最优化理论
第五章：无约束最优化方法

人工智能学院
智能感知与图像理解教育部重点实验室

无约束最优化方法

1 坐标转换法

2 单形替换法

3 无约束优化算法总结

无约束最优化方法

1 坐标转换法

2 单形替换法

3 无约束优化算法总结

坐标轮换法

一、坐标轮换法简介

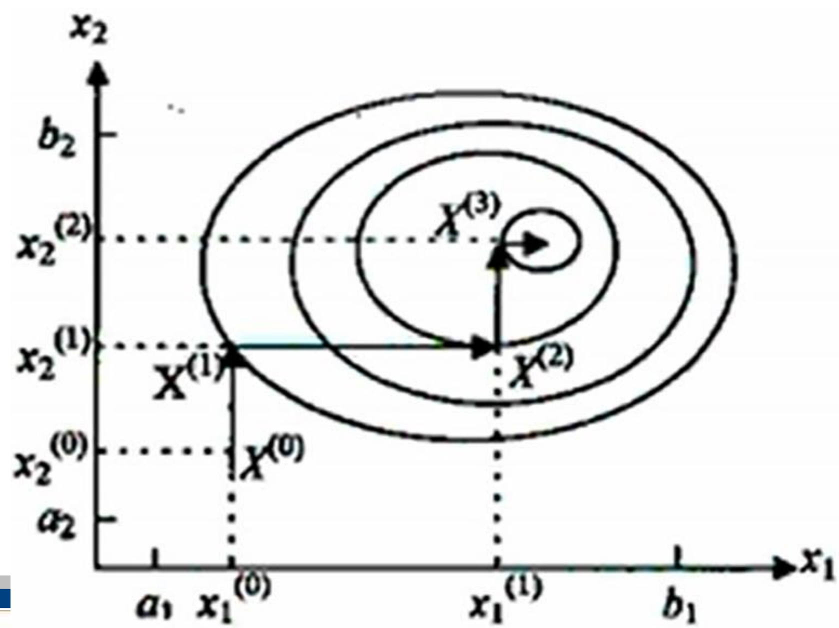
坐标轮换法也叫变量轮换法，是一种经典的直接优化法，由D’esopo于1959年提出，思想非常简单，将**多变量**的优化问题轮流地转化成**单变量**的优化问题，即依次沿各个**坐标方向**上作一维搜索，且每次搜索都以前一次搜索的最好点作为起点。

二、坐标轮换法基本原理

下面以二元函数为例来进行分析，然后给出n元函数的一般算法。

设某二元函数$S=f(X)=f(x_1,x_2)$，其等高线见下图。设极值存在的区间为

$a_1\le x_1\le b_1$，$a_2\le x_2\le b_2$，其中$e_1=(1,0)^T$，$e_2=(0,1)^T$分别表示x_1，x_2的坐标方向。



二、坐标轮换法基本原理

二元函数的搜索过程如下：

1. 设从$X^{(0)}=(x_1^{(0)},x_2^{(0)})$出发，先固定$x_1=x_1^{(0)}$不变，求以$x_2$为单变量的目标函数

$$f(x_2)=f(x_2^{(0)}+\lambda_2)=f(\lambda_2)$$

的极值，即利用一维搜索法求以λ_2为单变量的目标函数极值，从而可出最佳步长因子λ_2^1，得

$$X^{(1)}=(x_1^{(0)},x_2^{(0)}+\lambda_2^1)$$

$$S^{(1)}=f(X^{(1)})$$

二、坐标轮换法基本原理

二元函数的搜索过程如下：

2. 固定$x_2=x_2^{(1)}=x_2^{(0)}+\lambda_2^{1}$不变，求以$x_1$为单变量的目标函数$f(x_1)=f(x_1^{(0)}+\lambda_1)=f(\lambda_1)$的极值，即求以$\lambda_1$为单变量的目标函数极值，从而可定出最佳步长$\lambda_1^{1}$，得

$$X^{(2)}=(x_1^{(0)}+\lambda_1^{1},x_2^{(0)}+\lambda_2^{1})=f(x_1^{(1)},x_2^{(1)})\quad S^{(2)}=f(X^{(2)})$$

3. 因为$S^{(2)}$优于$S^{(1)}$，因此在下一轮坐标轮换搜索时搜索区间缩短为$x_1^{(1)}\le x_1\le b_1$，$x_2^{(1)}\le x_2\le b_2$。在新的区间中重复以上步骤，每次搜索不仅可使目标函数值得到改善，而且可以将所搜区间进一步缩短，直到达到给定精度为止。

二、坐标轮换法基本原理 搜索方向与步长

n元函数的搜索过程如下：

给定初始点x_0, 从x_0出发，依次沿坐标轴方向

e_i=(0,…,0,1,0,…,0)^T， $i=1,2,\dots,n$,

第i 个分量

进行最优一维搜索,即

$$\boxed{\begin{aligned} &f(x_k^{i-1}+t_k^i e_i)=\min_t f(x_k^{i-1}+te_i)\\ &x_k^i=x_k^{i-1}+t_k^i e_i, \quad i=1,\dots,n. \end{aligned}}$$

目标函数较复杂时，最佳步长可用0.618法或二次插值法

终止准则 设给定的允许误差$\varepsilon>0$,若$\left|X_k^n-X_{k-1}^n\right|\le\varepsilon$,则迭代终止，输出最优解$X^*=X_k^n$

循环 否则令k=k+1，从点 X_k^n 出发，再依次沿坐标轴方向e_i 进行最优一维搜索.

