组合优化与凸优化
第3章 非线性规划(Nonlinear)的数学模型

刘绍辉
计算机科学与技术学院 哈尔滨工业大学
shliu@hit.edu.cn
2022年春季

第3章 非线性规划(Nonlinear programming)

- 非线性规划的数学模型
- 非线性规划的直观理解
- 无约束问题的最优性条件
- 凸无约束问题的最优性条件
- 解非线性规划问题的基本思路
- 有约束问题的最优性条件简介
- 总结

第3章 非线性规划-数学模型

◆ **非线性规划比线性规划要困难**

- 没有统一的数学模型
- 有自己特定的适用范围，到目前为止，还没有适合于各种非线性规划问题的一般算法

◆ **非线性规划问题举例**

◆ **选址问题：设有n个市场，第j个市场位置为(p_j,q_j)，它对某种货物的需要量为$b_j(j=1,2,\cdots,n)$.现计划建立m个仓库，第i个仓库的存储容量为$a_i(i=1,2,\cdots,m)$. 试确定仓库的位置，使各仓库对各市场的运输量与路程乘积之和为最小**

第3章 非线性规划-数学模型

◆选址问题

- 设第i个仓库的位置为$(x_i, y_i)(i=1,2,\cdots,m)$.第$i$个仓库到第$j$个市场的货物供应量为$z_{ij}(i=1,2,\cdots,m, j=1,2,\cdots,n)$. 则第$i$个仓库到第$j$个市场的距离为
- $d_{ij}=\sqrt{\left(x_i-p_j\right)^2+\left(y_i-q_j\right)^2}$,目标函数为

$$\sum_{i=1}^{m}\sum_{j=1}^{n} z_{ij}\, d_{ij}=\sum_{i=1}^{m}\sum_{j=1}^{n} z_{ij}\sqrt{\left(x_i-p_j\right)^2+\left(y_i-q_j\right)^2}$$

- 约束条件是:
  - ✓每个仓库向各市场提供的货物量之和不能超过该仓库的存储容量
  - ✓每个市场从各仓库得到的货物量之和应等于它的需要量
  - ✓运输量不能为负数

第3章 非线性规划-数学模型

◆选址问题

➢最终该问题的数学模型可以表示为

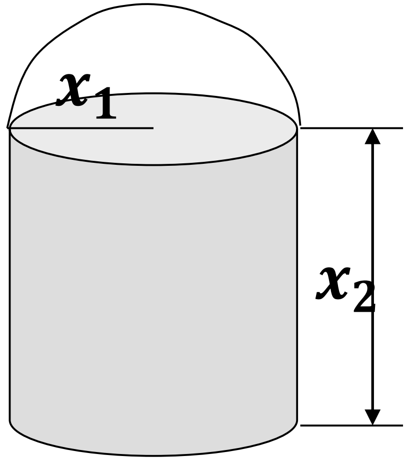
$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} z_{ij} \sqrt{(x_i - p_j)^2 + (y_i - q_j)^2},$$

$$s.t. \begin{cases} \sum_{j=1}^{n} z_{ij} \le a_i, i = 1, 2, \cdots, m \\ \sum_{i=1}^{m} z_{ij} = b_j, j = 1, 2, \cdots, n \\ z_{ij} \ge 0, i = 1, \cdots, m, j = 1, \cdots, n \end{cases}$$

第3章 非线性规划-数学模型

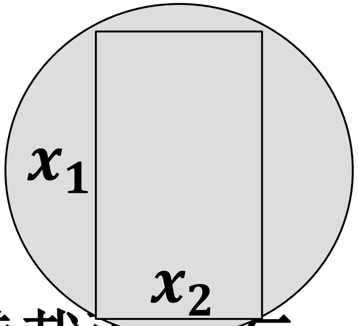
◆构件的表面积问题

- 一个半球形与圆柱形相接的构件，要求在构件体积一定的条件下，确定构件的尺寸使其表面积最小
- 设如图，圆柱形底半径为x_1,高为x_2,其表面积为
- $S=2\pi x_1^2+2\pi x_1x_2+\pi x_1^2=3\pi x_1^2+2\pi x_1x_2$
- 设其体积为V_0,则有
- $V_0=\frac{2}{3}\pi x_1^3+\pi x_1^2x_2$
- 故其数学模型为$\min S=3\pi x_1^2+2\pi x_1x_2, s.t.\begin{cases}\frac{2}{3}\pi x_1^3+\pi x_1^2x_2=V_0\\ x_1\geq 0, x_2\geq 0\end{cases}$



◆木梁设计问题

- 把圆形木材加工成矩形横截面的木梁，要求木梁高度不超过H,横截面的惯性矩($高度^2 \times 宽度$)不小于W,而且高度介于宽度与4倍宽度之间. 问如何确定木梁尺寸可使木梁成本最小.



