最优化理论与方法

研究生学位课

陈军华（副教授、主任）

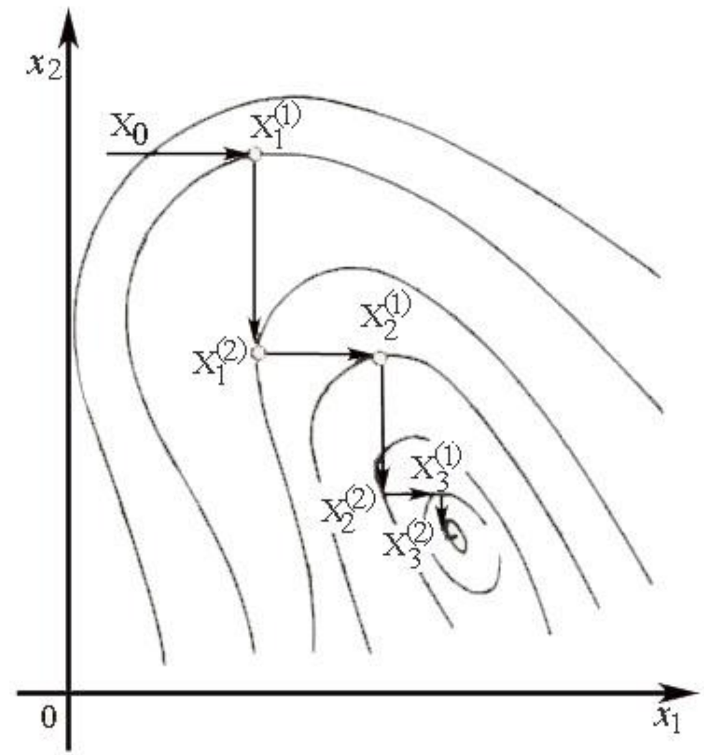
5 多变量无约束问题的直接法

- **直接法**
  - 直接法就是直接利用目标函数所反映的信息确定寻优方向，求得最优解，并不利用函数的解析性质，因此对函数的性质要求较少，函数可以不连续，或在某些点无导数等，但直接法的收敛速度一般较慢。

- **坐标轮换法**
  - 基本原理
  - 把含有n 个变量的优化问题轮换地转化为单变量（其它变量视为常量）的优化问题。
  - 先选定一个初始点 x_0 作为第一轮搜索的始点，依次沿n个坐标轴方向进行一维搜索，每次只在一个坐标轴方向上改变相应变量的值，其它$(n-1)$ 个变量均保持不变.

- **坐标轮换法**
  - 以两维函数寻优为例说明过程。
  - 设$\boldsymbol{x} \in \boldsymbol{E}^2$，取向量$\boldsymbol{e}_1 = (1,0)^T, \boldsymbol{e}_2 = (0,1)^T$,$\boldsymbol{x}^0$为起点。
  - $\min\limits_{\lambda} \boldsymbol{f}(\boldsymbol{x}^0 + \boldsymbol{\lambda}\boldsymbol{e}_1) = \boldsymbol{f}(\boldsymbol{x}^0 + \boldsymbol{\lambda}_1\boldsymbol{e}_1)$
  - $x_1^{(1)} = \boldsymbol{x}^0 + \boldsymbol{\lambda}_1\boldsymbol{e}_1$
  - 以$x_1^{(1)}$点出发沿第二坐标寻优。
  - $\min\limits_{\lambda} \boldsymbol{f}(x_1^{(1)} + \boldsymbol{\lambda}\boldsymbol{e}_2) = \boldsymbol{f}(x_1^{(1)} + \boldsymbol{\lambda}_2\boldsymbol{e}_2$
  - $x_1^{(2)} = x_1^{(1)} \quad + \boldsymbol{\lambda}_2\boldsymbol{e}_2$
  - 两个坐标方向轮换一遍后，重新

