XIDIAN UNIVERSITY

最优化理论
知识点复习和答疑

人工智能学院
智能感知与图像理解实验室

无约束最优化方法

1. 知识点复习
2. 答疑



无约束最优化方法

1 知识点复习

2 答疑



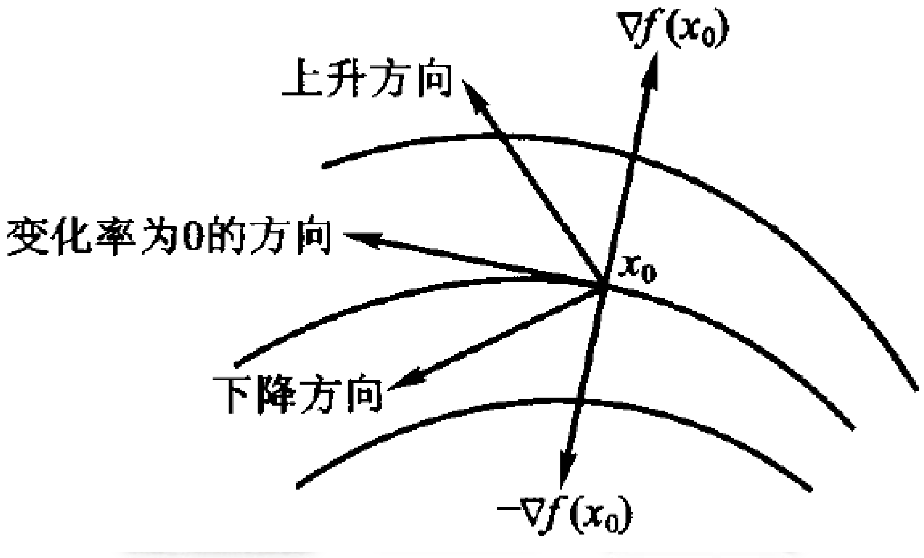
正定矩阵的判定<->函数的凹凸性

判定定理1：A正定的充要条件为A的特征值都大于零

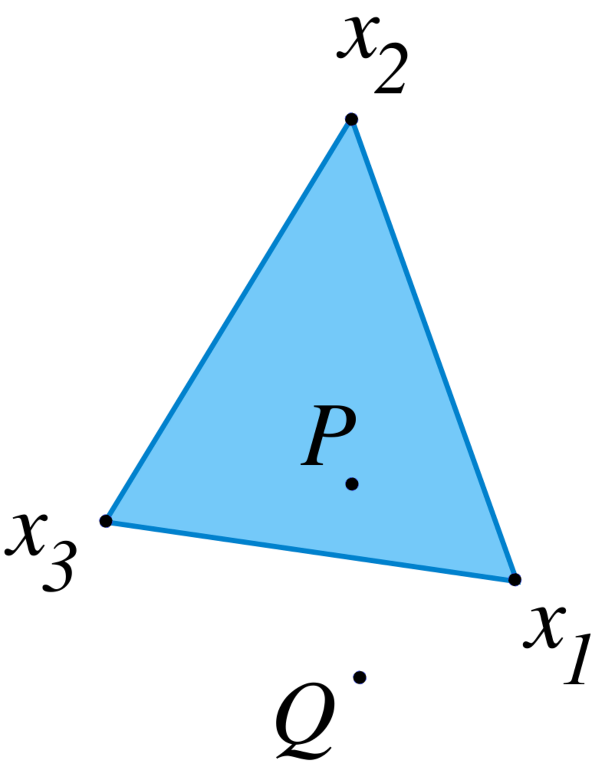
判定定理2：A正定的充要条件为A的所有顺序主子式都大于零

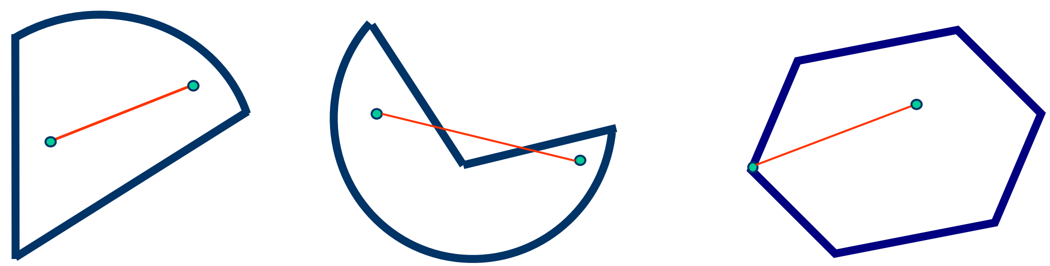
梯度与方向导数的结论

1. 梯度方向是函数值的最速上升方向；
2. 函数在与其梯度正交的方向上变化率为零；
3. 函数在与其梯度成锐角的方向上是上升的，
   而在与其梯度成钝角的方向上是下降的；
4. 梯度反方向是函数值的最速下降方向。