- 解：如图，则圆形木材半径$r = \sqrt{\left(\frac{x_1}{2}\right)^2 + \left(\frac{x_2}{2}\right)^2}$
- 木材长度无法改变，因此成本只与圆形木材的横截面积有关，故目标函数应为$S = \pi r^2 = \pi\left(\frac{x_1^2}{4} + \frac{x_2^2}{4}\right)$
- 约束条件为: $x_1 \le H, x_1^2 x_2 \ge W, x_1 \ge x_2, x_1 \le 4x_2$

第3章 非线性规划-数学模型

◆木梁设计问题的数学模型可表示为

$$
\min \pi\left(\frac{1}{4}x_1^2+\frac{1}{4}x_2^2\right), s.t. \begin{cases} x_1 \le H \\ x_1^2 x_2 \ge W \\ x_1 - x_2 \ge 0 \\ x_1 - 4x_2 \le 0 \\ x_1, x_2 \ge 0 \end{cases}
$$

◆上面这些问题中，均含有决策变量的非线性函数，因此都属于非线性规划问题

第3章 非线性规划-一般数学模型

◆一般非线性规划的数学模型可表示为

$$min\ f(x); s.t. \begin{cases} g_i(x) \geq 0, i = 1,2,\cdots,m \\ h_j(x) = 0, j = 1,2,\cdots,l \end{cases}$$

◆这里$x = (x_1, x_2, \cdots, x_n)^T \in R^n$,为$n$维向量，$f, g_i, h_j$都是$R^n \to R^1$的映射

◆与线性规划类似，把满足约束条件的解称为可行解(Feasible solution),若记$\aleph = \{x|g_i(x) \geq 0, i = 1,2,\cdots,m, h_j(x) = 0, j = 1,2,\cdots,l\}$,称$\aleph$为可行域(Feasible Region),此时，模型可简记为

$$min\ f(x), x \in \aleph$$

◆如果可行域$\aleph = R^n$,则成为无约束问题，否则称为有约束问题

$$min\ f(x)或者min_{x \in R^n} f(x)$$

◆回忆微积分中的拉格朗日乘子法，当时只有等式约束，凡是这种将等式约束化为无约束极值问题来求解的，都称为经典极值问题

第3章 非线性规划-一般数学模型

◆在极值问题中引入不等式约束，标志现代数学规划理论的开始

◆不等式使得极值问题的处理更复杂，但也使得部分经典极值问题解决不了的问题得到解决，从而扩大了极值问题的应用范围

◆注意这里的问题也可以等价写为

◆$\min f(x); s.t. \begin{cases} g_i(x) \le 0, i = 1,2,\cdots,m \\ h_j(x) = 0, j = 1,2,\cdots,l \end{cases}$,或者