坐标轮换法

三、坐标轮换法算法步骤

- 已知目标函数 $f(X)$，终止限 $\varepsilon > 0.$
- （1）任选取始点 $X^{(0)}$作为第一轮循环的初始点 $\varepsilon > 0.$
- （2）置搜索方向依次为 $e_1 = [1, 0,0, \cdots, 0]^T$

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

$$\cdots$$

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

- （3）按下式求最优步长并进行迭代计算

$$f(X_k^{(i-1)} + t_k^{(i)} e_i) = \min_t f(X_k^{(i-1)} + te_i)$$

$$X_k^{(i)} = X_k^{(i-1)} + t_k^{(i)} e_i$$

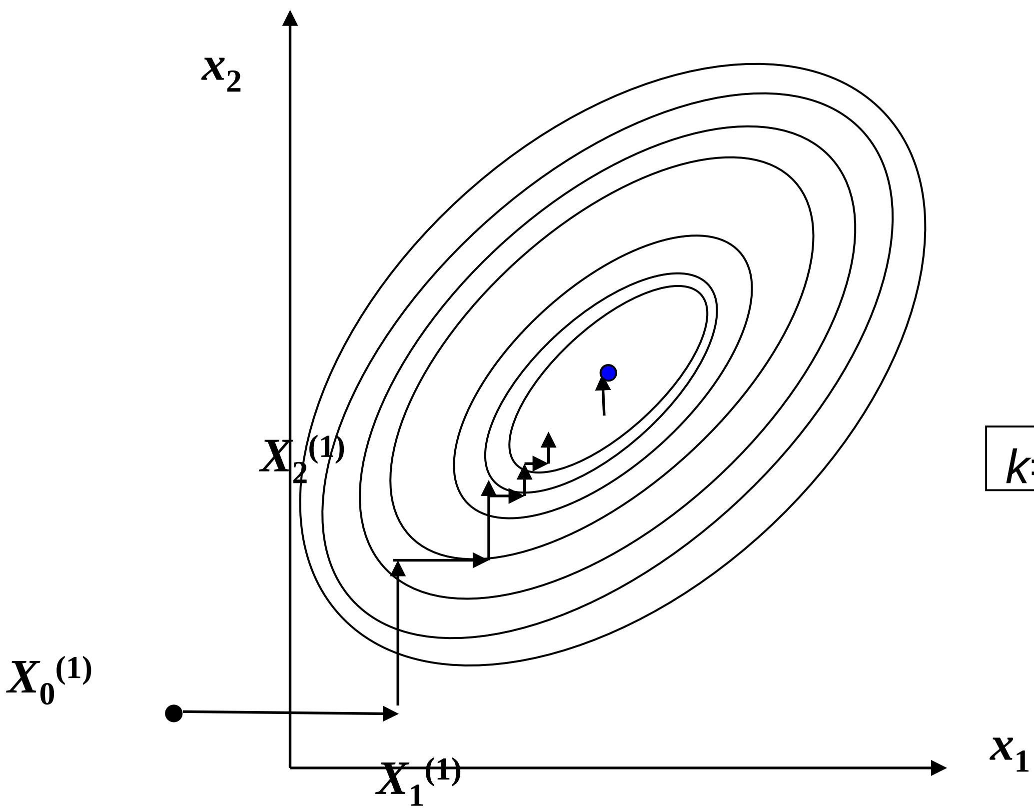
坐标轮换法

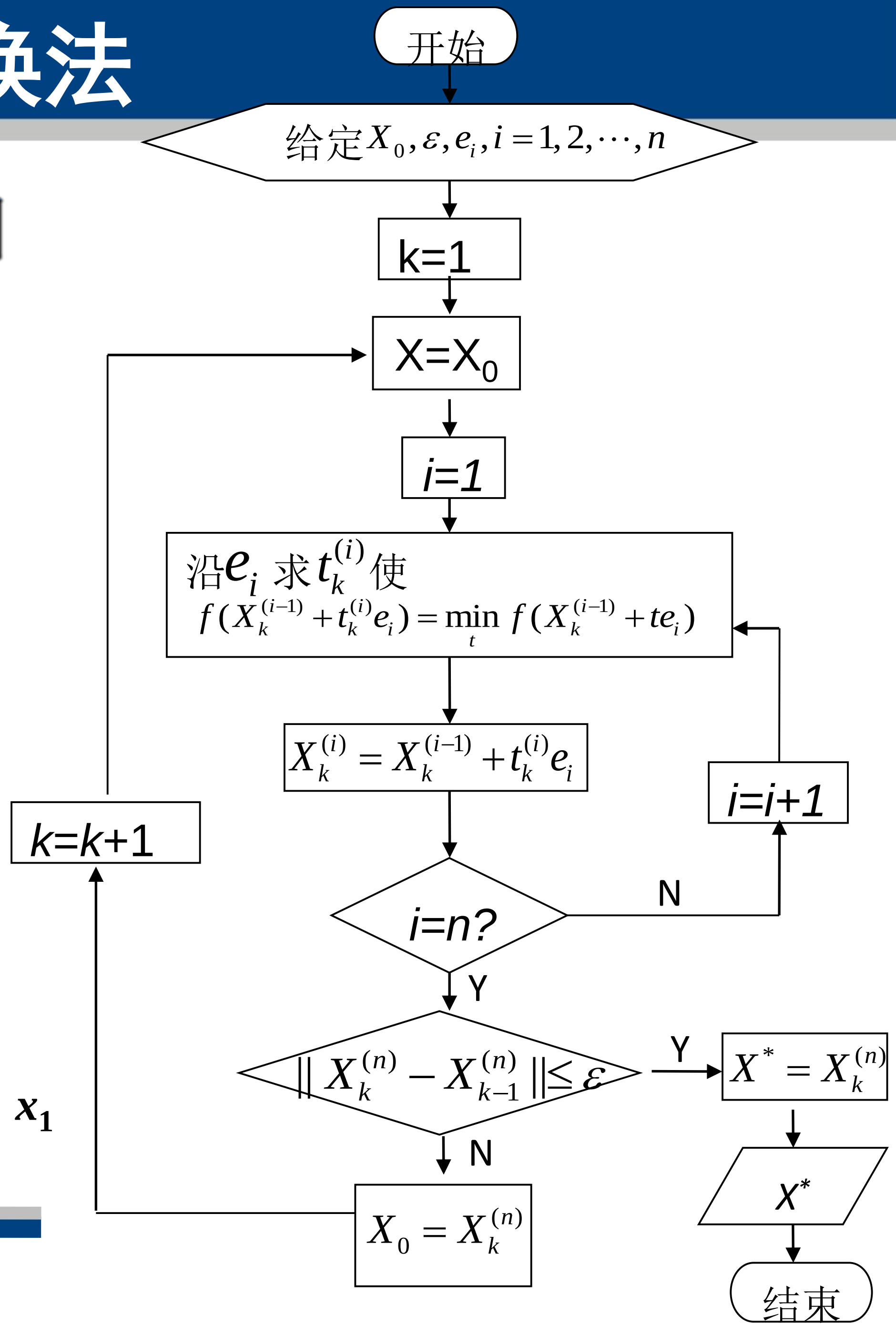
上式中，k 为循环次数，$k=1,2,\cdots;i$ 为该循环中一维搜索的序号，$i=1,2,\cdots,n;t_k^{(i)}$ 为利用一维搜索求出的最优步长.

- （4）如果 $i=n$，即转（5）;如果 $i<n$，则转（3）.
- （5）收敛性准则 $:\|X_k^{(n)}-X_{k-1}^{(n)}\|\le\varepsilon$，若满足判别式，即停止迭代，输出最优解 $X^*=X_k^{(n)}$ 及 $f^*=f(X^*)$;