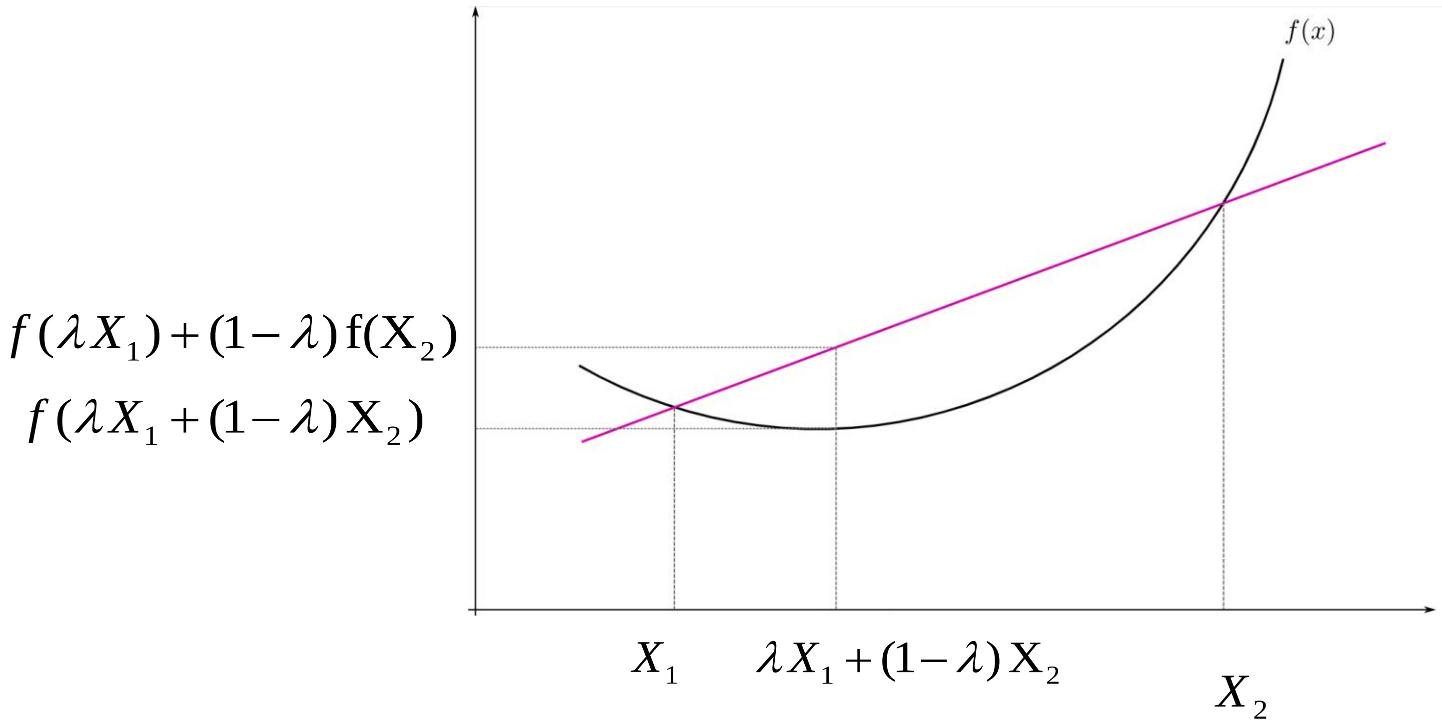
凸组合，凸集的定义





凸函数的定义及其判定

（Hesse矩阵半正定<->凸函数）



线性规划的任意形式转化为标准形式

$$\min f(X)=\sum_{j=1}^{n} c_j x_j$$

$$s.t.\begin{cases} \sum_{j=1}^{n} a_{ij} x_j = b_i, \ i=1,2,\ldots,m \\ x_j \geq 0, \ j=1,2,\ldots,n \end{cases}$$



单纯型法以及表格型单纯型法的运用

1 确定初始的基本可行解
2. 判断现行的基本可行解是否最优
3. 基本可行解的改进 ——基变换
4. 用初等变换求改进了的基本可行解——旋转运算



单纯型法以及表格型单纯型法的运用

C			C_B	C_N	θ
C_B	X_B	b	$x_1 \quad x_2 \quad \cdots \quad x_m$	$x_{m+1} \quad x_{m+2} \quad \cdots \quad x_n$	
c_1	x_1	b_1			θ_1
c_2	x_2	b_2	I	N	θ_2
.		.			.
.	.	.			.
c_m	x_m	b_m			θ_m
Z		$C_B b$	0	$C_N - C_B N$	

线性对偶问题

定义 设原线性规划问题为 **则称下列线性规划问题**

$$\min\ f(x)=c_1x_1+c_2x_2+\cdots+c_nx_n$$

$$s.t\quad\begin{cases}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n\ge b_1\\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n\ge b_2\\ \cdots\\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n\ge b_m\\ x_j\ge 0\left(j=1,2,\cdots,n\right)\end{cases}$$

$$\max g(y)=b_1y_1+b_2y_2+\cdots+b_my_m$$

$$s.t\quad\begin{cases}a_{11}y_1+a_{21}y_2+\cdots+a_{m1}y_m\le c_1\\ a_{12}y_1+a_{22}y_2+\cdots+a_{m2}y_m\le c_2\\ \cdots\\ a_{1n}y_1+a_{2n}y_2+\cdots+a_{mn}y_m\le c_n\\ y_i\ge 0\left(i=1,2,\cdots,m\right)\end{cases}$$

上述对偶问题称为*对称型对偶问题*。

原问题简记为(P)，对偶问题简记为(D)



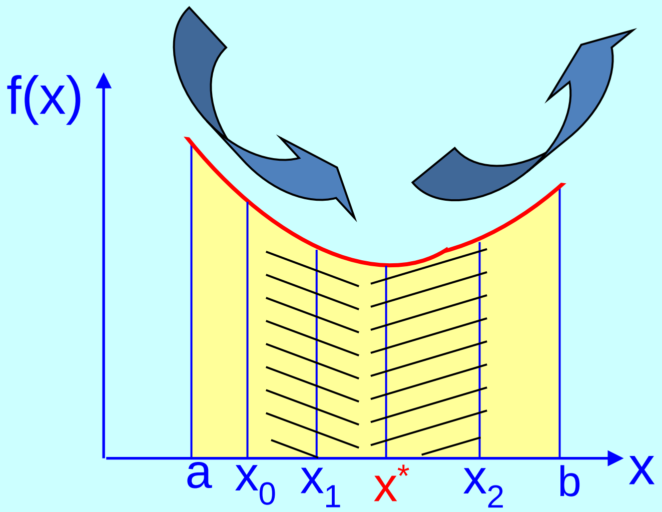
对偶关系对应表

原问题（对偶问题）	对偶问题（原问题）
min	max
目标函数系数	右边常数
右边常数	目标函数系数
约束条件系数矩阵	系数矩阵的转置
第i个约束条件为“$\geq$”型	第i个变量≥ 0
第i个约束条件为“$=$”型	第i个变量无限制
第i个约束条件为“$\leq$”型	第i个变量≤ 0
第j个变量≥ 0	第j个约束条件为“$\leq$”型
第j个变量无限制	第j个约束条件为“$=$”型
第j个变量≤ 0	第j个约束条件为“$\geq$”型

