

数字函数发生最佳法

路浩如
(浙江大学)

摘 要

本文提出了一种使数字函数发生过程最佳化的算法,称为最佳法。这种算法引起的误差在半步以内(精确性);曲线发生途径与起始点无关(唯一性);反方向发生曲线时,曲线路径不变(可逆性);同一算法可适用于二维曲线和三维曲线(统一性);算法简单可行(可行性)。最佳法是以跟踪邻近候选格点中函数绝对值最小的格点为准则追踪给定函数。按照函数差分值的大小排定第一坐标和第二坐标。文章对初始大方向的确定及函数发生过程中大方向的更换提出了简单易行的算法。

一、前 言

数字函数发生方法在世界各国特别是在欧美各国多采用数字微分分析法。我国和日本则多应用代数法。近年来出现一些引人注目的新方法,如日本安川电机所发表的“齐田法”、我国浙江大学独创的“正负法”等。

数字微分分析法由于采用“微分算法”进行积分,因而在增量有离散特点的数字函数发生中不可避免地形成积累误差。分析证明这种积累误差可以达到十分可观乃至无法允许的程度。因此在采用数字微分分析法进行函数发生时,如果不采取补偿措施,就只能适用于一些积累误差不太显著的情况。

1974年发表的齐田法其基本核心同样是“微分算法”,因而可以认为是微分分析法的一种重要的变型。关于微分分析法的误差分析完全适用于齐田法。

代数法是继微分分析法之后,在六十年代兴起的一种函数发生方法。代数法采用“差分算法”以避免积累误差,使误差限制在一步之内。

代数法大都局限于研究直线和圆弧的发生,而且对不同的函数没有统一的算法。浙江大学研究的“正负法”则在这两个方面上发展了代数法的基本思想,其基本特点和代数法一样,基本误差也限制在一步之内。这里所说的基本误差不包括正负法所采用的曲线变向算法引起的误差。后者在斜曲线(函数中包含有 x 与 y 相乘项的情况)的特定场合是比较显著的。

根据以上概述,仅从误差这一点来说,差分算法是消除积累误差的有效途径。但代数法和正负法存在以下缺点:

1) 不是以误差确定走步方向. 它判定的方向常是扩大误差的方向. 例如, 当前点正好是函数上一点时, 走步方向并不是沿着曲线的切线方向, 而是沿着法线方向. 这种步法的误差虽不超过一步, 但却形成了不必要的偏离.

2) 函数发生途径不是唯一的. 它与起点及函数发生方向有关, 因而是不可逆、不对称的.

3) x 坐标与 y 坐标分别单独走步, 影响函数发生的精度和速度.

4) 不能用于三维函数的发生, 只能在不同的 z 值下重复发生二维函数.

二、最佳化的要求

数字函数发生最佳化的基本要求是:

(1) 数字函数发生的最佳化算法应能使走步误差达到可以实现的最低极限 (精确性).

(2) 函数发生的途径与起始点的位置无关, 即从不同起始点出发发生的曲线应该是完全重合的(唯一性).

(3) 反方向发生曲线时, 曲线的路径不变, 即从不同方向发生曲线应该完全重合(可逆性).

(4) 同一算法能适用于二维曲线及三维曲线的发生(统一性).

(5) 算法必须简单切实可行(可行性).

在满足以上诸项要求时, 发生一对称的函数必有对称的步法(对称法).

实现以上最佳化要求的关键在第一项. 事实上在离散格点坐标系统上出现一个格点的充分必要距离是一步. 因此走步误差的最低极限应在 $1/2$ 步范围之内. 图 1 和图 2 分别表示在允许走步误差为一步和 $1/2$ 步时曲线发生可能选取格点的范围. 显然走步误差为一步的数字函数发生方法无法实现最佳化. 而把走步误差限制在 $1/2$ 步之内时, 曲线发生的途径将只有一种方案, 它是唯一的也是最佳的方案.