若不满足，则令 $k=k+1$，转（3）.

坐标轮换法

四、坐标轮换法算法流程图





五、坐标轮换法实例

例1：用坐标轮换法求 $\min f(X)=\left(x_1-2\right)^4+\left(x_1-2x_2\right)^2$，初始点 $X_0=[0,3]^T, \varepsilon=0.04.$

解：从初始点出发，依次沿方向 $\mathrm{e}_1=(1,0)^T, \mathrm{e}_2=(0,1)^T$ 搜索，以第一步为例，从 X_0 出发，沿 e_1 方向搜索，求得 $X_1^{(1)}$ 点

$$f\left(X_0+te_1\right)=\left[\left(x_0^{(1)}+te_1\right)-2\right]^4+\left[\left(x_0^{(1)}+te_1\right)-2x_0^{(2)}\right]^2=(t-2)^4+(t-6)^2$$

令 $\dfrac{\mathrm{d}}{\mathrm{d}t} f\left(X_0+te_1\right)=0$，即 $\dfrac{\mathrm{d}}{\mathrm{d}t}[(t-2)^4+(t-6)^2]=0$

得 t=3.13，即取 $t_1^{(1)}$=3.13.于是有

$$X_1^{(1)}=X_0+t_1^{(1)}e_1=\begin{bmatrix}0\\3\end{bmatrix}+3.13\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}3.13\\3\end{bmatrix}$$