对偶问题之间的相互联系

对偶定理、互补松弛定理


搜索极小点所在区间：进退法



利用单谷（峰）函数性质求最优值：对分法

步骤1：计算 $x_0 = \dfrac{a+b}{2}$

步骤2：若 $f'(x_0) < 0$，令 $a = x_0$，转步骤3；

若 $f'(x_0) > 0$，令 $b = x_0$，转步骤3；

若 $f'(x_0) = 0$，停止, $x^* = x_0$。

步骤3：若 $|b-a| < \varepsilon$，则 $x^* = \dfrac{a+b}{2}$，停止，否则，转步骤1.

优点：计算量较少，而且总能收敛到一个局部极小点。

缺点：收敛速度较慢



Newton切线法

步骤1：给定初始点 $x_1 \in R, \varepsilon > 0$,令 $k = 1$。

步骤2：计算$f'(x_k),\ f''(x_k)$。

步骤3：若$\left|f'(x_k)\right| < \varepsilon$，停止，$x^* \approx x_k$，否则转步骤4。

步骤4：计算 $x_{k+1} = x_k - \dfrac{f'(x_k)}{f''(x_k)}$

令$k = k+1$，转步骤2。

特点：收敛速度快，局部二阶收敛。

缺点：须计算二阶导数，工作量大；对初始点要求高，要求初始点离极小点不太远，否则有可能使极小化发散或收敛到非极小点；局部收敛。

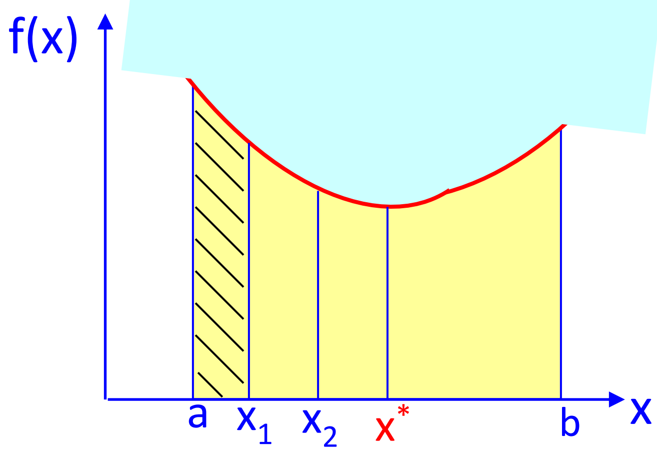

黄金分割法

定理： 设 f：$R \to R$ 在$[a, b]$上是单谷(峰)函数$a \le x_1 < x_2 \le b$ 那么

1° 若 $f(x_1) \ge f(x_2)$，则 $x^* \in [x_1, b]$,如左下图

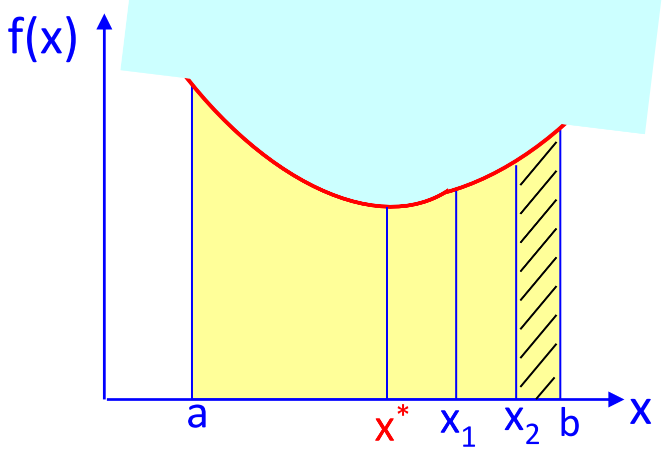
2° 若 $f(x_1) < f(x_2)$，则 $x^* \in [a, x_2]$, 如右下图

（I） 若$f(x_1) \ge f(x_2)$，$x^* \in [x_1, b]$



(I) 消去$[a, x_1]$

（II）若$f(x_1) < f(x_2)$, $x^* \in [a, x_2]$



(II) 消去$[x_2, b]$

抛物线插值法

1. 寻找满足如下条件的点（进退法寻找），成为两头大中间小的点：$x_1 < x_2 < x_3$，$f(x_1) > f(x_2), f(x_2) < f(x_3)$