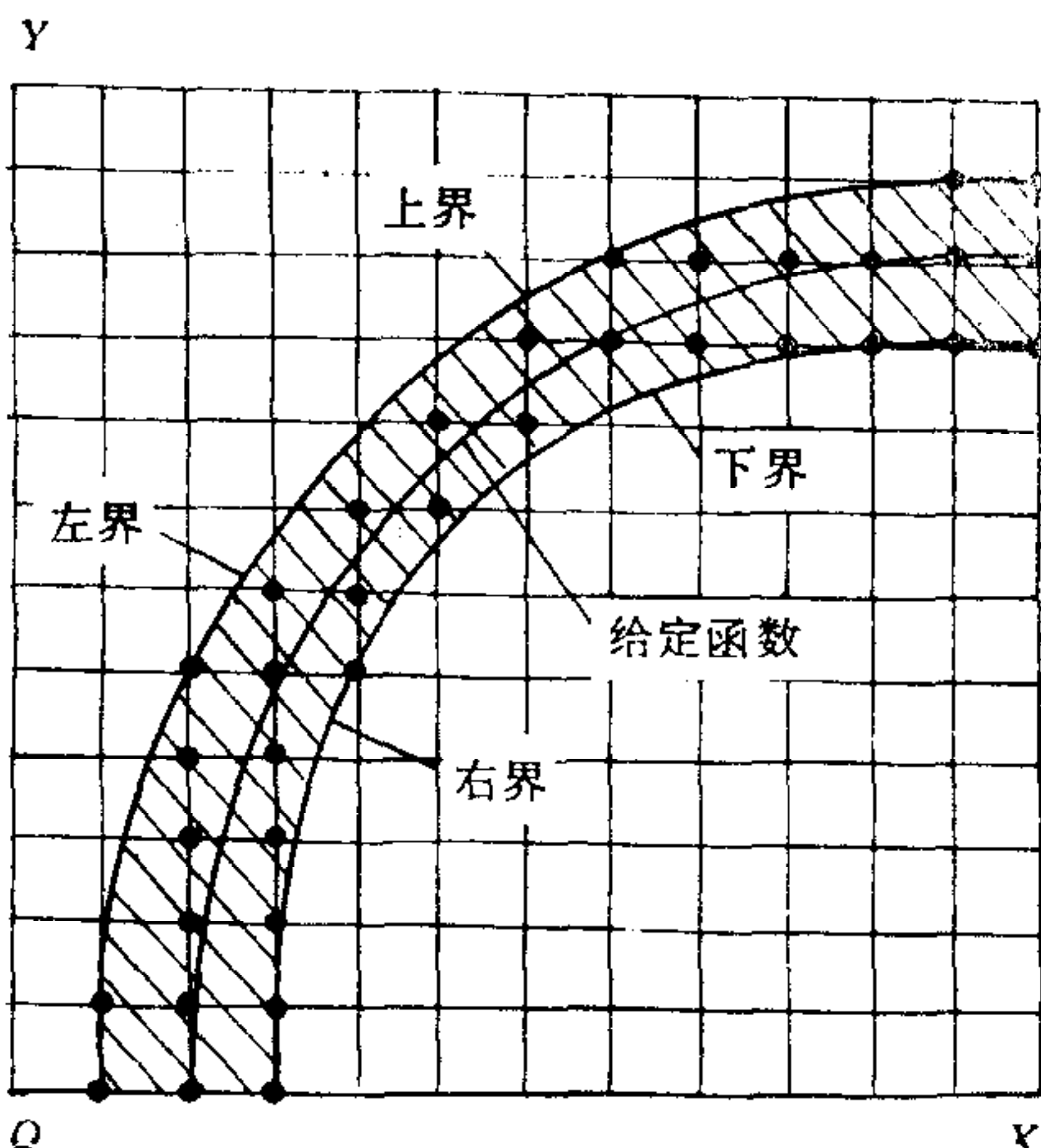


图 1

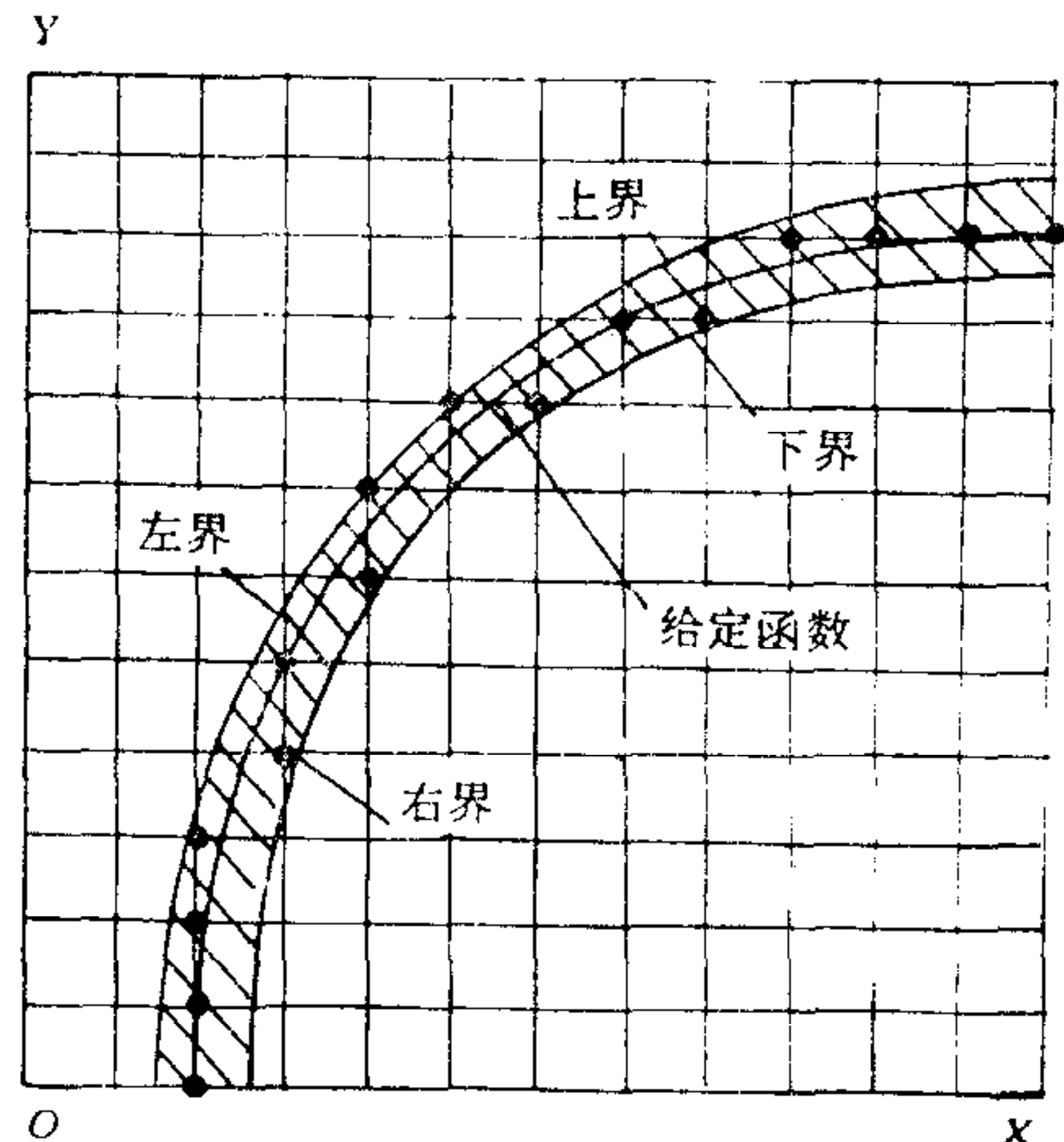


图 2

本文的目的就在于发展一种只需简单运算而使函数发生最佳化的算法原理，命名为“最佳法”。

三、函数的差分

对任意二维函数 $f(x, y)$ ，在 x, y 平面上的任意一点都可以计算其函数值。因此函数 $f(x, y) = 0$ 在该平面上所表示的曲线是平面上函数值为零的所有点的集合。由于数字函数发生中只能以离散格点的集合来逼近给定函数，如果这一点集始终保持每一点上的函数值最近似零，那么该点集所形成的曲线必将是函数的最佳近似，从而函数发生的过程就成为追踪 $|f(x, y)| = \min$ 各离散格点的过程。它是在给定步长下函数的最佳发生过程。

在函数发生过程中，当前点的函数值是已知的，函数值的计算就归结为从当前点向某一坐标或两个坐标方向前进进一步的函数值计算。由此，采用差分算法是合宜的。

为了便于应用，根据前进方向是在坐标的正方向还是负方向，把它们区别为“向前”和“向后”。于是有相应的“向前差分”和“向后差分”。设当前点的 x 坐标为 x_i ，向前一步有 x_{i+1} ，向后一步则有 x_{i-1} 。定义各次差分为：

$$\left. \begin{array}{l} \text{向前一次差分} \quad \Delta_{x_i} = f_{x_{i+1}} - f_{x_i} \\ \text{向前二次差分} \quad \Delta_{x_i}^{(2)} = \Delta_{x_{i+1}} - \Delta_{x_i} \\ \text{向前三次差分} \quad \Delta_{x_i}^{(3)} = \Delta_{x_{i+1}}^{(2)} - \Delta_{x_i}^{(2)} \\ \dots\dots\dots \end{array} \right\}, \quad (1)$$

$$\left. \begin{array}{l} \text{向后一次差分} \quad \nabla_{x_i} = f_{x_{i-1}} - f_{x_i} \\ \text{向后二次差分} \quad \nabla_{x_i}^{(2)} = \nabla_{x_{i-1}} - \nabla_{x_i} \\ \text{向后三次差分} \quad \nabla_{x_i}^{(3)} = \nabla_{x_{i-1}}^{(2)} - \nabla_{x_i}^{(2)} \\ \dots\dots\dots \end{array} \right\}. \quad (2)$$

向前差分和向后差分之间存在着以下转换关系：

$$\nabla_{x_i}^{(n)} = (-1)^n \Delta_{x_{i-n}}^{(n)}, \quad (n \geq 1, \text{ 整正数}). \quad (3)$$

$$\Delta_{x_i}^{(n)} = (-1)^n \nabla_{x_{i+n}}^{(n)}, \quad (n \geq 1, \text{ 整正数}). \quad (4)$$

如果考虑对于同一点的向前差分和向后差分，则可从(3),(4)式得到以下转换式：

$$\nabla_{x_i}^{(n)} = \sum_1^k \left[(-1)^k \frac{(k-1)!}{(k-n)!(n-1)!} \Delta_{x_i}^{(n)} \right], \quad (5)$$

$$\Delta_{x_i}^{(n)} = \sum_1^k \left[(-1)^k \frac{(k-1)!}{(k-n)!(n-1)!} \nabla_{x_i}^{(n)} \right]. \quad (6)$$

从当前点的函数值及各次差分容易得到走步时的函数值及各次差分。例如从 x_i 进至 x_{i+1} 时有：

$$\left. \begin{array}{l} f_{x_{i+1}} = f_{x_i} + \Delta_{x_i} \\ \Delta_{x_{i+1}} = \Delta_{x_i} + \Delta_{x_i}^{(2)} \\ \Delta_{x_{i+1}}^{(2)} = \Delta_{x_i}^{(2)} + \Delta_{x_i}^{(3)} \end{array} \right\} \quad (7)$$

$$\left. \begin{array}{l} \dots\dots\dots \\ \nabla_{x_{i+1}} = \nabla_{x_i} - \nabla_{x_{i+1}}^{(2)} \\ \dots\dots\dots \end{array} \right\}$$

对于给定函数 $f(x, y) = 0$ ，它的各次差分可以根据泰勒公式得到。向前差分时，取坐标增量 $\Delta x = 1$ ；向后差分时，取 $\Delta x = -1$ ，于是有：

$$\left. \begin{array}{l} \Delta_x = f'_x + \frac{1}{2} f''_x + \frac{1}{6} f'''_x + \dots \\ \Delta_x^{(2)} = \Delta'_x + \frac{1}{2} \Delta''_x + \frac{1}{6} \Delta'''_x + \dots \\ \Delta_x^{(3)} = \Delta_x^{(2)'} + \frac{1}{2} \Delta_x^{(2)''} + \frac{1}{6} \Delta_x^{(2)'''} + \dots \\ \dots\dots\dots \\ \nabla_x = -f'_x + \frac{1}{2} f''_x - \frac{1}{6} f'''_x + \dots \\ \nabla_x^{(2)} = -\nabla'_x + \frac{1}{2} \nabla''_x - \frac{1}{6} \nabla'''_x + \dots \\ \dots\dots\dots \end{array} \right\} \quad (8)$$

