“最优化方法” 课程介绍

1、参考书籍：

《最优化理论与算法》，陈宝林，清华大学出版社。

《非线性规划》，陈开明，复旦大学出版社。

2、公共邮箱：

optimal_ecust@163.com

密码：**optimal2012**

基 础 知 识

一. 引言
二. 极值最优化问题的经典方法
三. 最优化问题实例
四. 最优化问题及基本概念
五. 梯度与Hesse阵
六. Taylor展开式
七. 凸集与凸函数
八. 极小点的判定条件
九. 算法及相关概念
十. 中止条件
十一. 收敛速度

一. 引言

1. 最优化定义

最优化是从所有可能方案中选择最合理方案以达到最优目标的一门学科。

（1） 达到最优目标的方案：最优方案（最优解）

（2） 搜寻最优方案的方法：最优化方法

最优化问题：寻求某些变量的取值使其符合某些限制条件，并使某个目标函数达到最大值或最小值的问题。

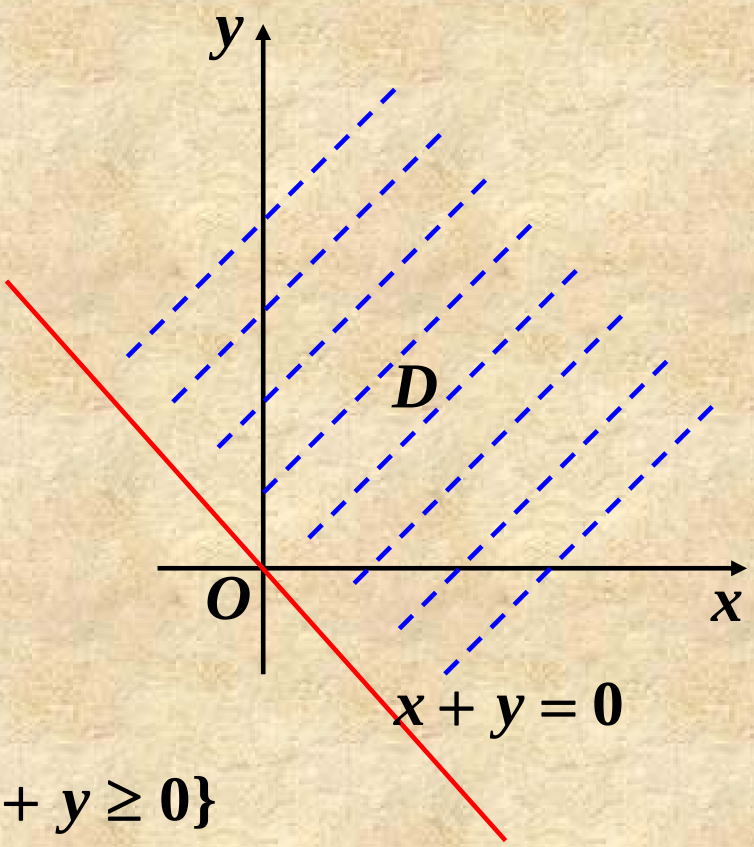
一般的数学形式为：

$$\begin{aligned} &\min \quad f(x) \\ &s.t. \quad x \in D \end{aligned}$$

其中 D 为可行域。

例.
$$\begin{aligned} &\min \quad z = x^2 + y^2 \\ &s.t. \quad x + y \geq 0 \end{aligned}$$

$$\begin{aligned} &\min \quad z = x^2 + y^2 \\ &s.t. \quad (x, y) \in D = \{(x, y) \mid x + y \geq 0\} \end{aligned}$$



2. 包含内容：

最优化又称数学规划： LP、NLP、DP、IP

- 最优化问题
  - 单目标
    - 静态问题
      - 无约束问题
      - 有约束问题
        - 线性规划
        - 非线性规划
    - 动态问题
  - 多目标