五、坐标轮换法实例

例1：用坐标轮换法求 $\min f(X)=(x_1-2)^4+(x_1-2x_2)^2$，初始点 $X_0=[0,3]^T, \varepsilon=0.04.$

再从 $X_1^{(1)}$ 出发，沿方向 e_2 搜索，求得 $X_1^{(2)}$ 点

$$f\left(X_1^{(1)}+te_2\right)=\left(x_1^{(1)}-2\right)^4+\left[x_1^{(1)}-2\left(x_1^{(2)}+te_2\right)\right]^2$$

$$=1.13^4+[3.13-2\times3-2t]^2$$

令 $\frac{\mathrm{d}}{\mathrm{d}t}f\left(X_1^{(1)}+te_2\right)=0$

可得 t=-1.44，即取 $t_1^{(2)}$=-1.44.于是有

$$X_1^{(2)}=X_1^{(1)}+t_1^{(2)}e_2=\begin{bmatrix}3.13\\3\end{bmatrix}-1.44\begin{bmatrix}0\\1\end{bmatrix}=\begin{bmatrix}3.13\\1.56\end{bmatrix}$$

五、坐标轮换法实例

例1：用坐标轮换法求$\min f(X)=\left(x_1-2\right)^4+\left(x_1-2x_2\right)^2$，初始点$X_0=[0,3]^T, \varepsilon=0.04$.

终止判别

$$\left\|X_1^{(2)}-X_0\right\|=\sqrt{(3.13-0)^2+(1.56-3)^2}=3.45>\varepsilon$$

因终止条件不满足，需继续迭代，取$X_0=X_1^{(2)}$，进行第二轮循环迭代，各轮迭代计算数据见表1，最优化解为$X^*=[2.22,1.12]^T$.

五、坐标轮换法实例

例1：用坐标轮换法求 $\min f(X)=(x_1-2)^4+(x_1-2x_2)^2$，初始点 $X_0=[0,3]^T,\varepsilon=0.04$.

表1

循环迭代序号	X_0	$e_1=[1,0]^T$		$e_2=[0,1]^T$		$f(X_k^{(2)})$	$\lVert X_k^{(2)}-X_0\rVert$	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

五、坐标轮换法有关说明

坐标轮换法的优点：算法简单，容易实现；

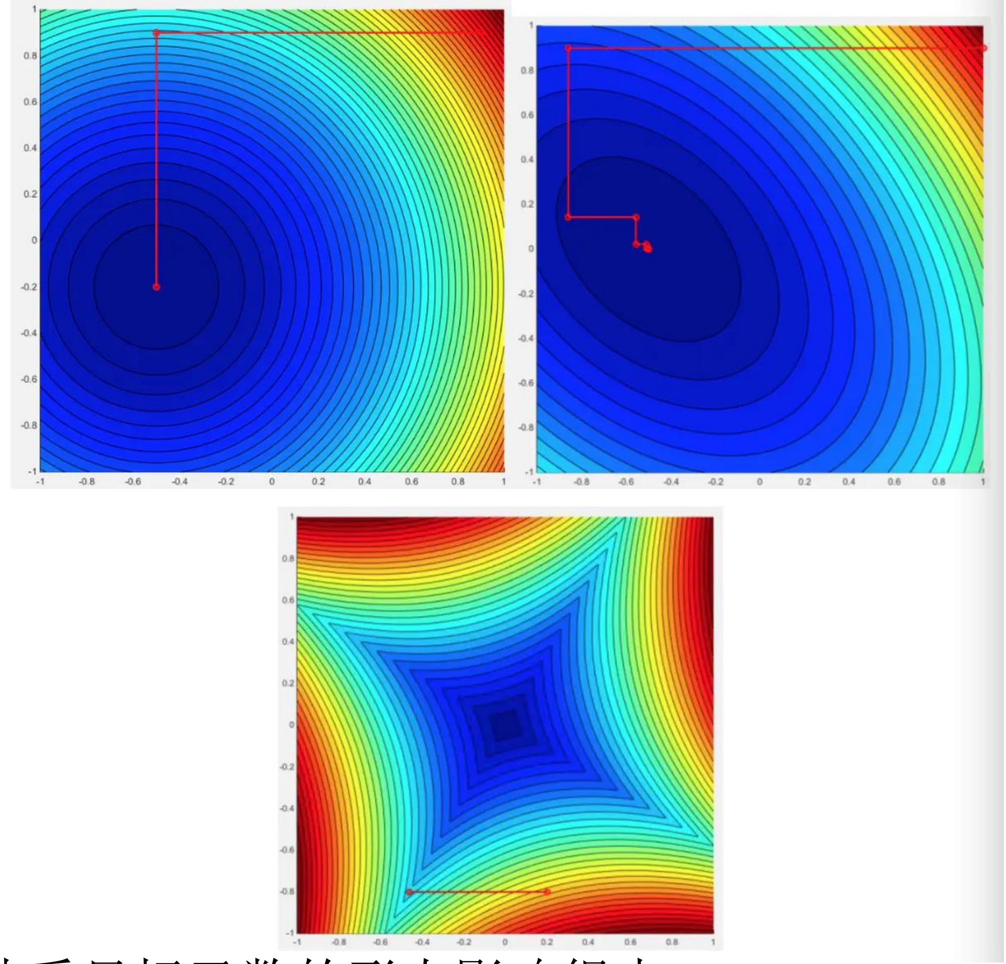
坐标轮换法的缺点：当维数增加时，效率明显下降，收敛慢，以振荡方式逼近最优解。只适用于变量数目少且目标函数中变量间的交互作用弱的情况。