对 n 阶多项式函数，其 n 次差分成为常数，更高次的差分为零。在假定起始坐标为 $(x_0, y_0) = (0, 0)$ 时，函数值和各次差分就只由函数各系数决定。得到起始点之函数值及差分值，便可由(7)或(8)式得到走步后的各值。

在硬件受到限制的条件下，可以只保留前者或后者作为当前点，必要时利用(5)或(6)式进行变换。

四、最佳准则

1. 第一坐标

从某当前一点追踪 $|f(x, y)| = \min$ 时，候选格点达八个(图 3)。但利用函数发生中的其它信息，例如曲线走向、函数导数等将显著减少追踪过程的运算。曲线斜率小于 1 时，函数发生的走步只有二种可能(如图 4)。图 5 表明曲线斜率大于 1 时也只有二种可能。

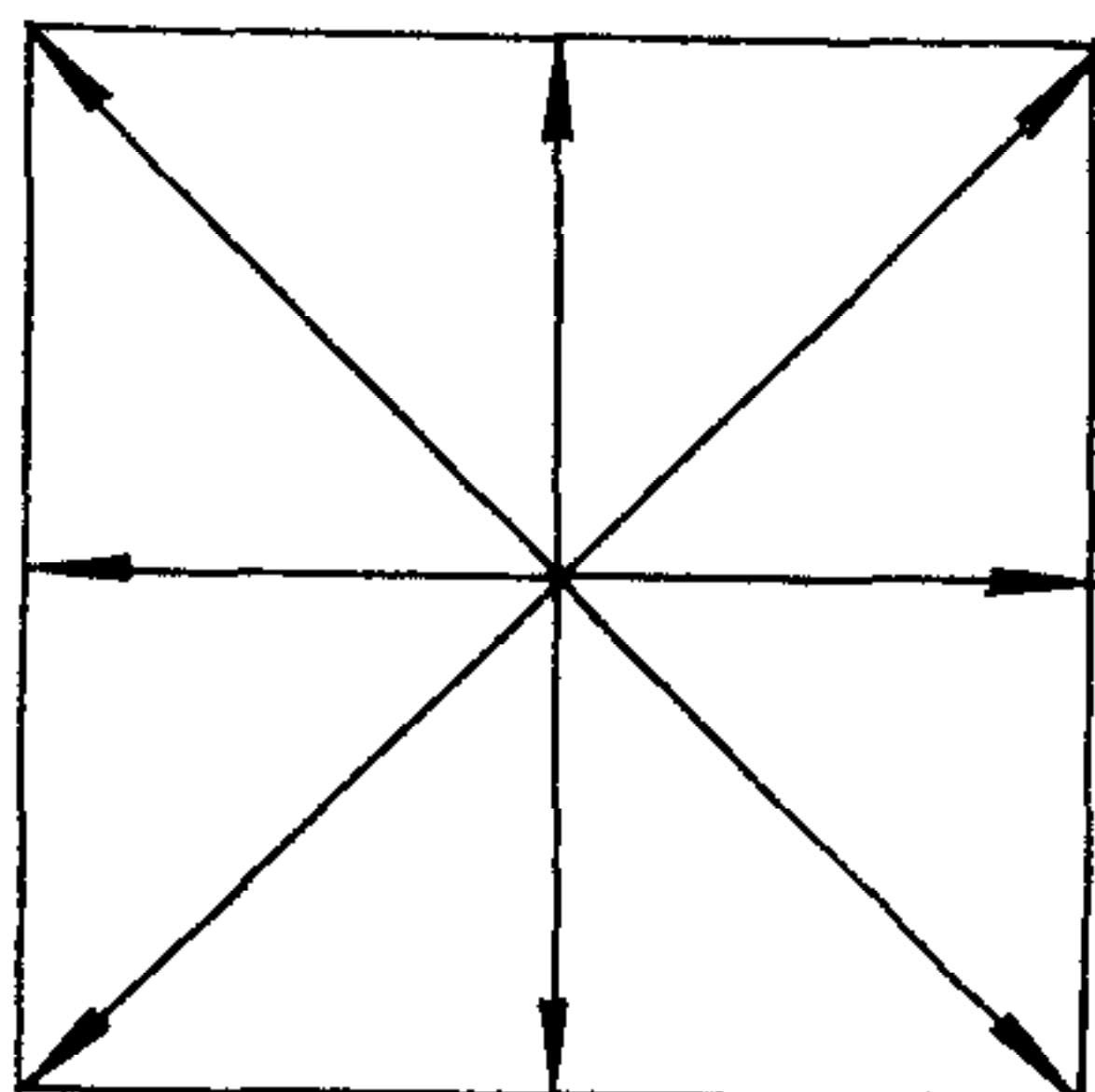


图 3

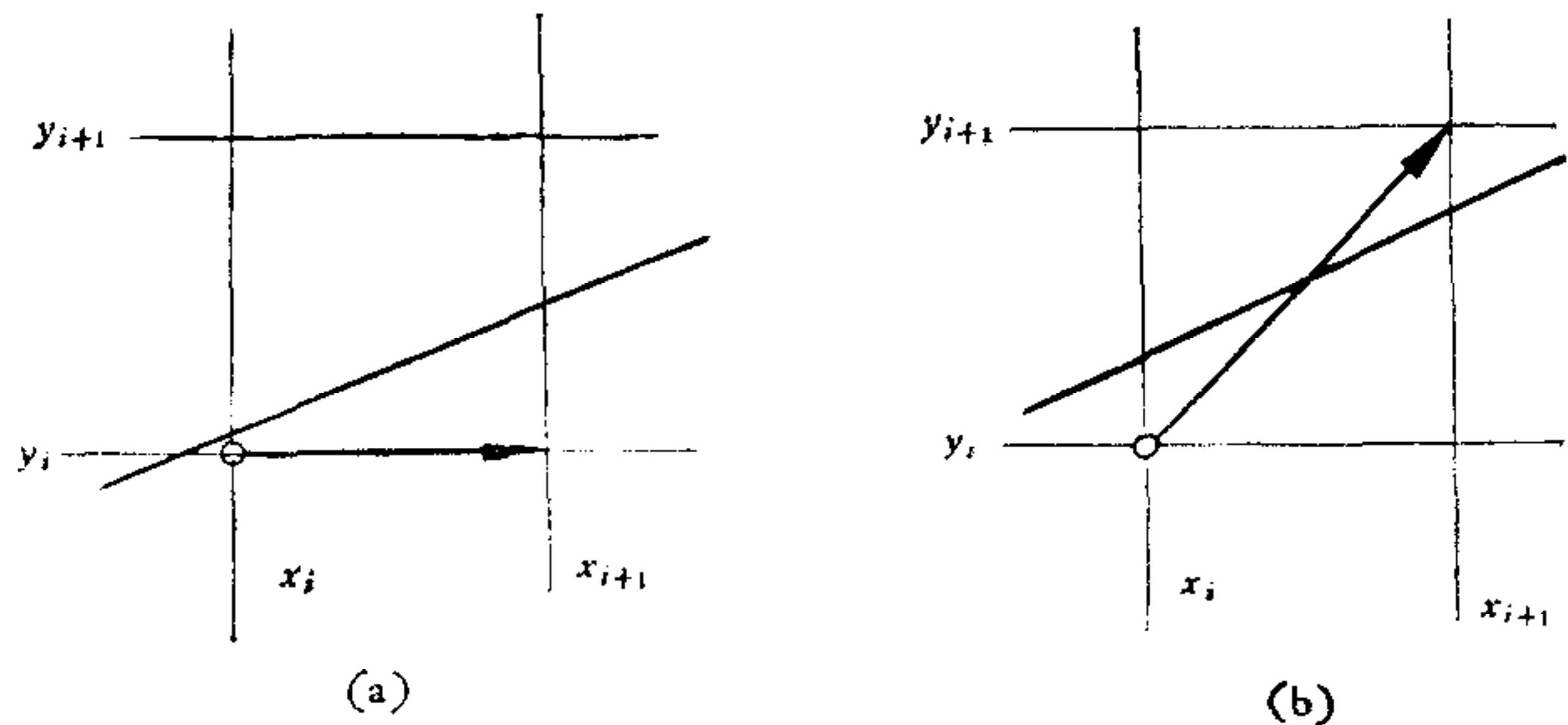


图 4