二. 极值最优化问题的经典方法

1. 求驻点的方法

(1) $$\min_{x \in R^n} f(x) \Rightarrow \nabla f(x) = 0$$

$$\Leftrightarrow \begin{cases} \dfrac{\partial f}{\partial x_1} = 0 \\ \vdots \\ \dfrac{\partial f}{\partial x_n} = 0 \end{cases}$$

(2) $\min \quad f(x_1, x_2, \ldots x_n)$

$$s.t. \begin{cases} h_1(x_1, x_2, \ldots x_n) = 0 \\ \qquad\vdots \\ h_l(x_1, x_2, \ldots x_n) = 0 \end{cases}$$

$$\Rightarrow \quad \min_{\substack{x \in R^n \\ \lambda \in R^l}} L(x, \lambda) = f(x) - \sum_{i=1}^{l} \lambda_i h_i(x)$$

$$\Rightarrow \begin{cases} \dfrac{\partial L}{\partial x} = \nabla f(x) - \lambda \nabla h(x) = 0 \\ \dfrac{\partial L}{\partial \lambda} = h(x) = 0 \end{cases}$$

三. 最优化问题实例

例1 多参数曲线拟合问题

已知热电阻 R 依赖于温度 t 的函数关系为

$$R = x_1 \exp\left(\frac{x_2}{t + x_3}\right) \quad 。$$

其中 x_1, x_2和 x_3是待定参数。通过试验，得到 t和 R 的15组数值。

i	t_i	R_i
1	50	34780
2	55	28610
3	60	23650
。。。	。。。	。。。
15	120	3307

试确定函数关系。

利用最小二乘思想，可将其化为无约束最优化问题：

$$\min f(x_1, x_2, x_3) = \sum_{i=1}^{15}[R_i - x_1 \exp(\frac{x_2}{t_i + x_3})]^2$$

例2 生产计划问题

某厂有 m 种资源，数量为 $b_1, b_2, \cdots, b_m$。这些资源可用来生产 n 种产品 $1, \cdots, n$。生产单位第 j 种产品需要的第 i 种资源的量为 a_{ij}。单位第 j 种产品的利润为 c_j。问：工厂应如何安排生产计划才能使总利润最大?

建立模型：

设第 j 种产品的产量为 x_j。总利润为 C。

模型： $$\max \quad C = \sum_{j=1}^{n} c_j x_j$$

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

例3 运输问题

已知有 m 个生产点 B_i，可供应某种商品的量分别为 $b_i(i=1,\cdots,m)$，有 n个销地 C_j，销量分别为 $c_j(j=1,\cdots n)$。从 B_i 到 C_j的单位运价为 a_{ij}。问：应如何安排运输方案才能使总运费最小？

运费		销	地			产量
		1	2	…	n	
产地	1	a_{11}	a_{12}	…	a_{1n}	b_1
	2	a_{21}	a_{22}	…	a_{2n}	b_2
	…	…	…	…	…	…
	m	a_{m1}	a_{m2}	…	a_{mn}	b_m
销量		c_1	c_2	…	c_n	

在产销平衡条件下，要求总运费最小的调运方案，可得如下模型。

设 x_{ij} 表示从第 i 个产地向第 j个销地的运量。则有

$$\min \quad z = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_{ij}$$

$$s.t. \begin{cases} \sum_{j=1}^{n} x_{ij} = b_i & (i = 1, 2, \dots, m) \\ \sum_{i=1}^{m} x_{ij} = c_j & (j = 1, 2, \dots, n) \\ x_{ij} \geq 0 & \begin{pmatrix} i = 1, \dots, m; \\ j = 1, \dots, n. \end{pmatrix} \end{cases}$$

