机数的方法模拟这个过程，结果成员 3 和成员 4 各被复制 1 次，成员 1 被复制 2 次，成员 2 没有被复制，被淘汰（表 2.2）。

表 2.2　参数编码及再生过程

编号	群体成员	参数编码	$f(x)$	$F(x)$	生存概率区间	期望后代数	实际个数
1	164	010100100	0.176	0.555	0.000～0.555	2.222	2
2	313	100111001	0.706	0.109	0.555～0.663	0.436	0
3	139	010001011	0.379	0.297	0.663～0.960	1.188	1
4	419	110100011	1.032	0.040	0.960～1.000	0.160	1

与希望的复制数($QF(x)$)比较，基本得到了期望的结果：最好的成员 1 被两次复制，处于

需要全本请访问 www.ertongbook.com