注：若目标函数汇总没有自变量的乘积项，即一个变量的变化与其他变量无关，称之为无交互作用，这时该方法最有效。若目标函数中有自变量乘积项，则为弱交互作用，这时坐标轮换法的效率很低。

坐标轮换法

五、坐标轮换法有关说明

图（c）目标函数中出现了脊线，称为强交互作用，此时如果某一步搜索正好落在脊线上，则无论向哪个方向搜索，目标函数值均增大，搜索过程终止于脊线上，坐标轮换法无效。



坐标轮换法受目标函数的形态影响很大

（a）搜索有效，两次收敛；（b）搜索低效；（c）等值线出现脊线，搜索

无约束最优化方法

1 坐标转换法

2 单形替换法

3 无约束优化算法总结

单形替换法

一、单形替换法简介

单纯形替换法是Spendley、Hext 和Himsworth 于1962年提出;Nelder 和 Mead 1965年改进，也是一种不使用导数的求解无约束极小化问题的直接搜索方法，与前面几种方法不同的是，单纯形替换法不是利用搜索方向从一个点迭代到另一个更优的点，而是从一个单纯形迭代到另一个更优的单纯形。

二、单形替换法基本原理

•现以求二元函数的极小点为例,说明单纯形法形成原理.

• 设二元函数 $f(X)=f(x_1,x_2)$ 在 x_1-x_2 平面上取不在同一条直线上的三个点 X_1，X_2，X_3,并以它们为顶点构成一单纯形——三角形．算出各顶点的函数值 $f(X_1)$，$f(X_2)$，$f(X_3)$，比较其大小，现假定比较后有

$$f(X_1)>f(X_2)>f(X_3)$$

这说明点 X_1 最差，点 X_3 最好，点 X_2 次差.

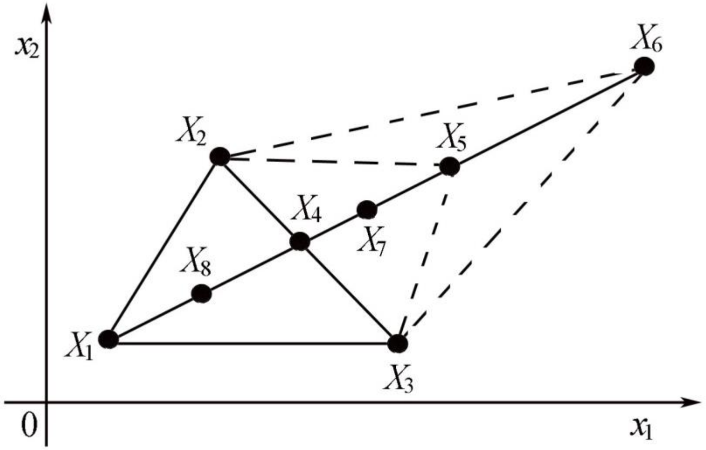
• 为了寻找极小点，一般来说应向最差点的反对称方向进行搜索.

单形替换法

•以 X_4 记为 X_2X_3 的中点（如图所示），在 X_1X_4 的延长线上取点 X_5，使

$$X_5 = X_4 + (X_4 - X_1) = 2X_4 - X_1$$

X_5 称为 X_1 关于 X_4 的反射点.



•算出 X_5 的函数值 $f(X_5)$，可能出现以下几种情形:

单形替换法

- (1) $f(X_5) < f(X_3)$

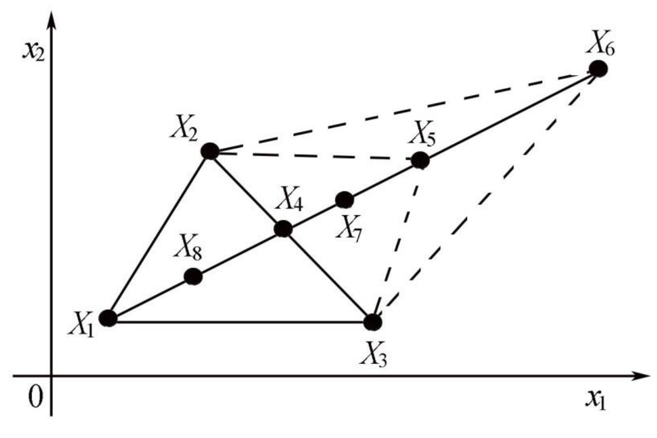
这说明搜索方向正确，可进一步扩大效果，继续沿X_1X_5向前搜索，也就是向前扩张．这时取

$$X_6 = X_4 + \alpha(X_4 - X_1)$$

其中α为扩张因子，一般取$\alpha = 1.2 \sim 2.0$.