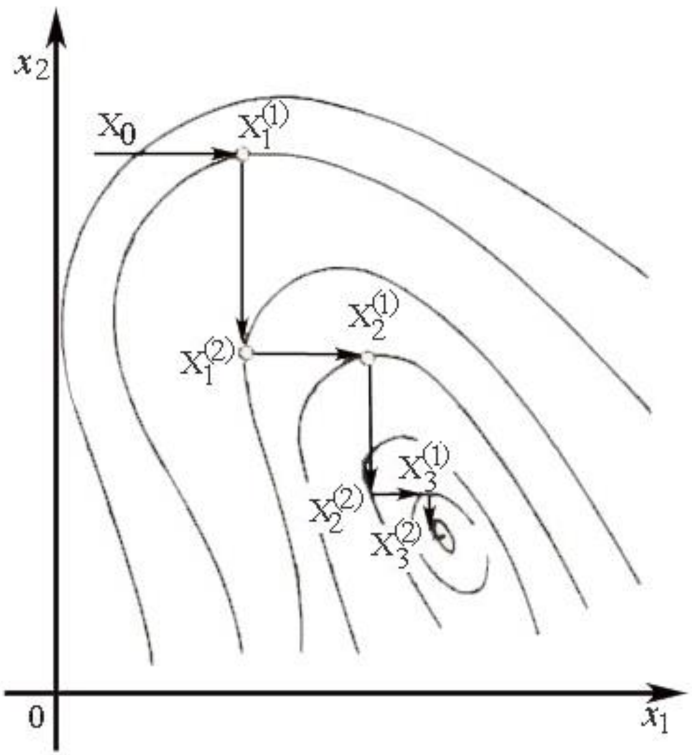
再轮换，取$x_1 \ = x_1^{(2)}$为新起点



- **坐标轮换法**
  - $\min_{\lambda} f(x_1^{(2)} + \lambda e_1) = f(x_1^{(2)} + \lambda_1 e_1)$
  - $x_2^{(1)} = x_1^{(2)} \quad + \lambda_1 e_1$
  - 及：
  - $\min_{\lambda} f(x_2^{(1)} + \lambda e_2) = f(x_2^{(1)} + \lambda_2 e_2)$
  - $x_2^{(2)} = x_2^{(1)} \quad + \lambda_2 e_2$

$$\vdots$$

直到$\|x_n - x_{n-1}\| < \varepsilon$



坐标轮换法算法步骤

- 1：取$x_1^1 \in E^n$，置坐标轴方向：

$$e_1 = [1, 0, 0, \cdots, 0]^T$$
$$e_2 = [0, 1, 0, \cdots, 0]^T$$
$$e_n = [0, 0, 0, \cdots, 1]^T$$

- 令 $j = 1, k = 1, x_k^j = x_1^1$;

- 2：求单变量极值问题的最优解

$$\min_{\lambda} f(x_k^j + \lambda e_j) = f(x_k^j + \lambda_j e_j)$$
$$x_k^{j+1} = x_k^j + \lambda_j e_j$$

- 3：判断是否满足$j = n$，若 $j = n$，转下一步，否 j++，转2.
- 4：检验是否满足收敛性判别准则

$\left\| x_k^{n+1} - x_k^1 \right\| < \varepsilon$,否则令 $x_{k+1}^1 = x_k^{n+1}$， $\mathrm{k}++, j = 1$,转2.

- **例5_1**
  - 用坐标轮换法求$\min f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$

$x^0 = (0,3)^T, \varepsilon = 0.04.$

- **例5_1**
  - 用坐标轮换法求$\boldsymbol{minf}(\boldsymbol{x}) = (\boldsymbol{x_1} - 2)^4 + (\boldsymbol{x_1} - 2\boldsymbol{x_2})^2$

$\boldsymbol{x}^0 = (0,3)^T, \varepsilon = 0.04.$

解：取向量$\boldsymbol{e_1} = (1,0)^T, \boldsymbol{e_2} = (0,1)^T$

$$\boldsymbol{x}^0 + \lambda_1 \boldsymbol{e_1} = (\lambda_1, 3)^T$$

  - $\boldsymbol{f}(\boldsymbol{x}^0 + \lambda_1 \boldsymbol{e_1}) = [\lambda_1 - 2]^4 + [\lambda_1 - 6]^2$
  - 求 $\min\limits_{\lambda} \boldsymbol{f}(\boldsymbol{x}^0 + \lambda \boldsymbol{e_1})$
  - $\frac{d\boldsymbol{f}(\boldsymbol{x}^0 + \lambda_1 \boldsymbol{e_1})}{d\lambda_1} = 0 \quad \rightarrow \quad \lambda_1 = 3.13$

$$x_1^{(1)} = \boldsymbol{x}^0 + \lambda_1 \boldsymbol{e_1} = \begin{bmatrix} 0 \\ 3 \end{bmatrix} + 3.13 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3.13 \\ 3 \end{bmatrix}$$