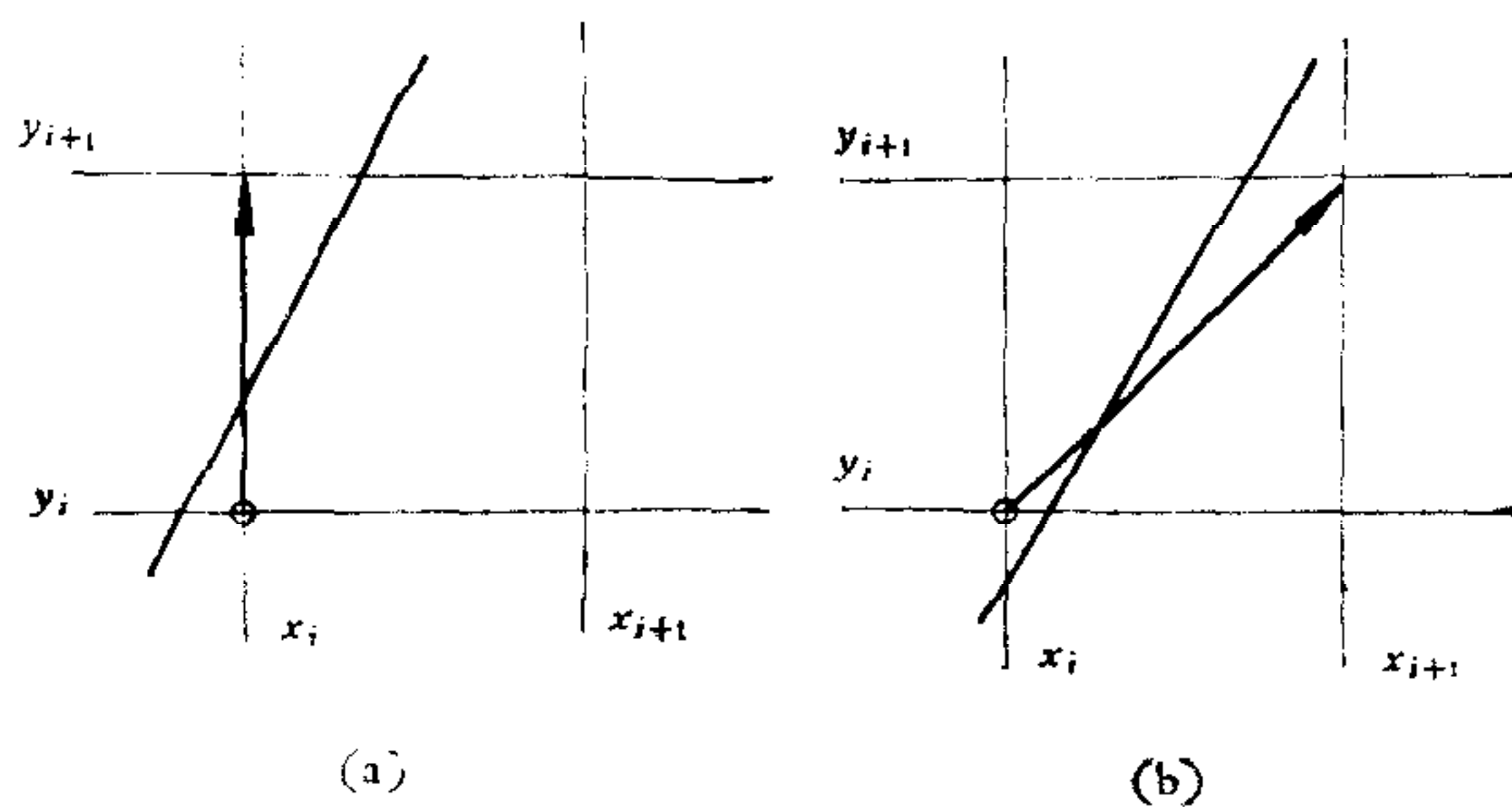


图 5

于是问题可归结为在当前点前进方向的二个候选格点中判别哪一点使 $|f(x, y)| = \min$ 。如此逐点追踪,就能获得给定函数 $f(x, y) = 0$ 的最佳逼近曲线。图 6 就是据此逼近给定函数格点集合的直观示例。黑色圆点为满足最佳逼近原则的格点。

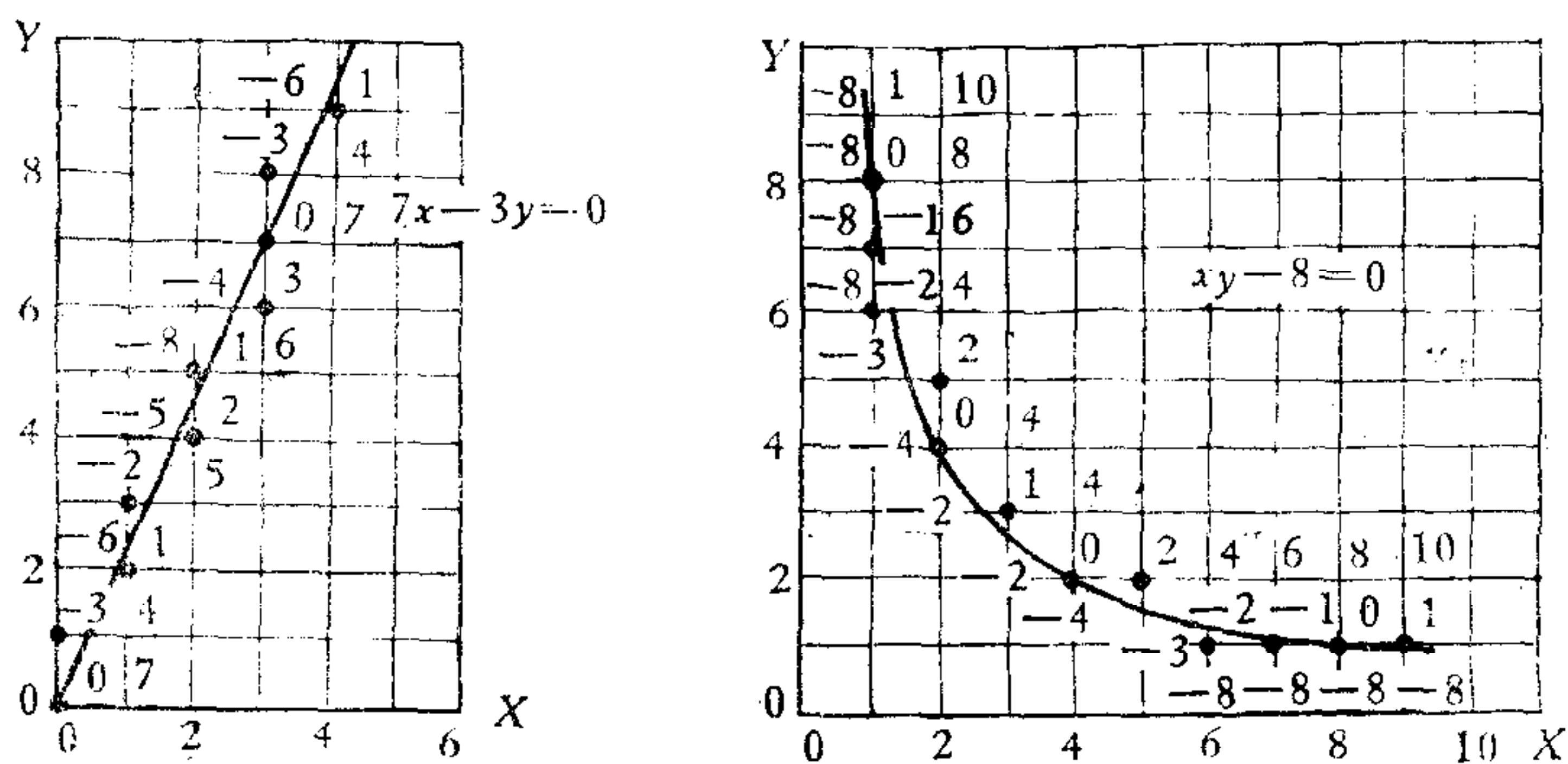


图 6

当曲线的大方向 x, y 轴走步趋势已经先行判明的情况下,函数导数绝对值 $\left| \frac{dy}{dx} \right| < 1$ 时,曲线走向更接近于 x 轴。这时,走步的情况只能是 x 方向单独走步,或 x, y 两个方向同时走步。绝不会出现 y 方向单独走步。同理,如果 $\left| \frac{dy}{dx} \right| > 1$,只可能 y 或 x, y 方向前进而不会有 x 方向单独走步。