四. 最优化问题及基本概念

1. 模型

$$\min\ f(x)$$

$$s.t.\ \begin{cases} g_i(x) \ge 0 & i = 1, \cdots, m \\ h_j(x) = 0 & j = 1, \cdots, p \end{cases} \qquad x = (x_1, \cdots, x_n)^T$$

$$\Leftrightarrow \quad \min\ f(x)$$

$$s.t.\ \begin{cases} g(x) \ge 0 \\ h(x) = 0 \end{cases} \qquad \begin{aligned} g(x) &= (g_1(x), \cdots, g_m(x))^T \\ h(x) &= (h_1(x), \cdots, h_p(x))^T \end{aligned}$$

$$\Leftrightarrow \quad \min\ f(x)$$

$$s.t.\ x \in D$$

其中 $D = \{x \mid g(x) \ge 0, h(x) = 0\}$称为可行域。

2. 基本概念：全局最优解与局部最优解

最优点：x^* ；　　最优值：$f(x^*)$； 最优解：$(x^*, f(x^*))^T$

邻域 $N(x_0,\delta)=\{x \mid \|x-x_0\|<\delta\}$

局部最优解 $x^*: f(x^*)< f(x), \forall x \in N(x^*,\delta) \cap D$

全局最优解 $x^*: f(x^*)< f(x), \forall x \in D$

3. 解法分类：

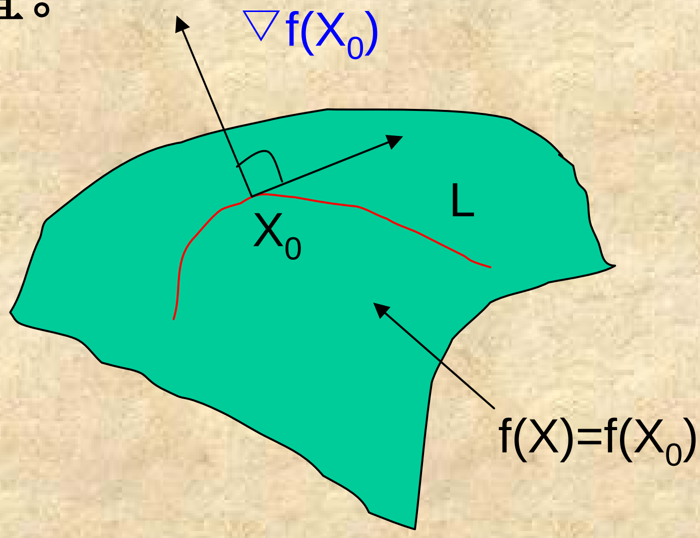
解析方法：利用函数的解析性质去构造迭代公式使之收敛到最优解（如牛顿法）。

直接方法：它对函数的分析性质如可微性没有要求，而是根据一定的数学原理来确定（如0.618法）。

五. 梯度与Hesse阵

1. 梯度 $\nabla f(X)=\left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\ldots,\frac{\partial f}{\partial x_n}\right)^T$

性质1： 函数在某点的梯度若不为**0**， 则必与过该点的等值线（面）垂直。



性质2： 梯度方向是函数具有最大变化率的方向，即函数值上升最快的方向。

2. 方向导数和下降（上升）方向

（1）方向导数：$\dfrac{\partial f(x_0)}{\partial p}=\lim\limits_{t\to 0^+}\dfrac{f(x_0+tp)-f(x_0)}{t\|p\|}$

函数 $f(x)$ 在点 x_0 处沿着方向 p 的方向导数。

（2）给定函数 $f(x)$ 和方向p，如果存在实数 $t>0$ ，使得对于任意的 $\delta\in(0,t)$，都有 $f(x_o+\delta p)<f(x_0)$ ，则称p为 $f(x)$ 在点 x_0 处的下降方向。

（3） 性质

(a) $\dfrac{\partial f(x_0)}{\partial p}=\nabla f(x_0)^T\cdot\dfrac{p}{\|p\|}$.