- **例5_1**

$x_1^{(1)} = \begin{bmatrix} 3.13 \\ 3 \end{bmatrix}$ 出发，沿$\boldsymbol{e}_2$方向搜索，求得$x_1^{(2)}$

- $x_1^{(2)} = x_1^{(1)} + \boldsymbol{\lambda}_2\boldsymbol{e}_2 = \begin{bmatrix} 3.13 \\ 3 + \boldsymbol{\lambda}_2 \end{bmatrix}$
- $\boldsymbol{f}\left(x_2^{(1)} + \boldsymbol{\lambda}_2\boldsymbol{e}_2\right) = (1.13)^4 + (3.13 - 6 - 2\boldsymbol{\lambda}_2)^2$
- 求 $\min\limits_{\boldsymbol{\lambda}} \boldsymbol{f}(x_2^{(1)} + \boldsymbol{\lambda}_2\boldsymbol{e}_2)$
- $\dfrac{d\boldsymbol{f}\left(x_2^{(1)} + \boldsymbol{\lambda}_2\boldsymbol{e}_2\right)}{d\boldsymbol{\lambda}_2} = 0 \quad \rightarrow \quad \boldsymbol{\lambda}_2 = -1.44$

$$x_1^{(2)} = x_1^{(1)} + \boldsymbol{\lambda}_2\boldsymbol{e}_2 = \begin{bmatrix} 3.13 \\ 1.56 \end{bmatrix}$$

- **例5_1**

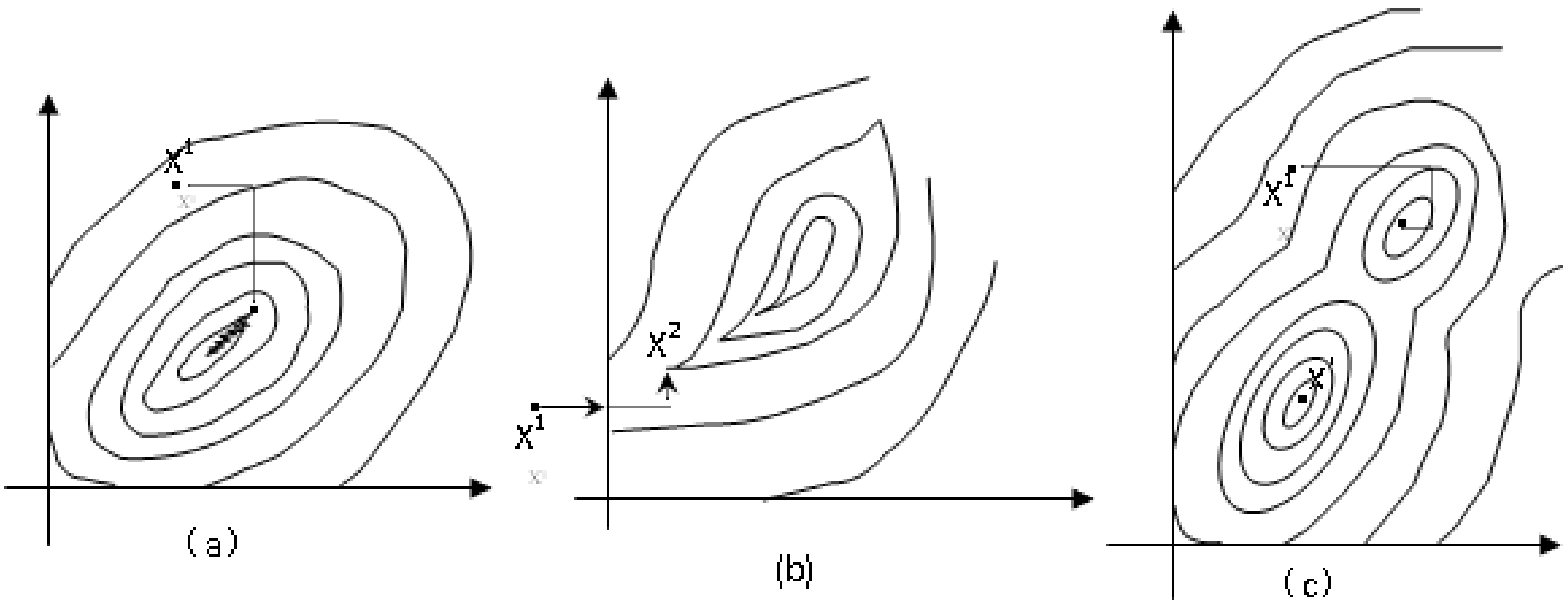
终止判别 $\left\| x_1^{(2)} - x_0 \right\| = \sqrt{(3.13-0)^2+(1.56-3)^2} = 3.45 > \varepsilon$