为了讨论方便,把曲线走向更为接近的坐标称为“第一坐标”,而把另一坐标称为“第二坐标”。以图 6 中直线 $7x - 3y = 0$ 的逼近点集为例, $\frac{dy}{dx} = \frac{7}{3} > 1$,故 y 是第一坐标, x 是第二坐标。

在数字函数发生中利用 $|\Delta y|$ 是否大于 $|\Delta x|$ 的概念来识别第一坐标较为方便。如果从函数上一点 (x_i, y_i) 在 x, y 方向各前进一步到 (x_{i+1}, y_{i+1}) ,则 $f(x_{i+1}, y_{i+1}) = \Delta x_i + \Delta y_i$ 。由于是追踪函数,必将力图使函数值最小,故 Δx_i 和 Δy_i 恒有相反的符号,它们相减而得函数值。若有 $|\Delta x| > |\Delta y|$,则 y 方向走步可恒定前进, x 方向则选择前进。故 y 坐标是第一坐标, x 为第二坐标。综上所述,在利用差分计算中可按以下原则确定第一坐标:

- 1) 若 $|\Delta x| > |\Delta y|$, 则取 y 为第一坐标, x 为第二坐标;
- 2) 若 $|\Delta y| > |\Delta x|$, 则取 x 为第一坐标, y 为第二坐标;

3) 若 $|\Delta_x| = |\Delta_y|$, 可任选 x 或 y 为第一坐标.

在函数发生过程中可以让第一坐标持续前进, 同时逐点判别何时应让第二坐标同时走步. 这种原理使走步判别极为简单, 且使第二坐标的选择走步总是与第一坐标走步同时进行. 这样既提高函数发生速度, 也提高其平滑性及精确性.

2. 大方向的确定

在数字函数发生过程中, 需要事先知道曲线发生的大方向, 加上第一坐标的信息, 候选格点数才降为 2 个. 因此如何确定大方向是至为重要的. 确定大方向包括函数发生开始时的起始大方向以及函数过极点后大方向的变换.

1) 函数发生时的起始大方向

曲线起始段的方向一经确定, 函数的发生就按既定顺序搜索前进. 由于引入了第一坐标的概念, 指定第一坐标起始方向后, 第二坐标方向便可自动确定, 故用 1 bit 即可指定起始大方向. 如果把顺时针和逆时针加以衍生和扩充, 则对于任何函数, 只要它具有正负区域的特性, 都可用“顺向”(函数的负区域在前进方向的右方)及“逆向”(函数的负区域在前进方向的左方)来指定起始方向. 图 7 是顺向的几个例子.

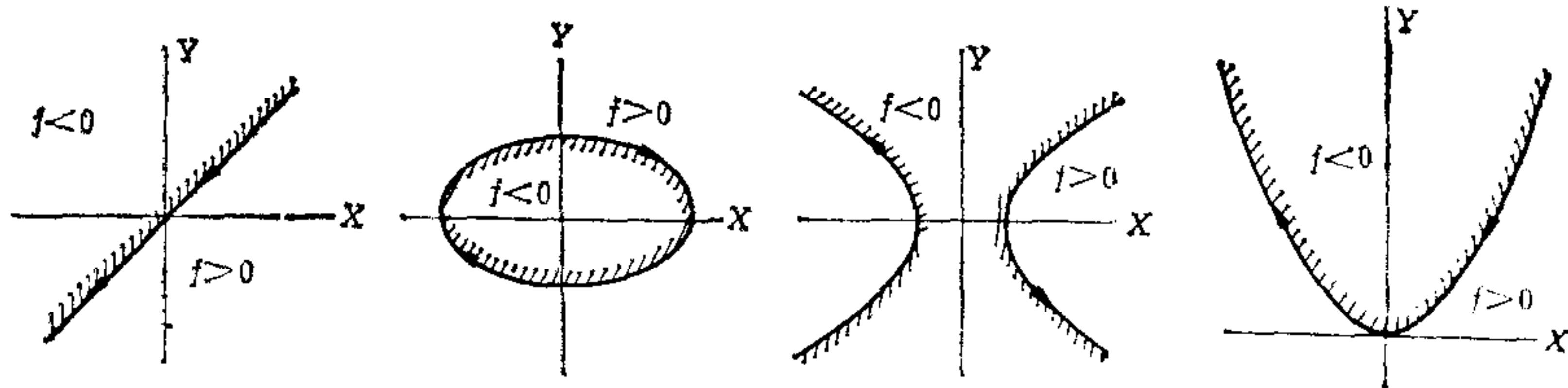


图 7

$$\begin{array}{cccc}
 Ax + By = 0 & Ax^2 + By^2 + C = 0 & Ax^2 + By^2 + C = 0 & Ax^2 + By = 0 \\
 A > 0, B < 0 & A > 0, B > 0, C < 0 & A > 0, B < 0, C < 0 & A > 0, B < 0
 \end{array}$$

给定“顺向”、“逆向”后, 大方向 $\pm x, \pm y$ 就确定了. 可以证明函数的区域分布与函数一次差分之间有着固有的关系(如图 8 所示). 可列出“顺向”及“逆向”时走步大方向与向前一次差分的符号关系如下:

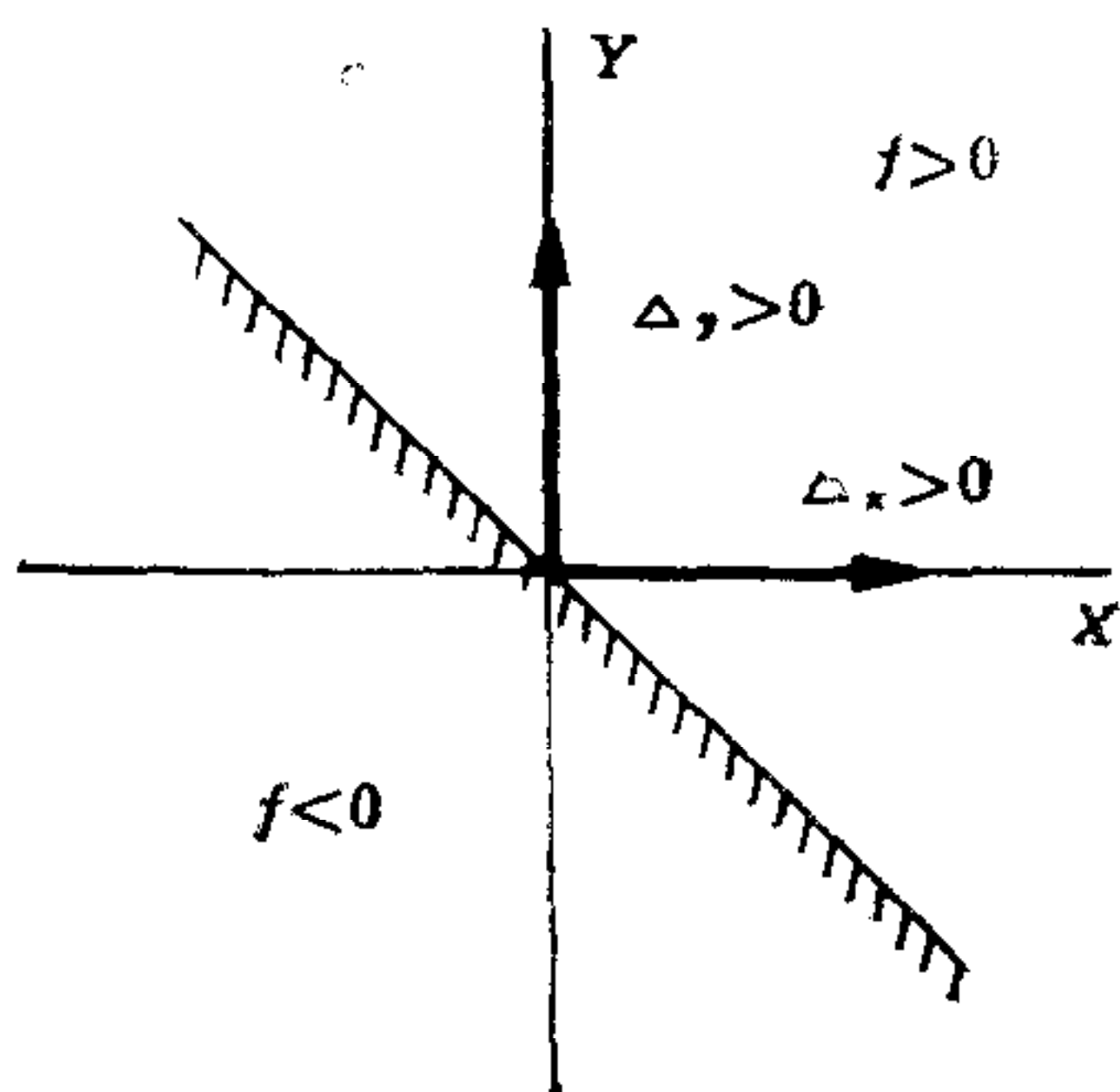


图 8

向前一次差分的符号关系如下:

	$\Delta_x < 0$	$\Delta_x > 0$		$\Delta_x < 0$	$\Delta_x > 0$
$\Delta_y > 0$	+X, +Y	+X, -Y	$\Delta_y > 0$	-X, -Y	-X, +Y
$\Delta_y < 0$	-X, +Y	-X, -Y	$\Delta_y < 0$	+X, -Y	+X, +Y
	顺向			逆向	

从这两张表可导出以下大方向关系式:

$$\left. \begin{array}{l}
 \text{“顺向”}: \text{sign}(X) = \text{sign}(\Delta_y) \\
 \text{sign}(Y) = -\text{sign}(\Delta_x)
 \end{array} \right\} \quad (9)$$

$$\left. \begin{array}{l}
 \text{“逆向”}: \text{sign}(X) = -\text{sign}(\Delta_y) \\
 \text{sign}(Y) = \text{sign}(\Delta_x)
 \end{array} \right\} \quad (10)$$

给定“顺向”“逆向”, 就可按(9)或(10)式自动确定走步起始大方向.

2) 函数过极点后大方向的变换

曲线经过极点时, 极点前后的斜率具有相反的符号, 故函数发生的大方向在过任一极

点后须加以变更。

根据逐点追踪 $|f(x, y)| = \min$ 的原则, 第二坐标走步仅在第二坐标一次差分与仅走第一坐标时的函数值异符号时成立(详见最佳准则这一节)。因此第二坐标的前进方向由第二坐标一次差分值的符号与仅走第一坐标后函数值符号的异同决定。如为异符号, 则维持原方向不变, 如为同符号则改变第二坐标方向。由于在一定的方向上第二坐标一次差分是不会改变符号的, 于是也可归结为: 仅走第一坐标的函数值变号时, 第二坐标换向。

如图 9 所示, 设第一坐标为 x , $\Delta_y > 0$, 点 1, 2, 3, 4, 5, 6 是发生曲线上的点。从点 1 看, A 点的 $f_x < 0$, 而 $\Delta_y > 0$, 故 y 走步应为正方向。在点 5, C 点函数值 $f_x > 0$, $\Delta_y > 0$ 故 y 走步应为负方向, 如此类推。或者, 仅就 f_x 之符号断定: 在图 10 曲线的前一区段中显然已有 $+X, +Y$ 的走向, 且总是以 x, y 同时走步与本区段相联接。这说明本区段曲线开始之前总有 $f_x < 0, \Delta_y > 0$ 。进入本区段后, 若 $f_x < 0$, 维持 y 正方向, 如 $f_x > 0$ 则使 y 走负方向。