(b) $\dfrac{\partial f(x_0)}{\partial p}<0 \Rightarrow$ p是下降方向; $\dfrac{\partial f(x_0)}{\partial p}>0 \Rightarrow$ p 是上升方向。

(c) $\nabla f(x_0)^T\cdot p<0\Rightarrow$ p是下降方向。

（4） 常用梯度公式

$$\nabla c = 0,\quad \nabla(b^T x) = b,\quad \nabla(x^T x) = 2x,\qquad \nabla(x^T A x) = 2Ax.$$

3. Hesse矩阵

（1）雅可比矩阵：设

$$g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix},$$

$$\nabla g(x) = \begin{pmatrix} \frac{\partial g_1(x)}{\partial x_1} \cdots \frac{\partial g_m(x)}{\partial x_1} \\ \vdots \qquad\qquad \vdots \\ \frac{\partial g_1(x)}{\partial x_n} \cdots \frac{\partial g_m(x)}{\partial x_n} \end{pmatrix} = \left(\frac{\partial g_j}{\partial x_i} \right)_{m \times n}$$

（2）海赛（**Hesse**） 矩阵：

$$\nabla^2 f(x) = \nabla(\nabla f(x)) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{pmatrix} = \left(\dfrac{\partial^2 f}{\partial x_i \partial x_j} \right)_{n \times n}$$

（3）其它

$$\nabla\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}=0 \qquad \nabla x=\begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}=I \qquad \nabla(Ax)=A^T$$

设 $\varphi(t)=f(x_0+tp)$, 则

$$\varphi'(t)=\nabla f(x_0+tp)^T p,$$

$$\varphi''(t)=p^T\nabla^2 f(x_0+tp)^T p。$$

例 已知 $f(x_1,x_2,x_3)=x_2^2-2x_1x_2+3x_3x_1-x_3^3$，求

（1）$\nabla f(x_1,x_2,x_3)|_{(1,2,1)}$，$\nabla^2 f(x_1,x_2,x_3)|_{(1,2,1)}$ 。

（2）判断向量 $d=(2,0,1)^T$ 是否为函数 $f(x_1,x_2,x_3)$ 在点 $(1,2,1)$ 的下降方向？

解: （1） $\because \dfrac{\partial f}{\partial x_1}=-2x_2+3x_3$，　$\dfrac{\partial f}{\partial x_2}=2x_2-2x_1$，

$$\frac{\partial f}{\partial x_3}=3x_1-3x_3^2.$$

$$\nabla f(x_1,x_2,x_3)|_{(1,2,1)}=(-2x_2+3x_3,2x_2-2x_1,3x_1-3x_3^2)^T|_{(1,2,1)}$$

$$=(-1,2,0)^T$$

$$\therefore \nabla^2 f(x_1, x_2, x_3)|_{(1,2,1)} = \left.\begin{bmatrix} 0 & -2 & 3 \\ -2 & 2 & 0 \\ 3 & 0 & -6x_3 \end{bmatrix}\right|_{(1,2,1)}$$

$$= \begin{bmatrix} 0 & -2 & 3 \\ -2 & 2 & 0 \\ 3 & 0 & -6 \end{bmatrix}$$

（2）

$$\because d^T \cdot \nabla f(x_1, x_2, x_3)|_{(1,2,1)} = (2,0,1) \cdot \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} = -2 < 0$$

所以向量 $d = (2,0,1)^T$ 是函数 $f(x_1, x_2, x_3)$ 在点 $(1,2,1)$ 的下降方向。

六. Taylor展开式

1: $$f(x+p)=f(x)+\nabla f(x)^T p+\frac{1}{2}p^T\nabla^2 f(x)p+o(\|p\|^2)$$

$$=f(x)+\nabla f(x)^T p+\frac{1}{2}p^T\nabla^2 f(\bar{x})p\,,\quad \bar{x}=x+\theta p\quad (0<\theta<1)$$

2: $$f(x)=f(x_0)+\nabla f(x_0)^T(x-x_0)+\frac{1}{2}(x-x_0)^T\nabla^2 f(x_0)(x-x_0)+\cdots$$