不满足，需要继续迭代，各轮迭代计算数据见下表，最优解为：

$$x^* = (2.24, 1.12)^T$$

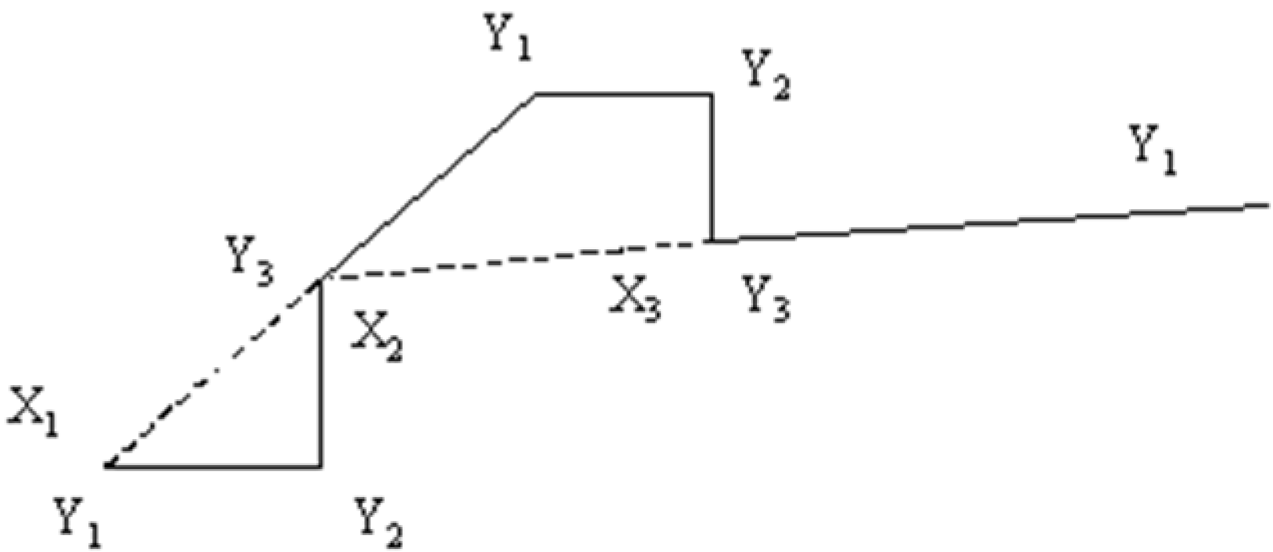
循环迭代序号	X_0	$e_1=[1,0]^T$		$e_2=[0,1]^T$		$f(X_k^{(2)})$	$\|X_k^{(2)}-X_0\|$	是否满足收敛准则
		$t_k^{(1)}$	$X_k^{(1)}$	$t_k^{(2)}$	$X_k^{(2)}$			
1	0.00，3.00	3.13	3.13，1.56	1.44	3.13，1.56	1.63	3.45	否
2	3.13，1.56	−0.50	2.63，1.56	−0.25	2.63，1.31	0.16	0.56	否
3	2.63，1.31	−0.19	2.44，1.31	−0.09	2.44，1.22	0.04	0.21	否
4	2.44，1.22	−0.09	2.35，1.22	−0.05	2.35，1.17	0.015	0.10	否
5	2.35，1.17	−0.06	2.29，1.17	−0.03	2.29，1.14	0.007	0.06	否
6	2.29，1.14	−0.04	2.25，1.14	−0.02	2.25，1.12	0.004	0.045	否
7	2.25，1.12	−0.03	2.22，1.12	−0.01	2.22，1.11	0.002	0.03	是

- **坐标轮换法说明**
  - 优点是算法简单，计算量小。其缺点是计算效率低，对高维问题尤为突出．因此，坐标轮换法通常用于维数较低的优化问题（一般 n<10）．
  - 对于可微函数，可能收敛于梯度为零的局部最优点。