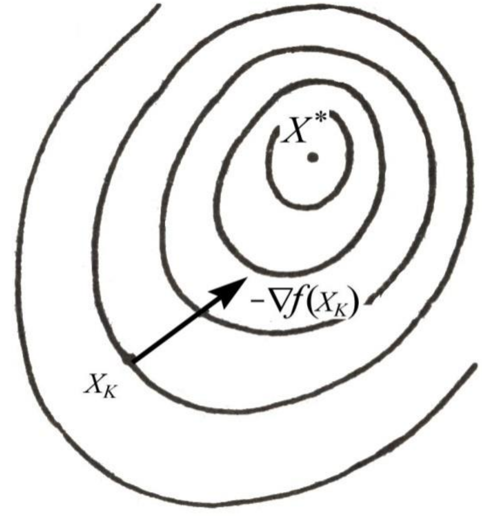
2. 两头大中间小，可得$a_2>0$，则 $\bar{x}$ 为$P(x)$的极小值点，且

$$\bar{x} \in [x_1, x_3]$$

3.若$|x_2 - \bar{x}| < \varepsilon$，则迭代结束，取 $x^* = \bar{x}$ ，否则在点 $x_1, x_2, x_3, \bar{x}$ 中，选取使$f(x)$ 最小的点作为新的x_2, 并使新的 x_1， x_3各是新的x_2近旁的左右两点，继续进行迭代，直到满足终止准则。

最速下降法基本原理

如图所示，假定我们 已经迭代了k 次获得了第k个迭代点 X_k ． 现在从X_k 出发，可选择的下降方向很多，一个非常自然的想法是沿最速下降方向（即负梯度方向）进行搜索应该是有利的，至少在 X_k 邻近的范围内是这样。因此，取搜索方向为 $P_k = -\nabla f(X_k)$



最速下降法迭代步骤

已知目标函数 $f(X)$ 及其梯度 $g(X)$，终止限 ε

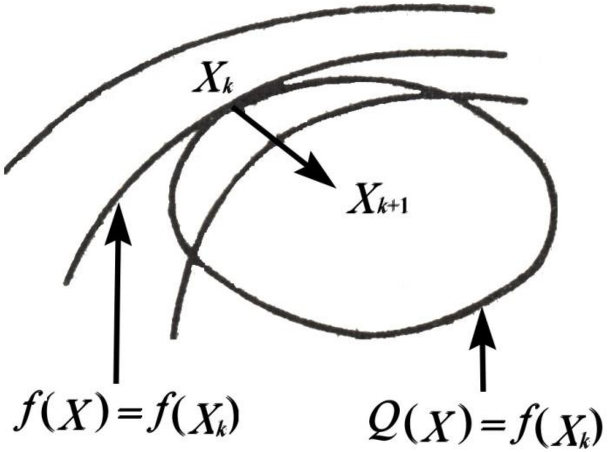
- （1）选定初始点 X_0，计算 $f_0 = f(X_0), g_0 = g(X_0)$. 置 $k = 0$.
- （2）作直线搜索：$X_{k+1} = ls(X_k, -g_k)$；计算

$$f_{k+1} = f(X_{k+1}),\ g_{k+1} = g(X_{k+1})$$

- （3）检测终止准则 $\|\nabla f(X_{k+1})\| \le \varepsilon$ 是否满足：若满足，则打印最优解 $X_{k+1}, f(X_{k+1})$，停止迭代；否则，置 $k = k+1$ 转(2).


Newton法基本原理

为寻求收敛速度快的算法，我们考虑在应用基本迭代公式 $X_{k+1}=X_k+t_kP_k$ 中，每轮迭代在迭代的起始点 X_k 处用一个适当的二次函数来近似该点处的目标函数，由此用点 X_k 指向近似二次函数极小点的方向来构造搜索方向 P_k (如图所示).



牛顿法迭代步骤

- 已知目标函数 $f(X)$ 及其梯度 $g(X)$, Hesse矩阵 $G(X)$,终止限 ε
- (1)选定初始点 X_0,计算 $f_0 = f(X_0), g_0 = g(X_0)$ 置 $k=0$.
- (2)计算 $G_k = \nabla^2 f(X_k)$.
- (3)由方程 $G_k P_k = -g_k$ 解出 P_k .
- (4)计算 $X_{k+1} = X_k + P_k,\ f_{k+1} = f(X_{k+1}), g_{k+1} = g(X_{k+1})$
- (5)判别终止准则是否满足：若满足，则打印最优解

  X_{k+1}, f_{k+1} 停机；否则置 $k=k+1$, 转(2).



修正Newton法迭代步骤

- 已知目标函数 $f(X)$ 及其梯度 $g(X)$, Hesse矩阵 $G(X)$, 终止限 ε.
- （1）选取初始点 X_0, 令 $k=0$.
- （2）求梯度向量．计算 $g(X_k)=\nabla f(X_k)$ ，若 $\|\nabla f(X_k)\|\le\varepsilon$，停止迭代输出 X_k．否则，转（3）．
- （3）构造Newton方向．计算 $G(X_k)^{-1}=[\nabla^2 f(X_k)]^{-1}$ ，取 $P_k=-[\nabla^2 f(X_k)]^{-1}\nabla f(X_k)$
- （4）进行一维搜索．求 t_k，使得

$$f(X_k+t_kP_k)=\min_{t\ge 0} f(X_k+tP_k)$$

令 $X_{k+1}=X_k+t_kP_k, k=k+1$ ，转（2）．



拟Newton法基本原理

- 人们希望寻找一种算法既可以保持Newton法收敛速度快的优点，又可以摆脱关于Hesse矩阵的计算，这就是本节要给大家介绍的拟牛顿法.

- 拟牛顿法是一种非常好的方法．其中**DFP算法**和**BFGS算法**，可以说直到目前为止在不用Hesse矩阵的方法中是最好的算法.