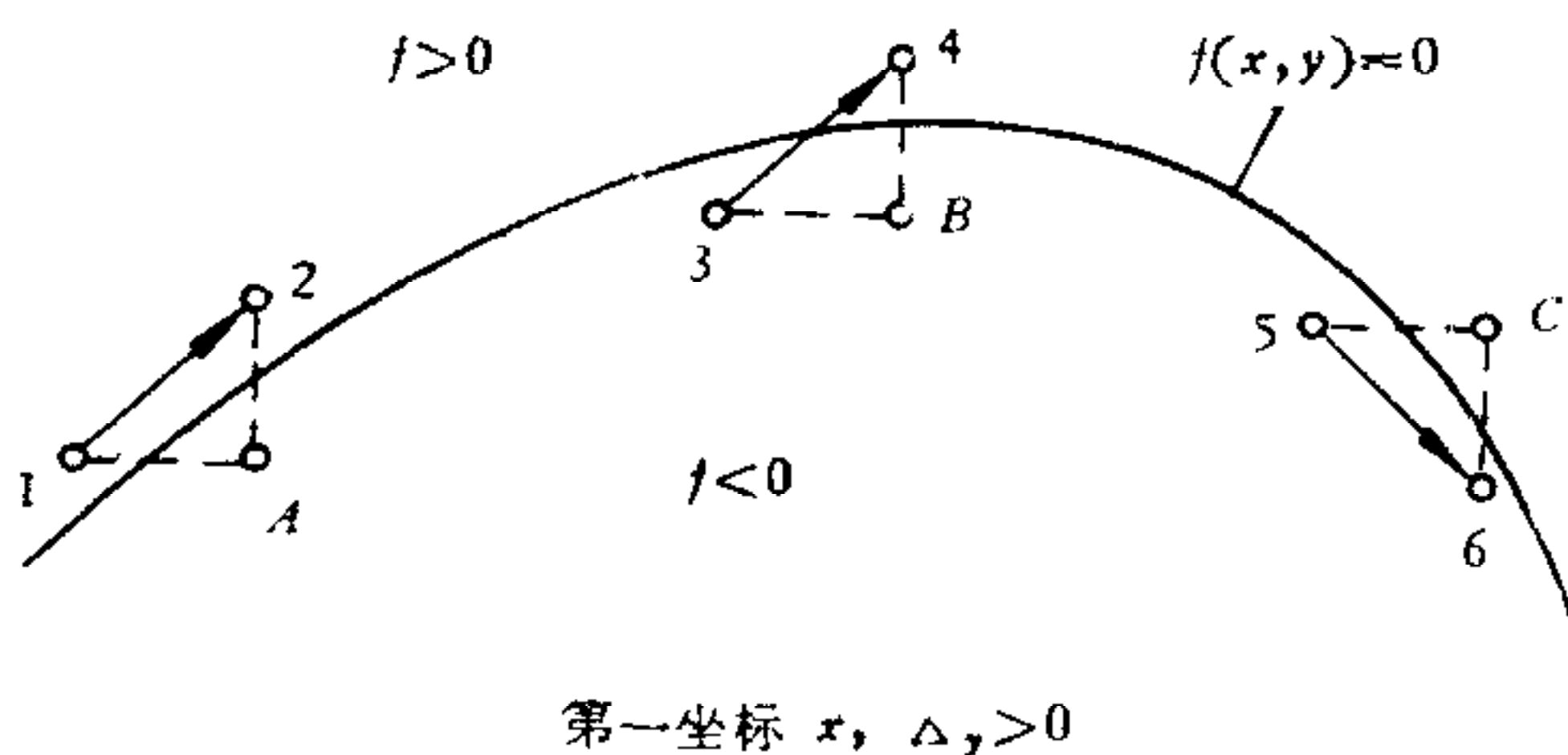


图 9

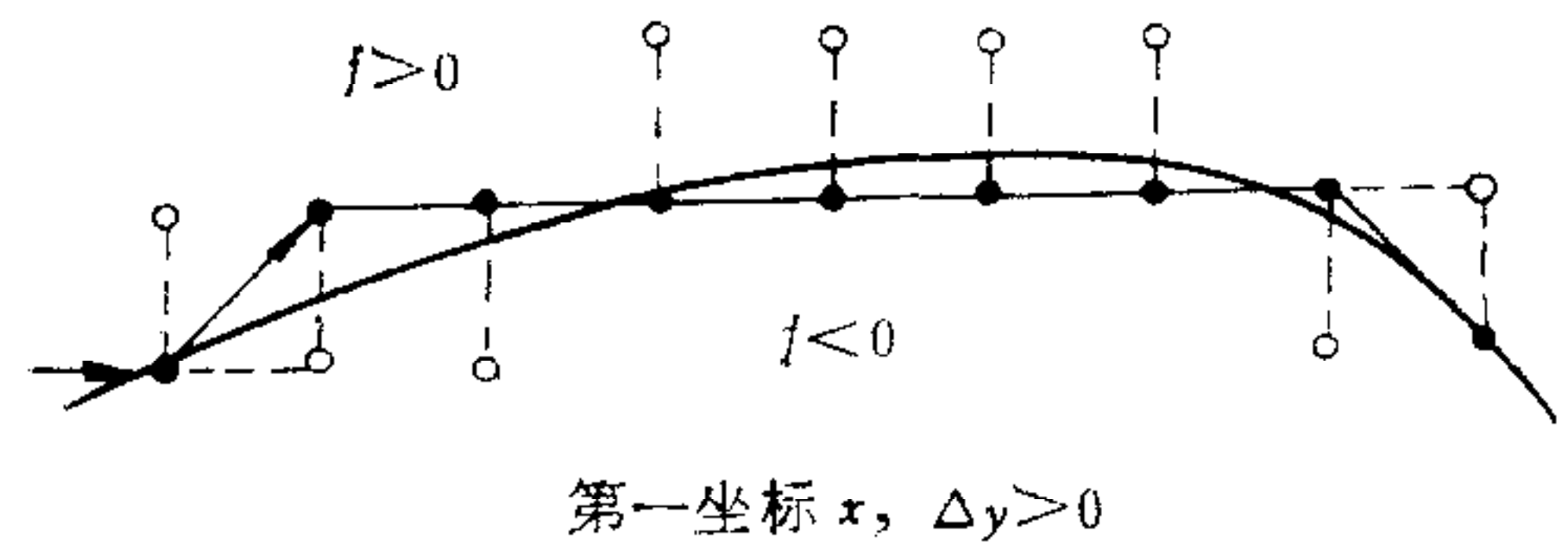


图 10

这一换向方法保证了在任何情况下都能正确走步, 既不会误判也不引入附加误差。由于逐步监视第二坐标走步方向的正确性, 故曲线发生逐点自行适应, 如图 10。从而使大方向判定正确无误。

3. 最佳准则

大方向和第一坐标确定之后剩下两个可供选择的候选格点, 即两种可能的走步情况: 1) 第一坐标单独走步; 2) 第一坐标、第二坐标同时走步。于是问题归结为判定什么条件下第二坐标走步能满足 $|f(x, y)| = \min$ 。这就是最佳准则的任务。

设函数发生的大方向是 $+X, +Y$ 方向, 第一坐标是 x 坐标。又设当前点函数值为 f , 第一坐标单独前进一步后的函数值为 f_x , 第二坐标同时前进一步后的函数值为 $f_{x,y}$, 两坐标正方向上一次差分为 Δ_x, Δ_y , 则有 $f_x = f + \Delta_x$, $f_{x,y} = f + \Delta_x + \Delta_y$ 。由于 $|f_{x,y}| < |f_x|$ 时必有 $(f + \Delta_x)$ 与 Δ_y 异号, 故有 1) $\text{sign}(f + \Delta_x) = -\text{sign}(\Delta_y)$, 且 $2|f + \Delta_x| < |\Delta_y|$, 得 $|f_x| < |f_{x,y}|$, 故第一坐标单独走步; 2) $\text{sign}(f + \Delta_x) = -\text{sign}(\Delta_y)$, 且 $2|f + \Delta_x| \geq |\Delta_y|$, 得 $|f_x| \geq |f_{x,y}|$, 故二个坐标同时走步。由此, 最佳准则可归纳为如下判别式:

$$\left. \begin{aligned} \text{sign}(f + \Delta_x) &= -\text{sign}(\Delta_y) \\ 2|f + \Delta_x| &\geq |\Delta_y| \end{aligned} \right\} \quad (11)$$

当满足判别式(11)时第一坐标、第二坐标同时前进。不满足时, 第一坐标单独走步。根据上述准则构成原理流程图 11。

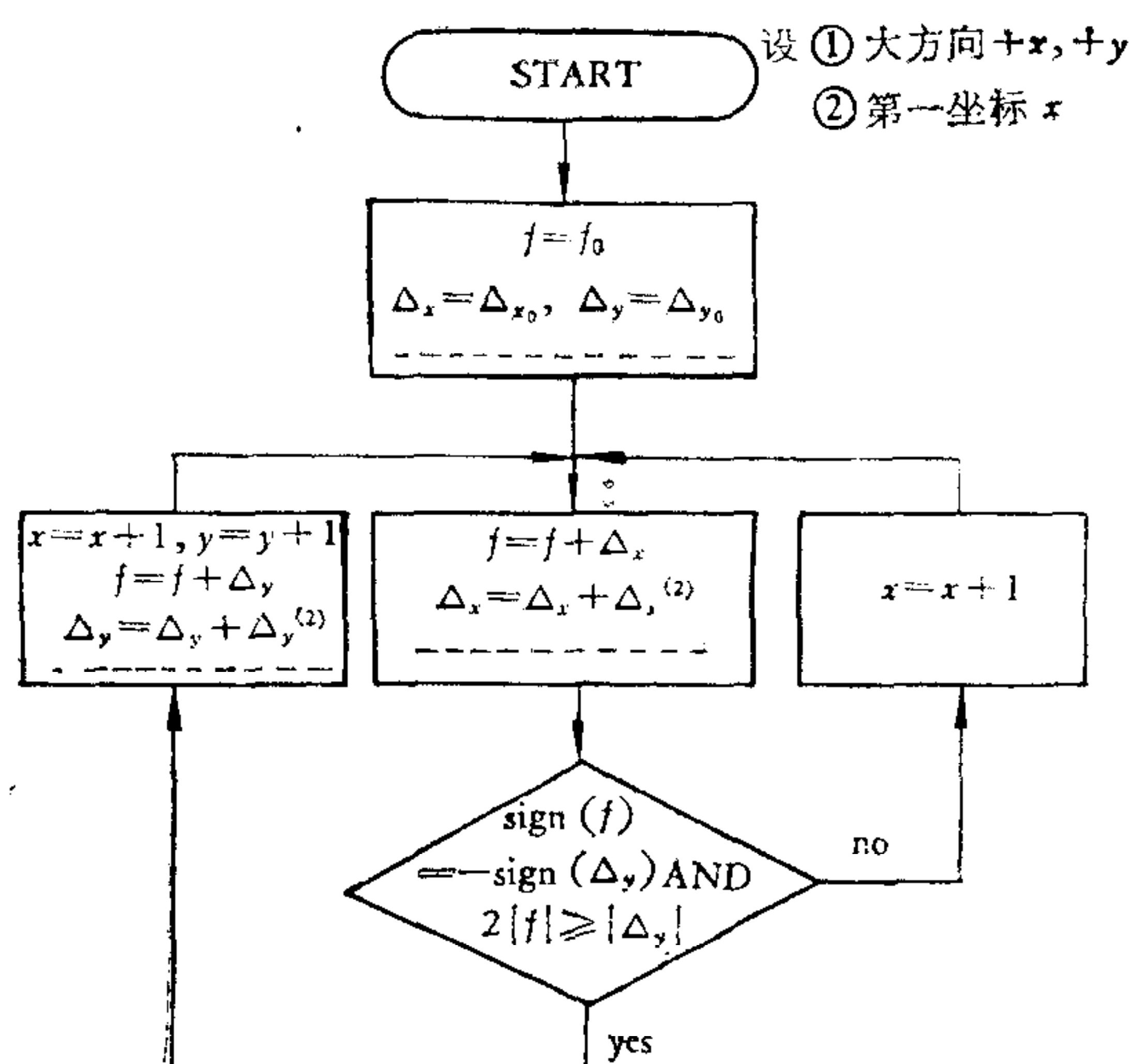


图 11

五、误差分析

为了分析方便,将误差定义为第一坐标值相同时,发生曲线上的格点与给定函数相应点之间在第二坐标方向的偏离.显然它大于法线方向上的偏离误差.