- **步长加速法**
  - 又称模式搜索法（**Pattern Search Method**）。
  - 由胡克（**Hooke**）和基夫斯（**Jeeves**）于**1961**年提出的。
  - 它不仅易于编制计算机程序，而且具有追循谷线加速移向最优点的性质。
  - 基本思想从几何上讲，就是寻找具有较小函数值的“山谷”，力图使迭代产生的序列沿“山谷”逼近极小点。

- **步长加速法–搜索过程**
  - 步长加速法由“探测移动”和“模式搜索”两个交替的动作构成。
  - 探测移动：依次沿n个坐标轴进行，用以确定新的基点和有利于函数值下降的方向。
  - 模式搜索：沿相邻两个基点连线方向进行，试图顺着“山谷”使函数值下降的更快。



1、探测移动

- 设当前基点x^k,目标函数值$f\left(x^k\right)$。记初始临时矢点为y^1（第一次取为初始迭代点x^1）。
- 沿第1个坐标方向e_1以某一步长δ进行探索，即在$y^1+\delta e_1$和$y^1-\delta e_1$这两点中寻求能使目标函数值下降的点，作为临时矢点y^2;
- 由此点出发沿另一坐标方向进行同样的探索，如能得到更好的点，就以该点代替前点作为新的临时矢点。如此沿各个坐标方向轮流探查一遍，最后的临时矢点为y^{n+1}
- 若$f\left(y^{n+1}\right)<f(x^k)$，转向模式搜索。否则，缩短步长，若步长满足精度要求，计算结束，得到近似极小点； 若步长不满足精度要求，以缩短后的步长继续进行探测移动。

2、模式搜索

- 令$x^{k+1}=y^{n+1}$，由基点x^k到新基点x^{k+1}构成第k个模式$x^{k+1}-x^k$。可以认为这是使目标函数值得以改善的最有利的移动方向，即在x^k处沿这一方向，目标函数值下降“最快”。显然，这一方向近似于目标函数的负梯度方向。
- 现假定在新基点x^{k+1}附近进行类似的探测，其结果可能和在x^k处的情形相同，故略去这步探测而第k个模式扩大α倍（加速），令

$$y^1=x^{k+1}+\alpha\left(x^{k+1}-x^k\right)=(1+\alpha)x^{k+1}-x^k,$$

- 完成模式搜索，转向以y^1为临时矢点的探测移动。

- **步长加速法的步骤**
  - 1、给定初始步长$\delta > 0$，加速因子$\alpha \geq 1$，缩减率$\beta \in (0,1)$，允许误差$\varepsilon > 0$及初始迭代点x^1，置$y^1 = x^1, k = 1, j = 1$；
  - 2、如果$f\left(y^j + \delta e_j\right) < f(y^j)$，令$y^{j+1} = y^j + \delta e_j$，转第4步；否则转第3步。
  - 3、如果$f\left(y^j - \delta e_j\right) < f(y^j)$，令$y^{j+1} = y^j - \delta e_j$，转第4步；否则，令$y^{j+1} = y^j$，转第4步。
  - 4、如果$j < n$，则置$j = j + 1$，转第2步；否则，转第5步；
  - 5、如果$f\left(y^{n+1}\right) < f(x^k)$，则转第6步；否则，转第7步；
  - 6、令$x^{k+1} = y^{n+1}$，$y^1 = x^{k+1} + \alpha\left(x^{k+1} - x^k\right) = (1 + \alpha)x^{k+1} - x^k$，置$k = k + 1, \mathrm{j} = 1$,转第2步；
  - 7、$\delta = \beta\delta, y^1 = x^k, x^{k+1} = x^k$，置$k = k + 1, j = 1$,转第2步。

- **例5_2**
  - 用步长加速法求$\min f(x) = (1 - x_1)^2 + 5(x_2 - x_1^2)^2$

$x^1 = (2,0)^T$,初始步长$\delta = 0.5$，加速因子$\alpha = 1$，缩减率$\beta = 0.5$，坐标方向$e_1 = (1,0)^T, e_2 = (0,1)^T$

- **例5_2**
  - 用步长加速法求$\min f(x) = (1 - x_1)^2 + 5(x_2 - x_1^2)^2$

$x^1 = (2,0)^T$,初始步长$\delta = 0.5$，加速因子$\alpha = 1$，缩减率$\beta = 0.5$，坐标方向$e_1 = (1,0)^T, e_2 = (0,1)^T$

解：先在x^1周围进行探测移动，令 $y^1 = (2,0)^T$,探测情况如下：

$$f(y^1) = 81, \qquad y^1 + \delta e_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{5}{2} \\ 0 \end{bmatrix}$$