拟Newton法迭代步骤

- 已知目标函数 $f(X)$ 及其梯度$g(X)$,问题的维数 n,终止限 ε.
- (1)选定初始点．计算 $f_0 = f(X_0), g_0 = g(X_0)$.
- (2)置 $H_0 = I, P_0 = -g_0, k = 0$.
- (3)作直线搜索 $X_{k+1} = ls(X_k, P_k)$;

  计算 $f_{k+1} = f(X_{k+1}), g_{k+1} = g(X_{k+1})$.
- (4)判别终止准则是否满足:若满足,则打印 X_{k+1}, f_{k+1},

  停机；否则转(5).
- (5)若 $k = n$, 则置 $X_0 = X_{k+1}, f_0 = f_{k+1}, g_0 = g_{k+1}$, 转(2);否则,转(6).
- (6)计算 $S_k = X_{k+1} - X_k,\ y_k = g_{k+1} - g_k,$

$$H_{k+1} = H_k + \frac{S_k S_k^T}{S_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k},\quad P_{k+1} = -H_{k+1} g_{k+1}$$

  置 $k = k+1$, 转(3).



共轭方向法原理

构成各种不同最优化方法，往往取决于如何从基本迭代公式 $X_{k+1}=X_k+t_kP_k$ 中确定搜索方向P_k。

最速下降法：计算步骤简单，但 $P_k=-\nabla f(X_k)$ 导致搜索路线出现锯齿状，**收敛速度慢**。

Newton法和修正Newton法：$P_k=-[\nabla^2 f(X_k)]^{-1}\nabla f(X_k)$使得**收敛速度快**，但**计算计算量大且要求Hesse矩阵正定，导致算法对初始点选择要求严格。**

因此需要寻找一种好的算法，它的收敛速度介于最速下降法和牛顿法之间，对于正定二次函数只需迭代有限次就可达到极小点，收敛速度快同时计算简单的算法——**共轭方向法**。



FR共轭梯度法计算步骤

Step 1. 选定初始点x^1。

Step 2. 如果$\|g_1\| \le \varepsilon$，算法停止，$x^* = x^1$；否则转Step 3;

Step 3. 取 $p^1 = -g_1, k=1$;

Step 4. 精确一维搜索找最佳步长λ_k，令 $x^{k+1} = x^k + \lambda_k p^k$;

Step 5. 如果$\|g_{k+1}\| \le \varepsilon$，算法停止，$x^* = x^{k+1}$，否转Step 6;

Step 6. 如果$k=n$，令 $x^1 = x^{k+1}, p^1 = -g_{k+1}$，算法停止，$k=1$，转Step 4;

否则转步骤7;

Step 7. 计算 $\alpha_k = \dfrac{\|g_{k+1}\|^2}{\|g_k\|^2}$，$p^{k+1} = -g_{k+1} + \alpha_k p^k$，$k = k+1$，转Step 4。