例　函数 $f(x)=e^x$ 在点 $x=0$ 的泰勒展开式为

$$e^x=1+\frac{1}{1!}(x-0)+\frac{1}{2!}(x-0)^2+\cdots+\frac{1}{n!}(x-0)^n+o(x^n)$$

$$=1+x+\frac{1}{2}x^2+\cdots+\frac{1}{n!}x^n+o(x^n)$$

七. 凸集与凸函数

1.凸集

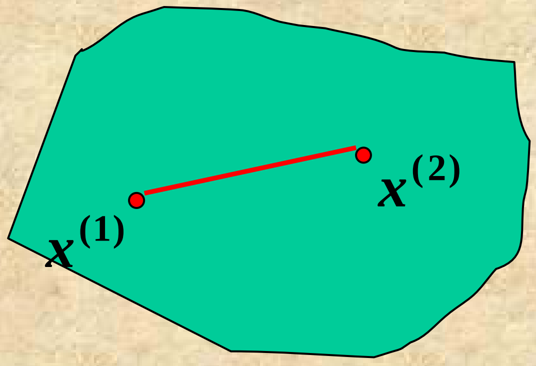
（1）凸组合:

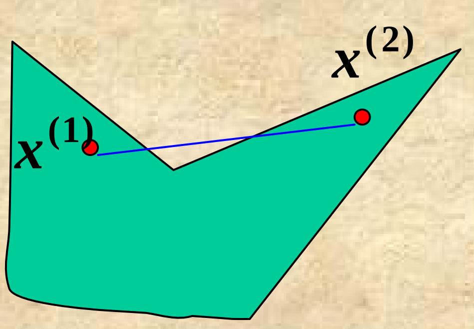
已知 $X \subset R^n$, 任取 k 个点 $x^{(i)} \in X$, 如果存在常数 $a_i \geq 0\ (i = 1, \cdots, k)$,

$\sum_{i=1}^{k} a_i = 1$, 使得 $\sum_{i=1}^{k} a_i x^{(i)} = \bar{x}$, 则称 $\bar{x}$ 为 $x^{(i)}\ (i = 1, \cdots, k)$ 的凸组合。

（2）凸集:

已知 $X \subset R^n$, 任取 $x^{(1)}, x^{(2)} \in X$ 及 $0 \leq \lambda \leq 1$, 如果 $\lambda x^{(1)} + (1-\lambda) x^{(2)} \in X$, 则称 X 为凸集。





性质　凸集的交仍为凸集。

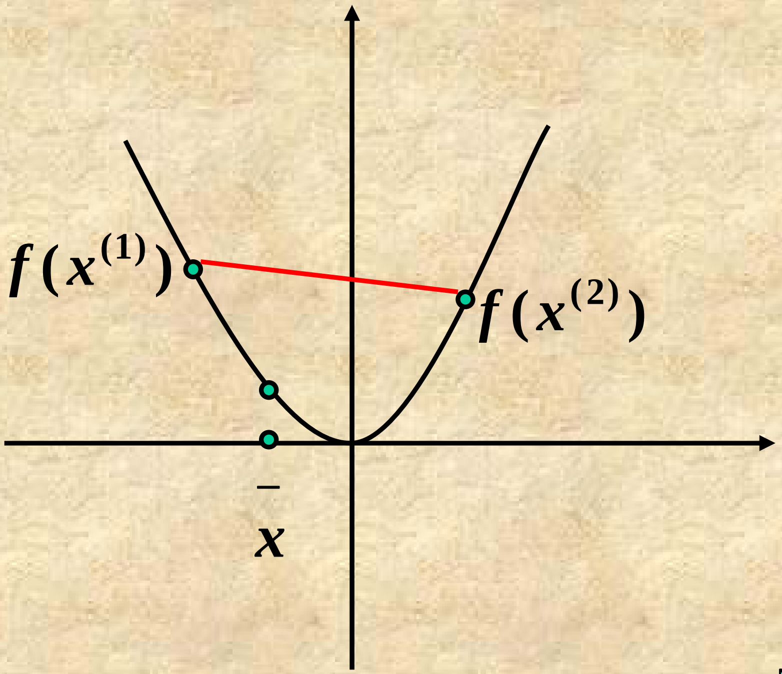
2.凸函数

设 $f: X \subset R^n \to R$，任取$x^{(1)}, x^{(2)} \in X$，如果$\forall a_1, a_2 \geq 0, \sum_{i=1}^{2} a_i = 1,$