- $f\left(y^1 + \delta e_1\right) = 197\frac{9}{16} > f(y^1)$,故失败。
- $y^1 - \delta e_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ 0 \end{bmatrix}$
- $f\left(y^1 - \delta e_1\right) = 25\frac{9}{16} < f(y^1)$,成功。

- **例5_2**
  - 因此，令$\boldsymbol{y}^2=\boldsymbol{y}^1-\boldsymbol{\delta}\boldsymbol{e}_1=\begin{bmatrix}\frac{3}{2}\\0\end{bmatrix}$，从$\boldsymbol{y}^2$出发，沿$\boldsymbol{e}_2$探测情况如下：
  - $$\boldsymbol{y}^2+\boldsymbol{\delta}\boldsymbol{e}_2=\begin{bmatrix}\frac{3}{2}\\0\end{bmatrix}+\frac{1}{2}\begin{bmatrix}0\\1\end{bmatrix}=\begin{bmatrix}\frac{3}{2}\\\frac{1}{2}\end{bmatrix},$$
  - $f\left(\boldsymbol{y}^2+\boldsymbol{\delta}\boldsymbol{e}_2\right)=15\frac{9}{16}<f(\boldsymbol{y}^2)$,成功。
  - 因此，令$\boldsymbol{y}^3=\boldsymbol{y}^2+\boldsymbol{\delta}\boldsymbol{e}_2=\begin{bmatrix}\frac{3}{2}\\\frac{1}{2}\end{bmatrix}$.

- **例5_2**
  - 第一轮探测完成后，由于$f(y^3) < f(y^1)$,因此得到第2个基点

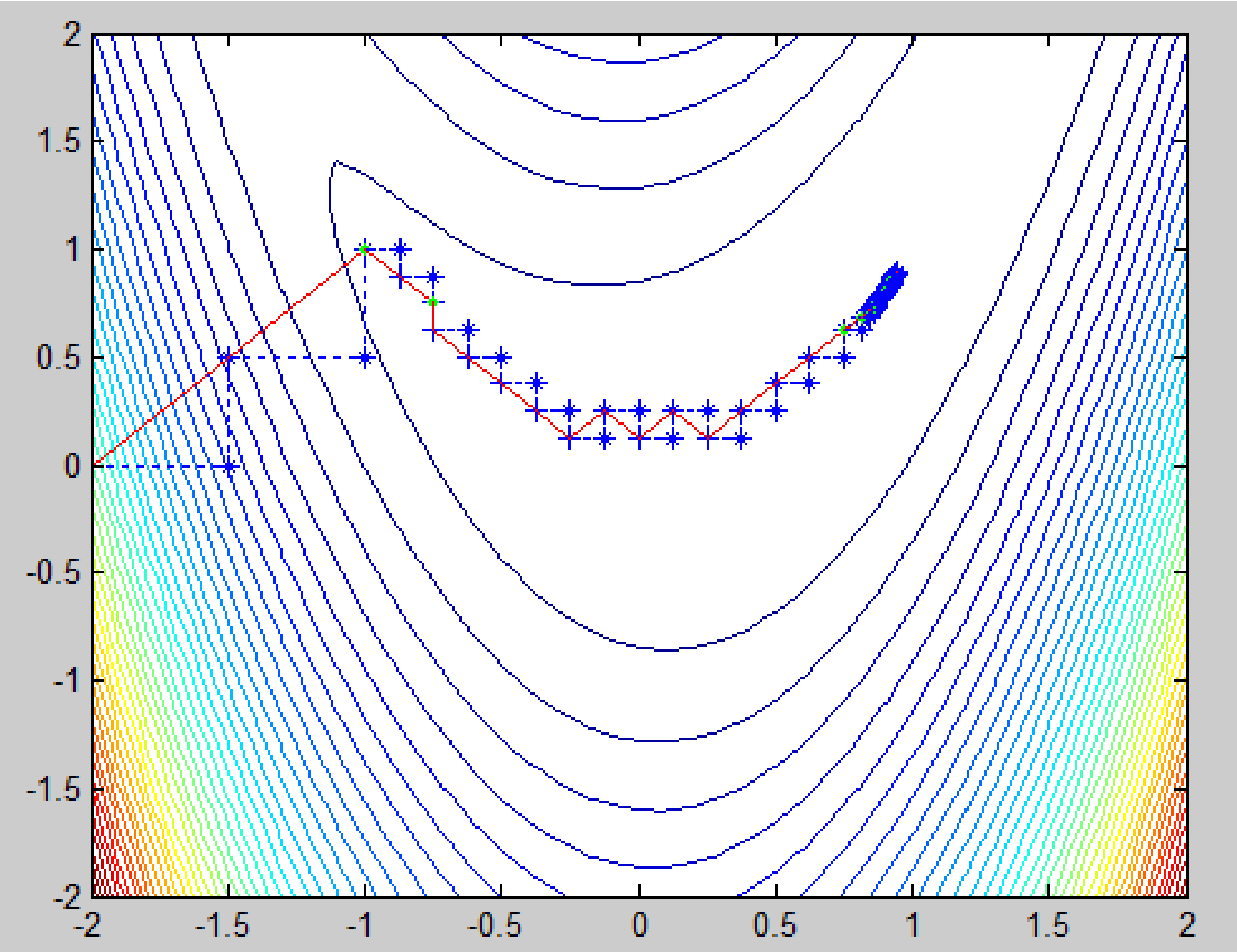
$$x^2 = y^3 = \begin{bmatrix} \frac{3}{2} \\ \frac{1}{2} \end{bmatrix}$$,再沿方向$x^2 - x^1$进行模式移动，令：