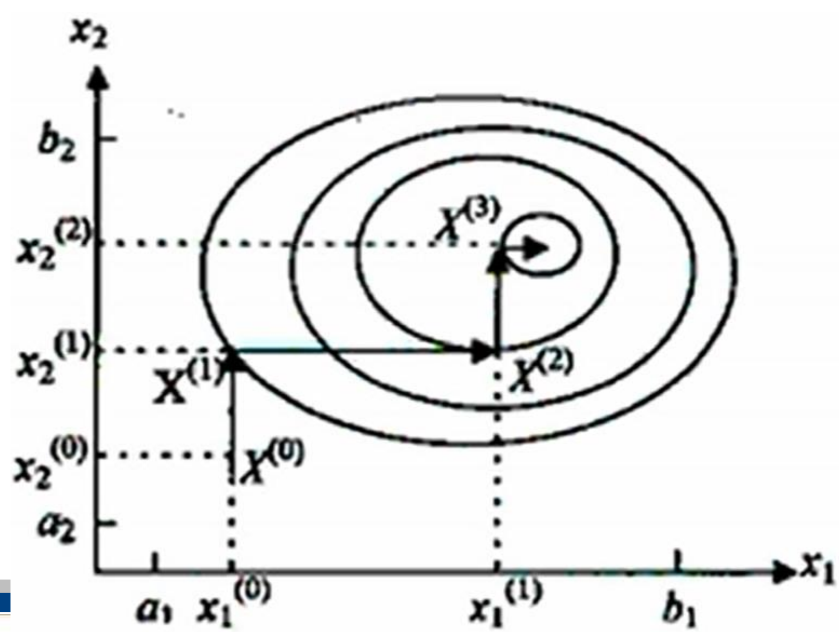


坐标轮换法基本原理

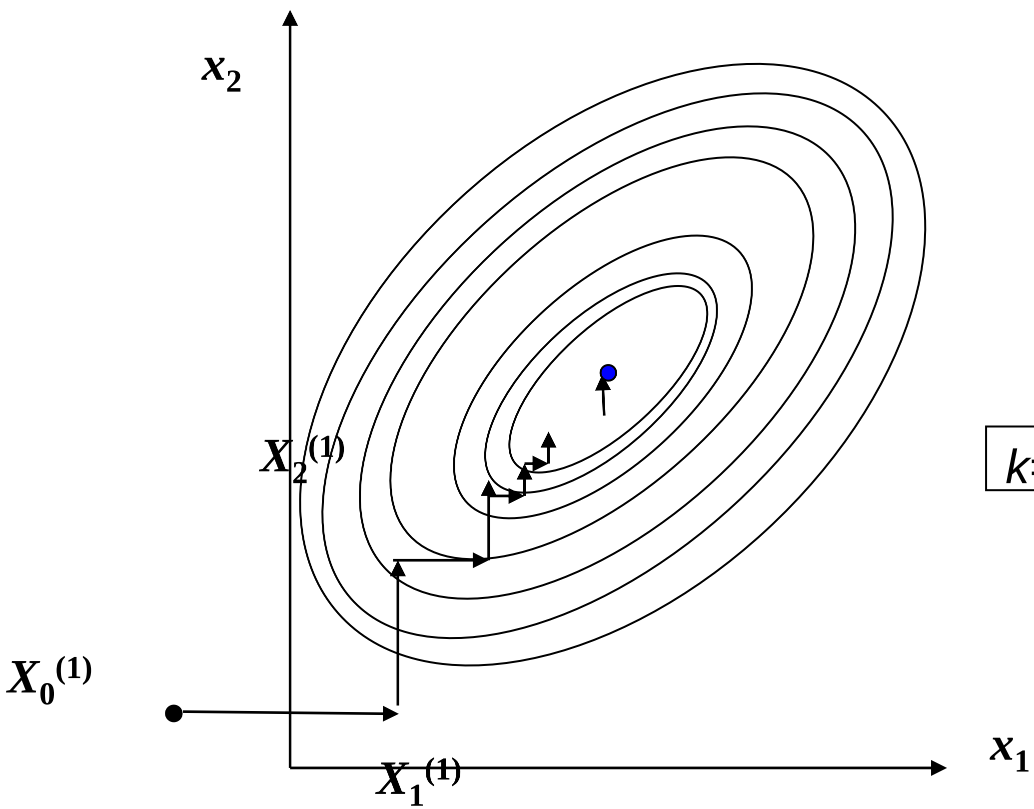
下面以二元函数为例来进行分析，然后给出n元函数的一般算法。

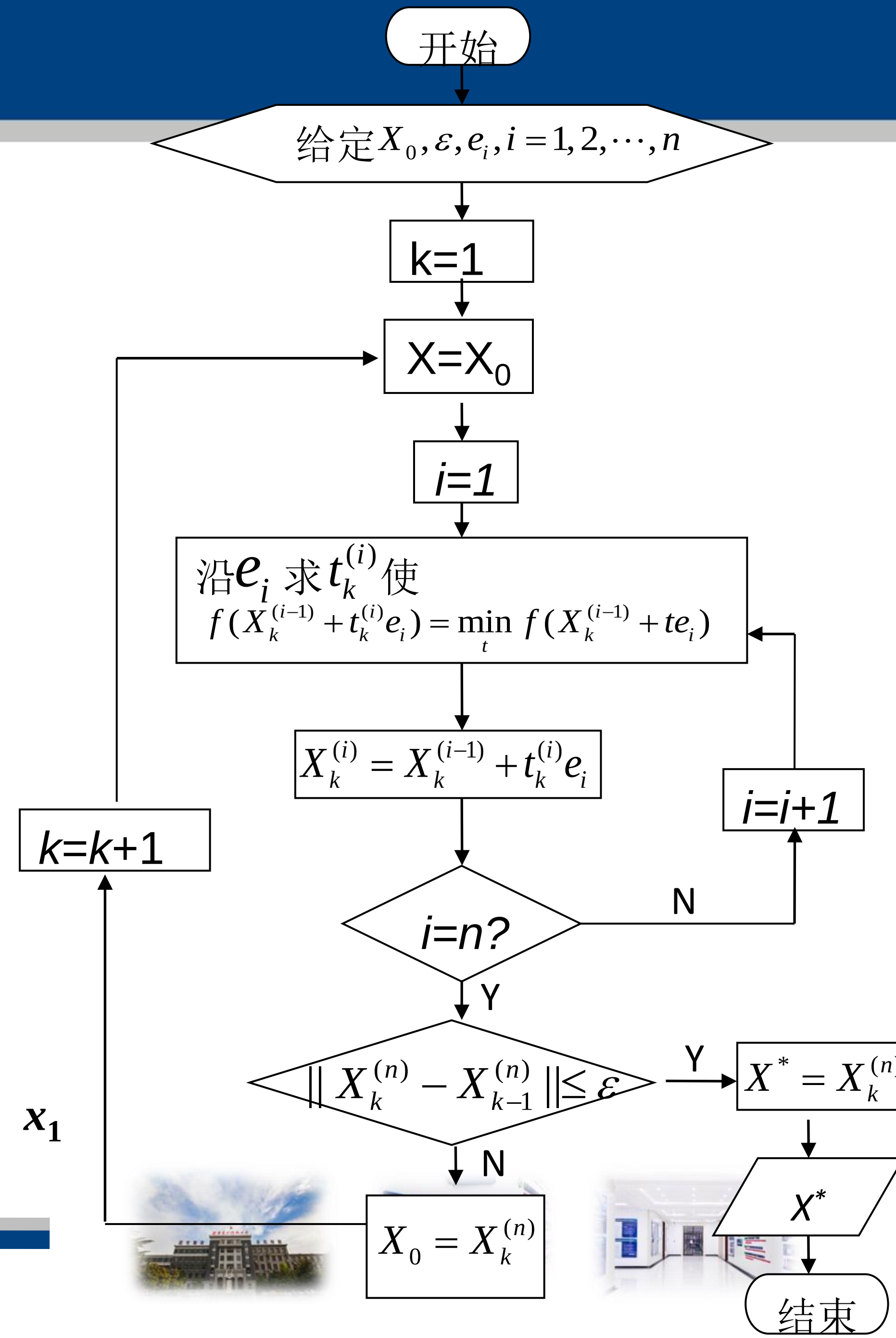
设某二元函数$S=f\left(X\right)=f\left(x_1,x_2\right)$，其等高线见下图。设极值存在的区间为

$a_1\le x_1\le b_1$，$a_2\le x_2\le b_2$，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。



坐标轮换法算法流程图


x_2
$X_2^{(1)}$
$X_0^{(1)}$
$X_1^{(1)}$
x_1


开始
给定$X_0, \varepsilon, e_i, i=1,2,\cdots,n$
k=1
$X=X_0$
$i=1$
沿e_i求$t_k^{(i)}$使
$f(X_k^{(i-1)}+t_k^{(i)}e_i)=\min_t f(X_k^{(i-1)}+te_i)$
$X_k^{(i)}=X_k^{(i-1)}+t_k^{(i)}e_i$
$i=n?$
N
$i=i+1$
Y
$\|X_k^{(n)}-X_{k-1}^{(n)}\|\le\varepsilon$
Y
$X^*=X_k^{(n)}$
N
$X_0=X_k^{(n)}$
$k=k+1$
X^*
结束