根据最佳准则,发生曲线上任一格点函数值 $|f|$ 必小于第二坐标上函数一次差分的二分之一.如果用下标 II 表示第二坐标各量,则最大误差 $\varepsilon_{1, \max}$ (正向)及 $\varepsilon_{2, \max}$ (负向)有如下关系式:

$$\begin{aligned} & \left| f'_{II} \varepsilon_{1, \max} + \frac{1}{2} f''_{II} \varepsilon_{1, \max}^2 + \frac{1}{6} f'''_{II} \varepsilon_{1, \max}^3 + \dots \right| \\ &= \frac{1}{2} \left| f'_{II} + \frac{1}{2} f''_{II} + \frac{1}{6} f'''_{II} + \dots \right| \end{aligned} \quad (12)$$

$$\begin{aligned} & \left| -f'_{II} \varepsilon_{2, \max} + \frac{1}{2} f''_{II} \varepsilon_{2, \max}^2 - \frac{1}{6} f'''_{II} \varepsilon_{2, \max}^3 + \dots \right| \\ &= \frac{1}{2} \left| -f'_{II} + \frac{1}{2} f''_{II} - \frac{1}{6} f'''_{II} + \dots \right| \end{aligned} \quad (13)$$

对于几种常用的典型函数的误差分析及实验,证明第二坐标的最大误差均为半步.

六、三维曲线

最佳法原理同样可用来发生三维函数.在发生三维曲线时正确选择坐标次序能使方程组中每一个方程都正确协调地满足最佳准则.

三维曲线的一般表示式为:

$$\left. \begin{aligned} f_1(x, y, z) &= 0 \\ f_2(x, y, z) &= 0 \end{aligned} \right\} \quad (14)$$

在坐标平面上的投影为

$$\left. \begin{aligned} f_A(x, y) &= 0 \\ f_B(y, z) &= 0 \\ f_C(z, x) &= 0 \end{aligned} \right\} \quad (15)$$

在发生上述曲线时为了便于确定在三维空间中的走步方向，应将 x, y, z 三个坐标区分为第一坐标、第二坐标和第三坐标，其方法与二维曲线类同。例如，如果方程组 (15) 有 $|\Delta_{A,x}| < |\Delta_{A,y}|, |\Delta_{B,y}| < |\Delta_{B,z}|$ 则应选 x 为第一坐标， y 为第二坐标， z 为第三坐标。用最佳准则发生三维曲线的流程图见图 12。

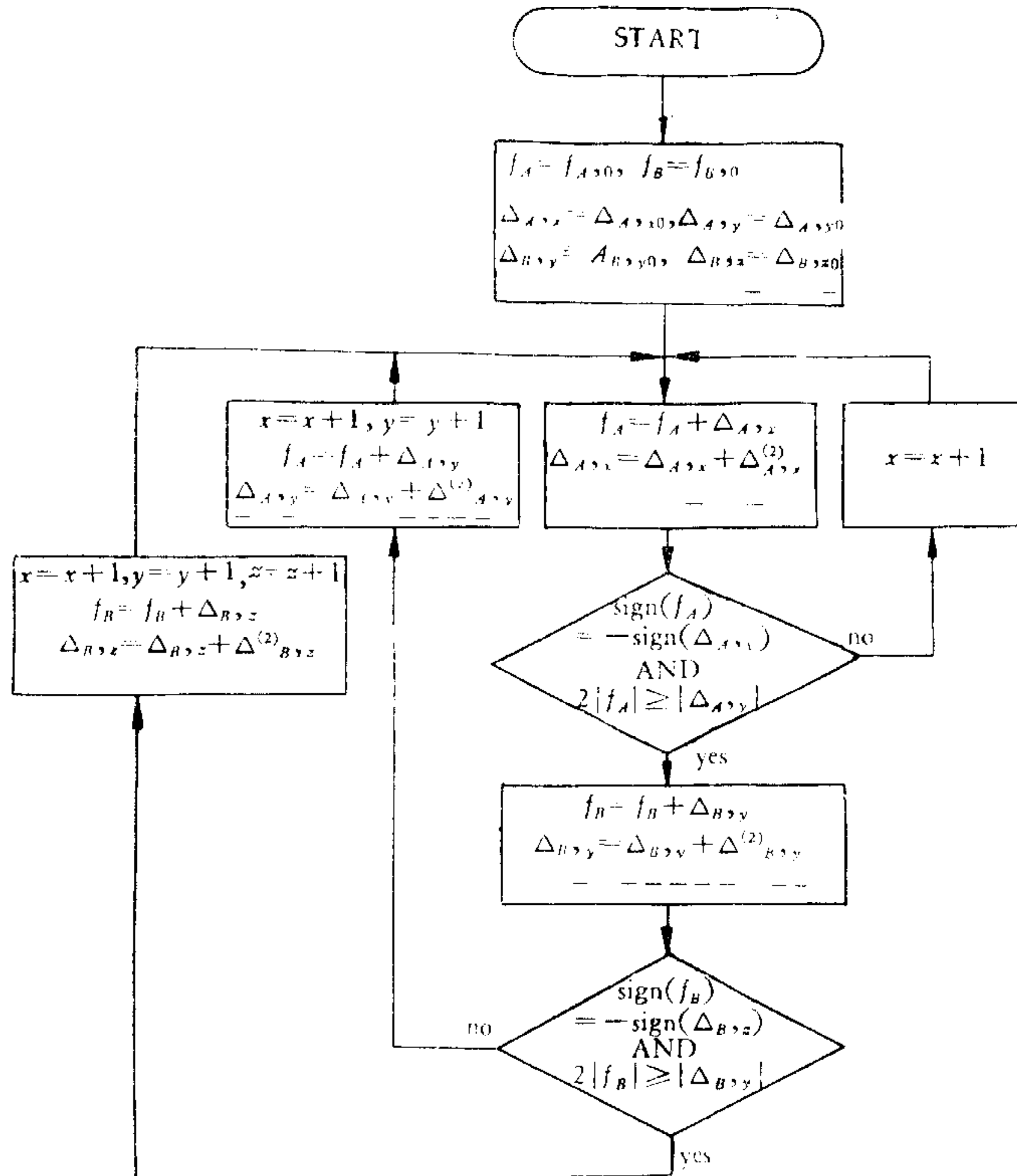


图 12

七、结 论

(1) 最佳法以追踪函数值最小来逼近给定函数，使走步误差在半步之内，因此函数发生的途径是唯一的、可逆的和对称的。

(2) 采取划分坐标次序、协同各坐标同时走步的方式，既使搜索的格点减到最少，(算法本身的运算量也相应减小)，又提供了最高限的输出效率。二维和三维曲线的发生也自然地统一。

(3) 充分运用差分算法的特点，使最佳法仅需进行加减运算及逻辑条件判别，这对软件实现或硬件实现都是合宜的。在硬件条件相同时，最佳法将显著提高函数发生的精度及速度。

参 考 文 献

- [1] 齐田伸雄, 安川电机, 38, No. 3.
[2] Fischer, H., Klinge, R., Schendel, C., Siemens-Z., H. 2, 1966.
[3] Kaiwa, T., Lnaba, S., Control Engineering, 9(1962), No. 3.

OPTIMAL METHOD OF DIGITAL FUNCTION GENERATION

LU HAORU

(Zhejiang University)

ABSTRACT

The investigated algorithm named Optimal Method can optimize the process of digital function generation: stepping errors due to the algorithm may be minimized to keep within 1/2 step (accuracy), the path of curve generation is independent of its starting point (uniqueness), the path of curve generation is just the same when its direction is reversed (reversibility), the algorithm suits both 2-dimensional and 3-dimensional curve generation (identity) and the algorithm is simple and practical in use (practicality).

As a criterion, Optimal Method traces the given function by searching its minimal absolute value among adjacent candidate mesh points. According to its values of difference functions the X- and Y-axes are assigned to be so called first and second axes or vice versa in order to make mesh point choice quite simple and easy. A compact algorithm is given to determine the beginning direction of function generation and also to change the direction so as to get an accurate and correct bending in the process of curve generation.