如果$f(X_6) < f(X_5)$ ，说明扩张有利，就可以点X_6代替点X_1构成新的单纯形$\{X_2, X_3, X_6\}$.

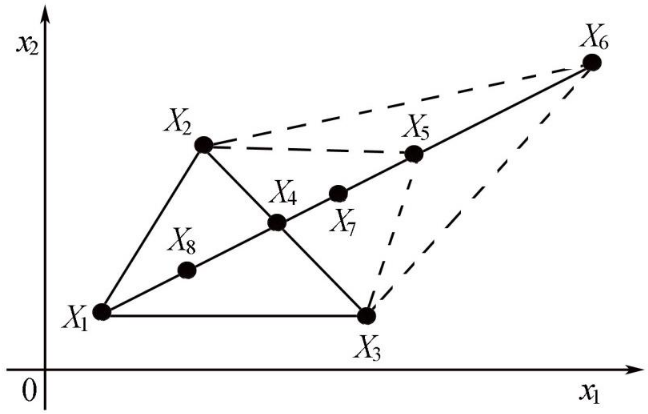
如果$f(X_6) > f(X_5)$，说明扩张不利，舍去点X_6，仍以点X_5代替X_1构成新的单纯形$\{X_2, X_3, X_5\}$



单形替换法

- (2) $f(X_3) < f(X_5) < f(X_2)$

  这说明搜索方向正确，但无须扩张，以X_5代替X_1构成新的单纯形$\{X_2, X_3, X_5\}$.



(3) $f(X_2) < f(X_5) < f(X_1)$

这表示X_5点走得太远，应缩回一些．若以β表示压缩因子，则有

$$X_7 = X_4 + \beta(X_5 - X_4) \qquad (5.35)$$

β 常取为0.5.以X_7代替x_1构成新的单纯形 $\{X_2, X_3, X_7\}$.

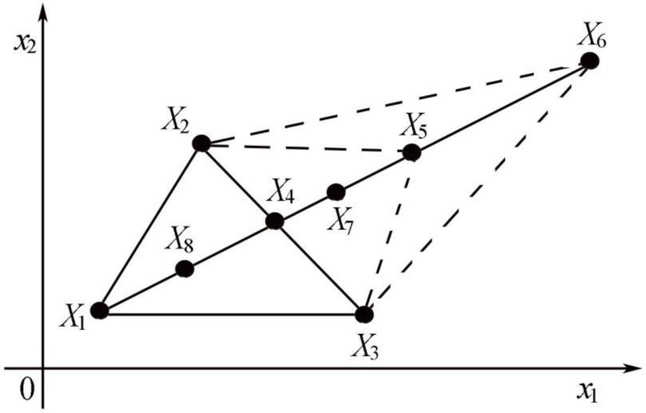
单形替换法

- (4) $f(X_5) > f(X_1)$

  这时应压缩更多一些，将新点压缩至X_1至X_4之间，令

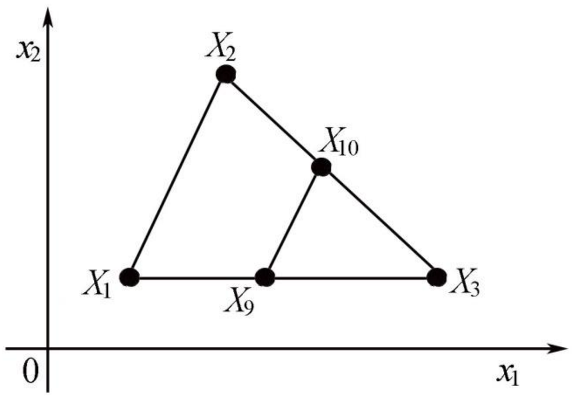
$$X_8 = X_4 - \beta(X_4 - X_1) = X_4 + \beta(X_1 - X_4) \qquad (5.36)$$

- 注意，将式（5.35）中的X_5代之以X_1，即可得式（5.36）．如果 $f(X_8) < f(X_1)$，则以X_8代替X_1 构成新的单纯形$\{X_2, X_3, X_8\}$，否则认为X_1X_4 方向上所有点的函数值$f(X)$都大于$f(X_1)$，不能沿此方向搜索.



单形替换法

- 这时，可以以X_3为中心进行缩边.若使顶点X_1和X_2向X_3移近一半距离，得新单纯形$\{X_3, X_9, X_{10}\}$,以此单纯形为基础再进行寻优.