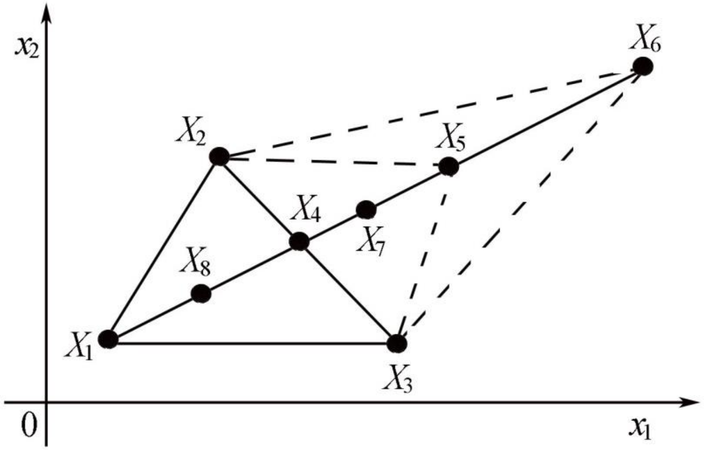
单形替换法基本原理

• 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．



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



外点罚函数法

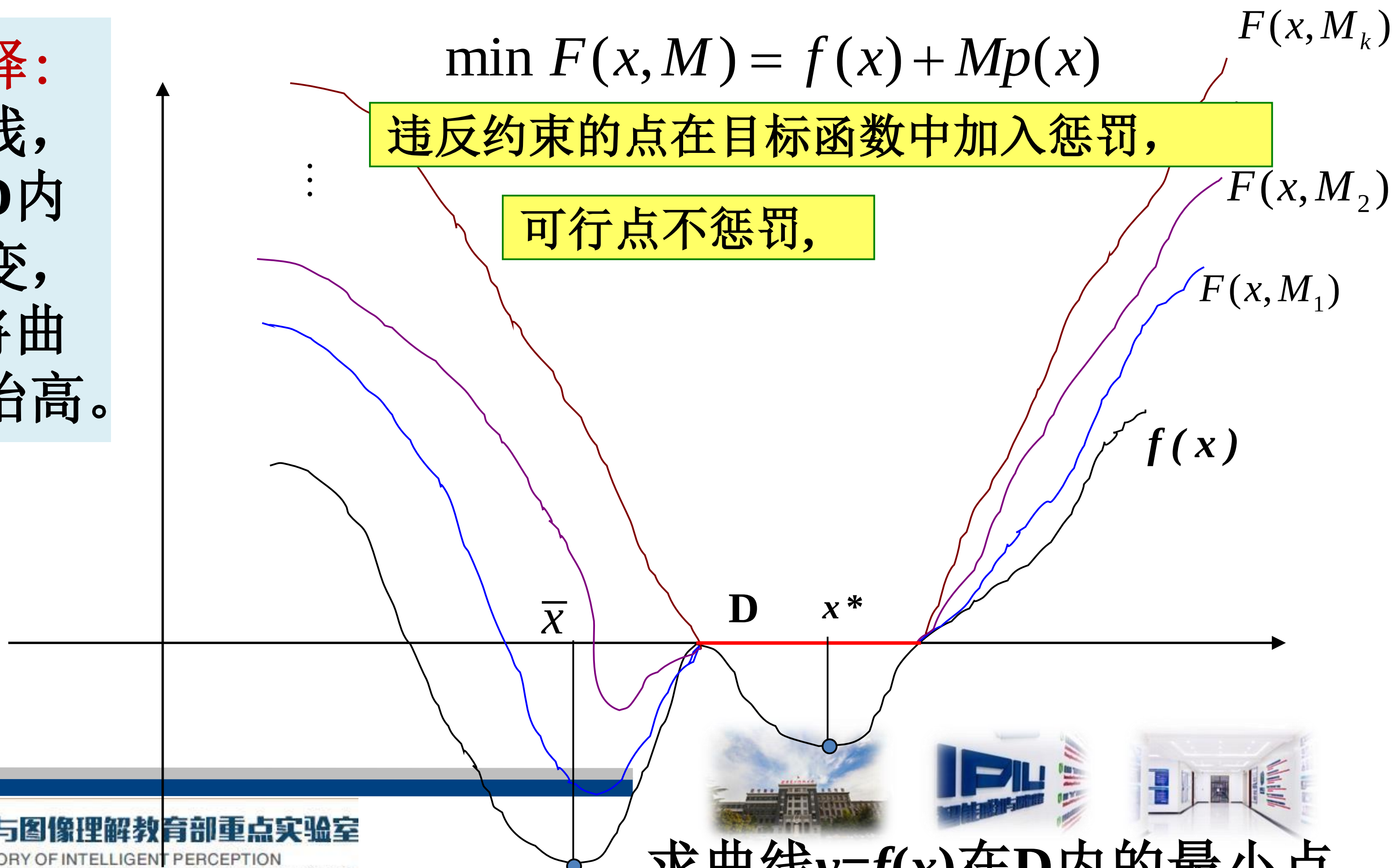
抬高到一定程度后，此曲线的无约束极小点就是原来函数在D内的最小点.