  - $$y^1 = x^2 + \alpha\left(x^2 - x^1\right) = 2x^2 - x^1 = 2\begin{bmatrix} \frac{3}{12} \\ \frac{1}{2} \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
  - 模式移动后，立即从得到的点 y^1 出发，进行第2轮探测移动。探测情况如下：

- **例5_2**
  - 先沿$\boldsymbol{e_1}$探测，这时有：
  - $f(y^1)=0,\ y^1+\delta e_1=\begin{bmatrix}1\\1\end{bmatrix}+\frac{1}{2}\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}\frac{3}{2}\\1\end{bmatrix},$
  - $f(y^1+\delta e_1)=8\frac{1}{16}>f(y^1)$,失败
  - $y^1-\delta e_1=\begin{bmatrix}1\\1\end{bmatrix}-\frac{1}{2}\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}\frac{1}{2}\\0\end{bmatrix},$
  - $f(y^1-\delta e_1)=3\frac{1}{16}>f(y^1)$,失败；
  - 由于沿$\boldsymbol{e_1}$的正反向探测均失败，故令$y^2=y^1=\begin{bmatrix}1\\1\end{bmatrix}$

- **例5_2**
  - 从$\boldsymbol{y}^2$出发沿$\boldsymbol{e}_2$探测的情况是：
  - $f\left(\boldsymbol{y}^2+\delta\boldsymbol{e}_2\right)=\frac{5}{4}>f\left(\boldsymbol{y}^2\right)$,失败
  - $f\left(\boldsymbol{y}^2-\delta\boldsymbol{e}_2\right)=\frac{5}{4}>f\left(\boldsymbol{y}^2\right)$,失败；
  - 由于沿$\boldsymbol{e}_2$的正反向探测均失败，故令$\boldsymbol{y}^3=\boldsymbol{y}^2=\begin{bmatrix}1\\1\end{bmatrix}$
  - 比较在$\boldsymbol{y}^3$和基点 $\boldsymbol{x}^2$处的函数值，由于$f\left(\boldsymbol{y}^3\right)=0<f\left(\boldsymbol{x}^2\right)=15\frac{9}{16}$,表明此次模式移动是成功的，因此得到新的基点：$\boldsymbol{x}^3=\boldsymbol{y}^3=\begin{bmatrix}1\\1\end{bmatrix}$

- **例5_2**
  - 从x^3出发沿方向$x^3 - x^2$进行模式移动。令：
  - $y^1 = x^3 + a(x^3 - x^2) = 2x^3 - x^2 = (\frac{1}{2}, \frac{3}{2})^T$
  - 然后，从y^1出发，进行探测移动。做下去就会发现此次模式移动仍失败。因此退回到基点x^3。减小步长，令：$\delta = \beta\delta = \frac{1}{4}$。
  - 再从$y^1 = x^3$开始，依次沿e_1, e_2探测。我们还会发现，在x^3周围的探测移动也是失败的，必须继续缩减步长。继续下去，必能得出结论，x^3是局部最优解。
  - 事实上，用解析方法容易求得x^3是此问题的最优解。


2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2