单形替换法

- 以上说明，不管哪种情况，我们都可以得到一个新的单纯形，其中至少有一顶点的函数值比原单纯形更小．如此继续，直至满足收敛终止准则．
- 在n维情况下，一个单纯形含有 $n+1$ 个顶点，计算工作量较大，但原理和上述二维情况相同．

三、单纯形法迭代步骤

- 已知设 X 为n维变量,目标函数为$f(X)$, 终止限为 $\varepsilon_1>0, \varepsilon_2>0$
- (1) 构成初始单纯形

在n维空间中选初始点X_0（离最优点越近越好）,从X_0 出发，沿各坐标方向以步长t得n个顶点 $X_i(i=1,2,\cdots,n)$.

这样选择顶点可保证向量组 $X_1-X_0, X_2-X_0, \cdots, X_n-X_0$ 线性无关．否则，就会使搜索范围局限在较低维的空间内，有可能找不到极小点．当然，在各坐标方向可以走不同的距离.

- 步长t的范围可取0.5~15.0，开始时常取 $t=1.5\sim2.0$，接近最优点时要减小，例如取0.5~1.0.

单形替换法

- (2) 计算各顶点的函数值

$$y_i = f(X_i),\quad i = 0, 1, 2, \cdots, n$$

比较各函数值的大小，确定最好点X_L、最差点X_H和次差点X_G，即

$$y_L = f(X_L) = \min_i y_i, i = 0, 1, 2, \cdots, n$$

$$y_H = f(X_H) = \max_i y_i, i = 0, 1, 2, \cdots, n$$

$$y_G = f(X_G) = \max_i y_i,\quad i = 0, 1, 2, \cdots, H-1, H+1, \cdots, n$$

单形替换法

- (3) 计算 X_H 之外各点的“重心” X_{n+1}，

$$X_{n+1}=\frac{1}{n}\sum_{i=0,i\neq H}^{n}X_i$$

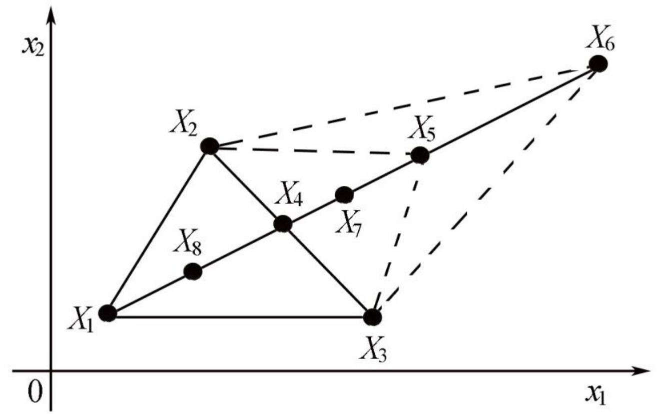
求出反射点

$$X_{n+2}=2X_{n+1}-X_H$$

- (4) 当 $f(X_{n+2})<f(X_L)$ 时，需要扩张，令

$$X_{n+3}=X_{n+1}+\alpha(X_{n+2}-X_{n-1})$$

如果 $f(X_{n+3})<f(X_{n+2})$，则以 X_{n+3} 代替 X_H 形成一新单纯形；否则，以代 X_{n+2} 替 X_H 构成新单纯形．然后转（8）．



单形替换法

- (5)当$f(X_L)\le f(X_{n+2})<f(X_G)$时，以$X_{n+2}$代替$X_H$构成新单纯形，然后转(8).
- (6)当$f(X_G)\le f(X_{n+2})<f(X_H)$时，则需要收缩．即令

$$X_{n+4}=X_{n+1}+\beta(X_{n+2}-X_{n+1}) \qquad (5.37)$$

以X_{n+4}代替X_H得新单纯形，并转(8).

- (7) 当$f(X_{n+2})\ge f(X_H)$时，令

$$X_{n+5}=X_{n+1}+\beta(X_H-X_{n+1})$$

如果$f(X_{n+5})\ge f(X_H)$, 则将单纯形缩边，可将向量X_i-X_L的长度缩小一半，即

$$X_i=X_L+\frac{1}{2}(X_i-X_L)=\frac{1}{2}(X_i+X_L),\ i=0,1,2,\cdots,n,\ i\ne L.$$

这样可得一新单纯形．否则就以X_{n+5}代替X_H得新单纯形．然后转(8).