几何解释：改造曲线，使其在D内保持不变，在D外将曲线逐步抬高。

$$\min F(x,M)=f(x)+Mp(x)$$

违反约束的点在目标函数中加入惩罚，

可行点不惩罚，



求曲线$y=f(x)$在D内的最小点

外点罚函数法

步骤1：给定初始点 x^0，初始罚因子$M_1>0$(可取$M_1=1$), 精度

$\varepsilon>0, k:=1.$

步骤2：以 x^{k-1} 初始点，求解无约束优化问题

$$\min F(x,M_k)=f(x)+M_k p(x)$$

得到极小点 $x^*(M_k)$，记为 x^k，其中

$$p(x)=\sum_{i=1}^{m}(\min\{g_i(x),0\})^2+\sum_{j=1}^{l}h_j^2(x)$$

用解析法求驻点 或者用无约束优化方法求解

步骤3：若$M_k p(x^k)<\varepsilon$，则停止计算，得到近似极小点 x^k；

否则，令$M_{k+1}=cM_k$，置k:=k+1，转步骤2。

$p(x^k)\to 0$

$x^k\in D$

$c\in[2,50]$

常取 $c\in[4,10]$


内点罚函数法

基本思想:

迭代点在可行域的内部移动，并对接近可行域边界的点施加惩罚(加入障碍)，距边界越近障碍越大，这相当于在可行域的边界上筑起一道很高的“围墙”，阻止迭代点穿越边界，从而将最优解“挡”在可行域内。内点罚函数法又称为障碍函数法。

内点法要求迭代点在可行域内部移动，初始点必须是内点，可行域的内部必须是非空的，内点法只能处理不等式约束，等式约束构成的集合内部为空。内点法只适合于不等式约束问题。

$$\min \quad f(x)$$
$$s.t. \quad g_i(x) \geq 0, \ i = 1, \ldots, m$$

（障碍墙）

内点罚函数法

步骤1： 给定初始点 $x^0 \in \operatorname{int} D$，初始罚因子 $r_1 > 0$（可取 $r_1 = 10$），精度 $\varepsilon > 0, k := 1.$

步骤2： 以 x^{k-1} 初始点，求解无约束优化问题 $\min F(x, r_k) = f(x) + r_k B(x)$

得到极小点 $x^*(r_k)$，记为 x^k， 其中

$$B(x) = \sum_{i=1}^{m} \frac{1}{g_i(x)} \quad \text{或者} \quad B(x) = -\sum_{i=1}^{m} \ln\left(g_i(x)\right)$$

步骤3： 若 $r_k B(x^k) < \varepsilon$，则停止计算，得到近似极小点 x^k，否则，令

$r_{k+1} = c r_k$，置k:=k+1，转步骤2。

$c = 0.1$

混合罚函数法

混合罚函数法是采用内点法和外点法相结合的，用内点法处理不等式约束，用外点法处理等式约束。可以用来求解含不等式和等式约束的优化问题。

混合罚函数法-迭代步骤

混合法具有内点法的特点，迭代过程在可行域之内进行，参数的选择同内点法。

Step1：任意给定初始点，要求满足不等式约束，初始障碍因子 $r_0, c<1$, 置k=1;

Step2：假设已获迭代点 $X_{\text{k-1}}$，以 $X_{\text{k-1}}$ 为初始点，求解 $\min F\left(x, r_k\right)$,

得到极小点 $X(r_k)$.

Step3：若 $\left\|X_{\text{k}} - X_{\text{k-1}}\right\| \le \varepsilon$（例如：$\varepsilon=10^{-6}$）则停止计算，得到近似极小点 $X(\mathrm{r}_k)$；

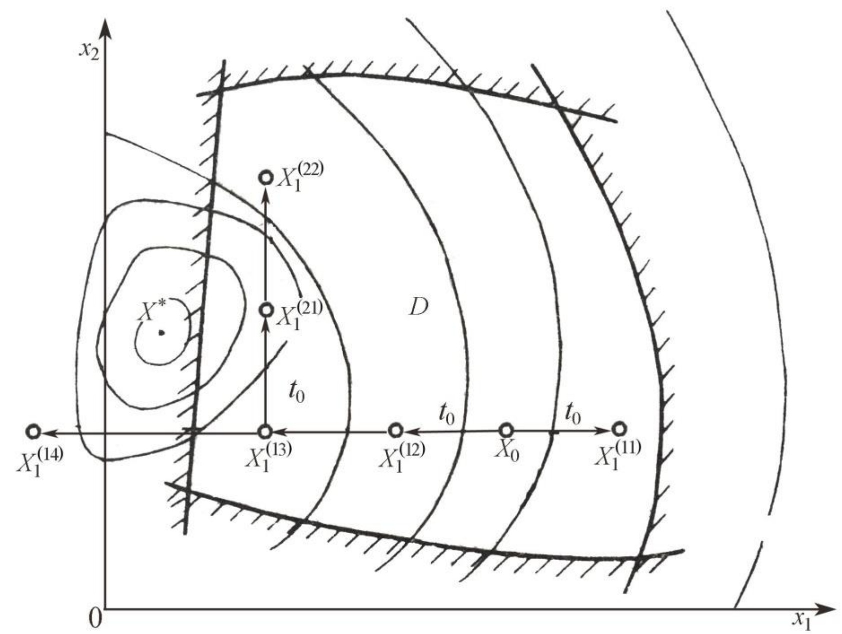
否则，转Step4

Step4： $r_{k+1} = c r_k$, k=k+1，转Step 2



约束坐标轮换法-基本原理

对于n维约束优化问题，依次沿坐标向量 $e_1,e_2,\ldots,e_n$ 的方向进行搜索时，由于只能在可行域内进行探索，故不宜采用最优步长，以免越出可行域。为此，通常利用加步搜索法来确定搜索步长，以求得一系列可行点，使目标函数值逐次下降，直至收敛到最优解。现以下图所示的二维情况进行说明:



无约束最优化方法

1 知识点复习

2 答疑

西安電子科技大學
XIDIAN UNIVERSITY

IPIU
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!