有 $f(a_1 x^{(1)} + a_2 x^{(2)})(<) \leq a_1 f(x^{(1)}) + a_2 f(x^{(2)})$，则称 f 为X上的(严格)凸函数。

例子：　$f(x) = x^2$

$\bar{x} = a_1 x^{(1)} + a_2 x^{(2)}$



水平集： $D_\alpha=\{x\,|\,f(x)\le\alpha, x\in X, f$ 是凸函数$\}$。

性质：水平集一定是凸集。

3. 凸函数的性质

定理. 凸函数的局部极小点就是全局极小点。

4. 凸函数的判断条件

定理1. $f(x)$ 是凸集X上的凸函数的充要条件是 $\forall x^{(1)}, x^{(2)}\in X$，有

$$f(x^{(2)})\ge f(x^{(1)})+\nabla f(x^{(1)})^T(x^{(2)}-x^{(1)})$$

定理2.设 $f(x)$ 在凸集X上有二阶连续偏导数，则 $f(x)$ 是凸函数的充要条件是 $\forall x\in X$，有 $\nabla^2 f(x)$ 半正定。

例：正定二次函数 $f(x)=\frac{1}{2}x^TAx+b^Tx+c$，其中$A$ 是正定矩阵。

$f(x)$ 是凸函数。

例 判断 $f(x)=x_1^2+3x_2^2+9x_3^2-2x_1x_2+6x_2x_3+2x_1x_3$ 的凸性。

解:

$$\nabla f=\begin{pmatrix}2x_1-2x_2+2x_3\\6x_2-2x_1+6x_3\\18x_3+6x_2+2x_1\end{pmatrix}$$

$$\therefore\quad \nabla^2 f=\begin{pmatrix}2&-2&2\\-2&6&6\\2&6&18\end{pmatrix}$$

$$\because\quad |2|=2>0,\qquad \begin{vmatrix}2&-2\\-2&6\end{vmatrix}=8>0,\qquad |\nabla^2 f|=0$$

$\therefore\quad f(x)$ 是凸函数。

5. 凸规划

（1） $\min \quad f(x)$

$s.t. \quad x \in D$

其中 $f(x)$ 是凸函数， D 是凸集。

（2） $\min \quad f(x)$

$$s.t. \quad \begin{cases} g_1(x) \le 0 \\ \quad \vdots \\ g_l(x) \le 0 \\ h_1(x) = 0 \\ \quad \vdots \\ h_k(x) = 0 \end{cases}$$

其中 $f(x), g_i(x)\ (i = 1, 2, \cdots, l)$ 是凸函数，

$h_j(x)\ (j = 1, 2, \cdots, k)$ 是线性函数。

八. 极小点的判定条件

（1）必要条件： $f(x^*)=\min f(x)\Rightarrow\nabla f(x^*)=0$

（2）充分条件： $$\left.\begin{array}{l}\nabla f(x^*)=0\\ \nabla^2 f(x^*)>0\end{array}\right.\Longrightarrow f(x^*)=\min f(x)$$

（3） 两个结论

- 正定二次函数 $f(x)=\frac{1}{2}x^TQx+b^Tx+c$ 有唯一的极值点：$x^*=-Q^{-1}b$。
- $\nabla^2 f(x^*)>0\Rightarrow$ 函数在极小点附近的等值线为近似的同心椭圆。

$$f(x)=f(x^*)+\nabla f(x^*)^T(x-x^*)+\frac{1}{2}(x-x^*)^T\nabla^2 f(x^*)(x-x^*)+o\left(\left\|x-x^*\right\|^2\right)$$

九. 算法及相关概念

1、迭代算法

集合 D上的迭代算法A:

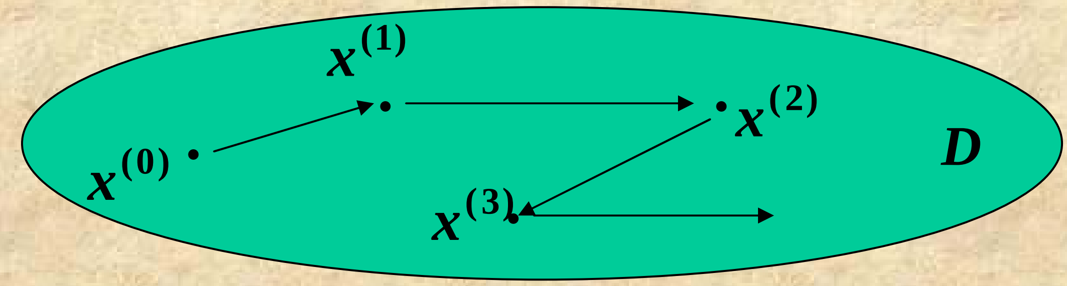
（1）初始点 $x^{(0)}$ ；

（2）按照某种规则A产生下一个迭代点 $x^{(k+1)} = A(x^{(k)})$ 。

（i）如果点列$\{x^{(k)}\}$收敛于最优解 x^*，则称算法 A 收敛。

（ii）如果 $f(x^{(0)}) > f(x^{(1)}) > \cdots > f(x^{(k)}) > \cdots$ ，则称算法A为

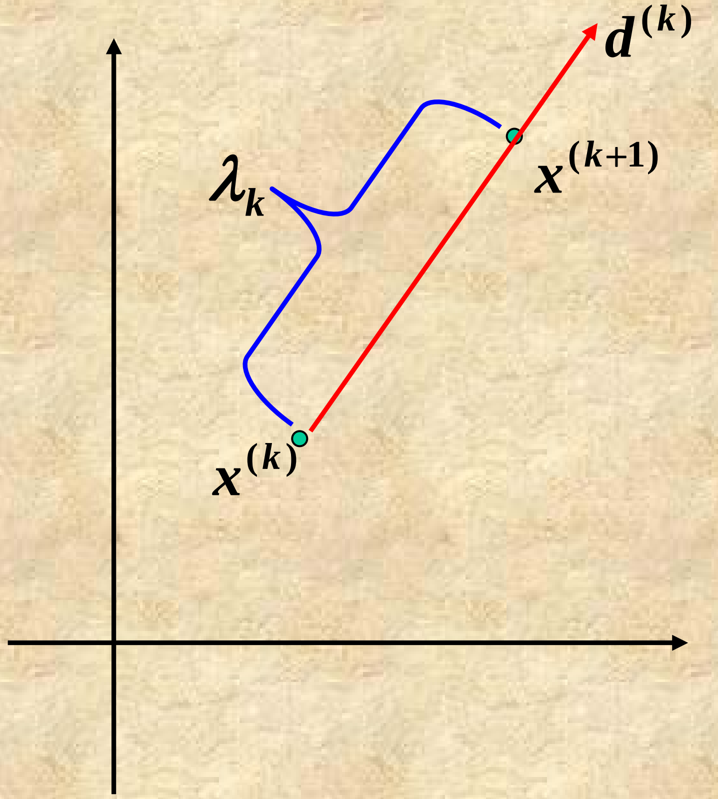
下降迭代算法。



已知当前迭代点 $x^{(k)}$，如何确定下一个迭代点$x^{(k+1)}$ ？

迭代格式：

$$x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$$



$d^{(k)}$:搜索方向；

λ_k ： 搜索步长。

2.下降迭代算法步骤

（1）给出初始点 $x^{(0)}$，令 $k=0$ ；

（2）按照某种规则确定下降搜索方向$d^{(k)}$;

（3）按照某种规则确定搜索步长 λ_k ，使得

$$f(x^{(k)}+\lambda_k d^{(k)})<f(x^{(k)})$$ ；

（4）令 $x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)}$, $k:=k+1$ ；

（5）判断 $x^{(k)}$是否满足停止条件。是则停止，否则转第2步。

搜索步长确定方法:

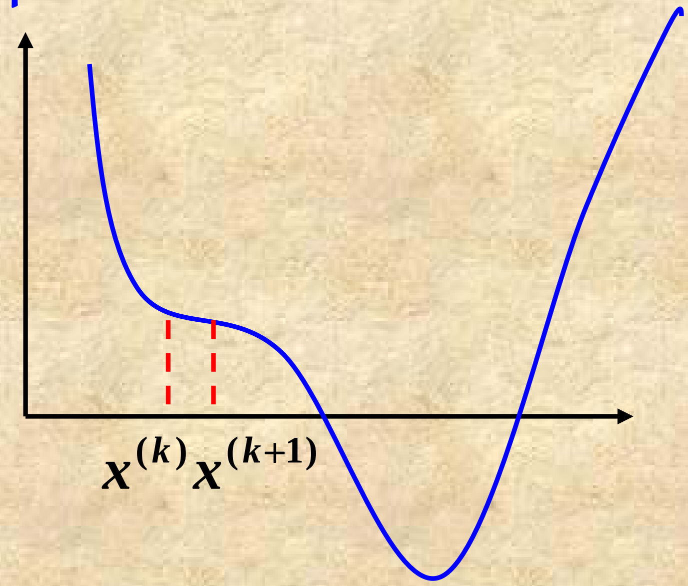
$$f(x^{(k)}+\lambda_k d^{(k)})=\min_{\lambda} f(x^{(k)}+\lambda d^{(k)})$$

称 λ_k 为最优步长，且有 $\nabla f(x^{(k)}+\lambda_k d^{(k)})^T d^{(k)}=0$ 。

十. 终止条件

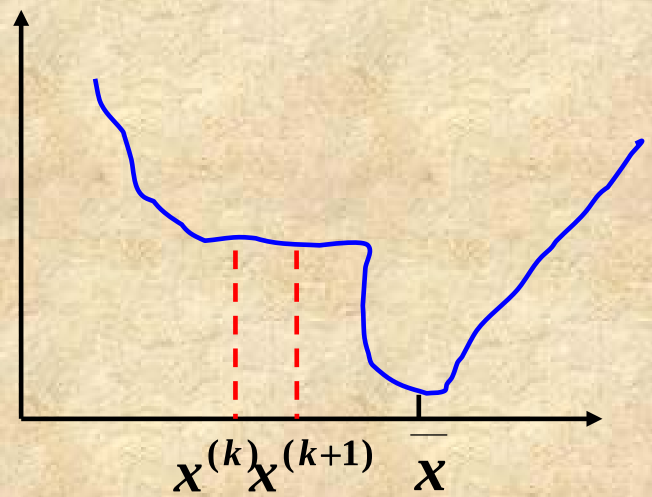
1. $$\left\| x^{(k+1)} - x^{(k)} \right\| < \varepsilon_1$$

2. $$\left| f(x^{(k+1)}) - f(x^{(k)}) \right| < \varepsilon_2$$



3. $$\frac{\left\| x^{(k+1)} - x^{(k)} \right\|}{\left\| x^{(k)} \right\|} < \varepsilon_3 \quad \left(\left\| x^{(k)} \right\| > \varepsilon \right)$$

4. $$\frac{\left| f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right) \right|}{\left| f\left(x^{(k)}\right) \right|} < \varepsilon_4 \quad \left(\left| f\left(x^{(k)}\right) \right| > \varepsilon \right)$$



$5. \|\nabla f(x^{(k)})\| < \varepsilon_5$

分析: 理论上可以认为 $\|\nabla f(x^{(k)})\| = 0$, 即 $x^{(k)}$ 是极值点。

$6. k \geq M$ (迭代次数的门槛值）。

十一. 收敛速度

1. 设算法A所得的点列为$\{x^{(k)}\}$，如果

$$\| x^{(k+1)} - x^{*} \| < \lambda \| x^{(k)} - x^{*} \|^{\alpha}, \quad \lambda, \alpha > 0.$$

则称$\{x^{(k)}\}$的收敛阶为α。

2. $\alpha = 1 \quad \Rightarrow$ 线性收敛

$1 < \alpha < 2 \Rightarrow$ 超线性收敛

$\alpha = 2 \quad \Rightarrow$ 二阶收敛