单形替换法

- (8) 收敛性检验，每次迭代得到新单纯形后，即应进行收敛性检验，如满足收敛指标，则迭代停止，X_L 即为所求的近似解．否则，继续进行迭代计算． 通常所用的收敛准则是

$$\sum_{i=0,i\neq L}^{n}[f(X_i)-f(X_L)]^2 \leq \varepsilon_1$$

或

$$\left|\frac{f(X_H)-f(X_L)}{f(X_L)}\right| \leq \varepsilon_2$$

式中 ε_1 和 ε_2 为预先给定的允许误差．

四、单形替换法实例

例3. 试用单形替换法求$f(X)=4(x_1-5)^2+(x_2-6)^2$的极小值。

解： 选$X_0=[8,9]^T$，并取$X_1=[10,11]^T$，$X_2=[8,11]^T$.这三点不在一条直线上，用它们作为初始单纯形的顶点，如下图所示。然后计算各定点的函数值：

$f(X_0)=45$，$f(X_1)=125$，$f(X_2)=65$，可知X_1为最差点，X_0为最好点。

以X_3表示X_0和X_2的重心，则

$$X_3=\frac{1}{n}\left(\sum_{i=0,i\neq H}^{n}X_i\right)=\frac{1}{2}\left(\begin{bmatrix}8\\9\end{bmatrix}+\begin{bmatrix}8\\11\end{bmatrix}\right)=\begin{bmatrix}8\\10\end{bmatrix}$$

反射点

$$X_4=2X_3-X_1=2\begin{bmatrix}8\\10\end{bmatrix}-\begin{bmatrix}10\\11\end{bmatrix}=\begin{bmatrix}6\\9\end{bmatrix}$$

$$f(X_4)=4(6-5)^2+(9-6)^2=13$$

单形替换法

例3. 试用单形替换法求$f(X)=4(x_1-5)^2+(x_2-6)^2$的极小值。

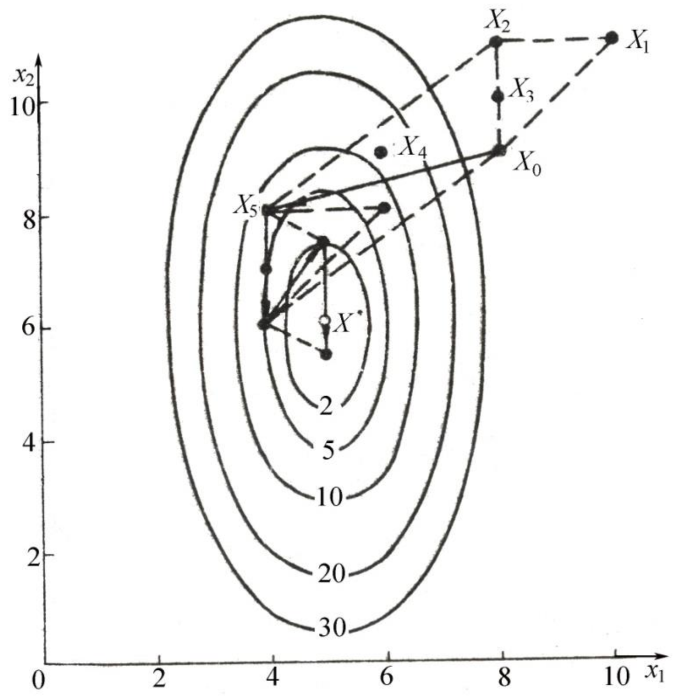
由于$f(X_4)<f(X_0)$，故需扩张，取α=2，则

$$X_5=X_3+2(X_4-X_3)=\begin{bmatrix}8\\10\end{bmatrix}+2\left(\begin{bmatrix}6\\9\end{bmatrix}-\begin{bmatrix}8\\10\end{bmatrix}\right)=\begin{bmatrix}4\\8\end{bmatrix}$$

$$f(X_5)=4(4-5)^2+(8-6)^2=8$$

因为$f(X_5)<f(X_4)$，故以X_5代替X_1，由X_5，X_0和X_2构成新单纯形，然后进行下一轮循环。该问题得最优解为X^*=$[5,6]^T$，$f(X^*)=0$.经32次循环，可把目标函数$f(X)$减少到1×10^{-6}。下图给出了前几次迭代得情形。

单形替换法



• 五、单纯形法有关说明

本算法上机占用内存很少，对变量不多且精度要求不高的问题此法很方便，但当变量个数多于10以上，此法就显得不十分有效.

无约束最优化方法

1 坐标转换法

2 单形替换法

3 无约束优化算法总结

无约束优化方法——间接法总结

1、梯度法

方向　　负梯度　　用到一阶导数

适合于精度不高或用于复杂函数寻找一个好的初始点

2、牛顿法

用到一阶导数和Hesse矩阵，具有二次收敛性

要求Hesse矩阵非奇异，且维数不宜太高

3、共轭梯度法

用到一阶导数，具有二次收敛性

4、变尺度法

收敛快，效果好，被认为是目前最有效的无约束优化方法。适用于维数较高，具有一阶偏导数的目标函数

无约束优化方法——直接法总结

1、坐标轮换法

计算效率较低

适合维数较低，目标函数无导数或导数较难求得

2、单形替换法

算法直观，收敛慢

第四次作业:

习题五（ P119 ）

第7，8题

西安电子科技大学
XIDIAN UNIVERSITY

智能感知与图像理解

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education

THE END

Thanks for your participation!