$\min f(x); s.t. \begin{cases} g_i(x) \ge 0 \\ i = 1,2,\cdots,m \end{cases}$ 或者

$\min f(x); s.t. \begin{cases} g_i(x) \le 0 \\ i = 1,2,\cdots,m \end{cases}$

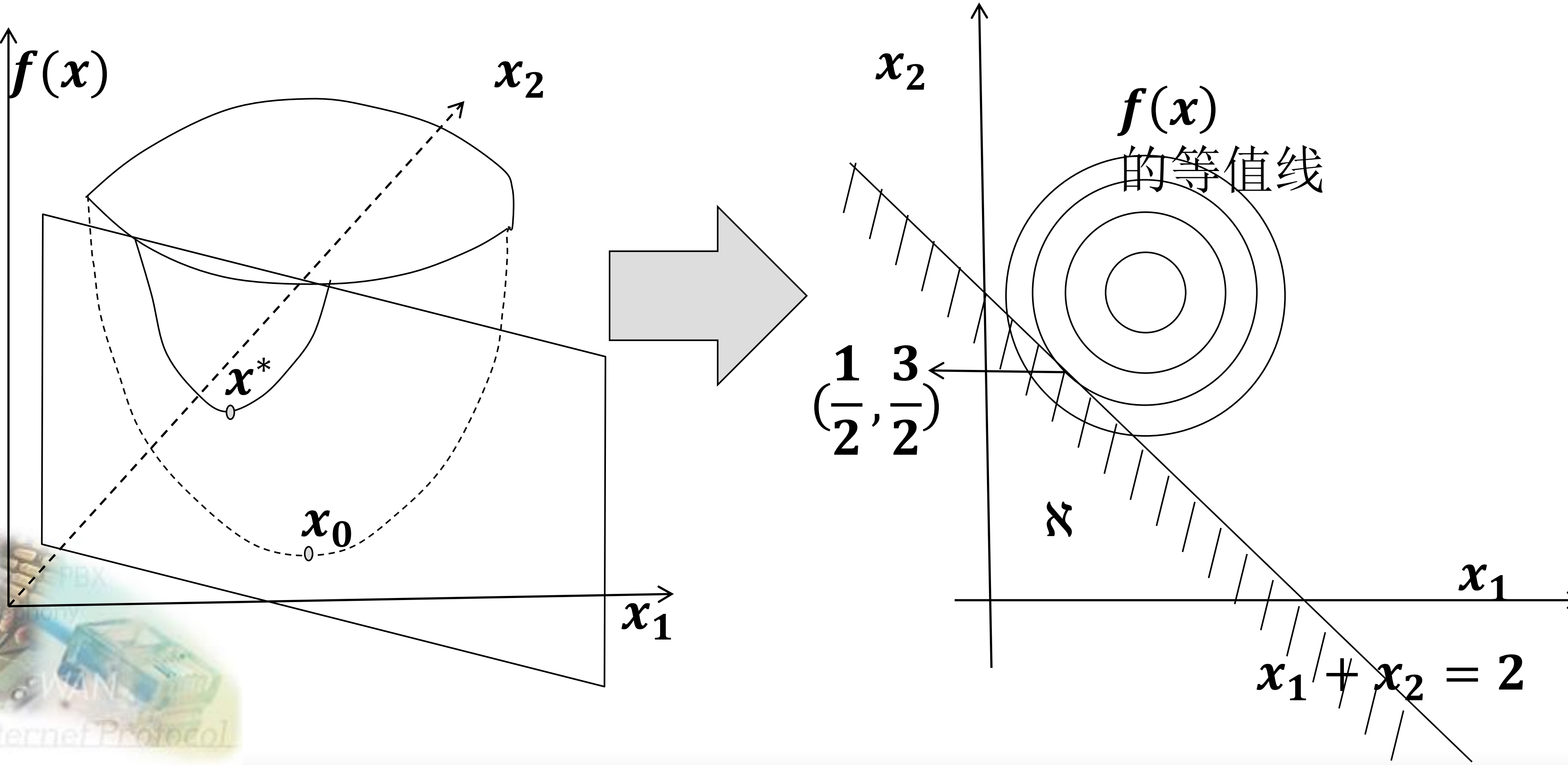
◆若所有的约束与目标函数均为凸函数，则成为凸规划问题，凸规划问题的解可简单判定是否是全局最优解

第3章 非线性规划(Nonlinear programming)

- 非线性规划的数学模型
- 非线性规划的直观理解
- 无约束问题的最优性条件
- 凸无约束问题的最优性条件
- 解非线性规划问题的基本思路
- 有约束问题的最优性条件简介
- 总结

第3章 非线性规划-直观理解

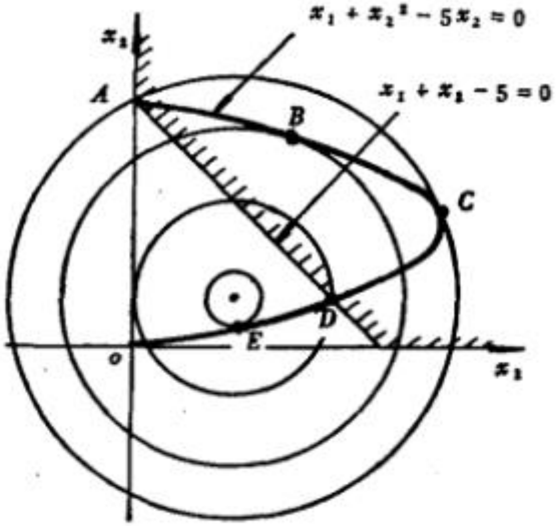
◆**例，非线性规划问题** $\min f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 2)^2; s.t. x_1 + x_2 - 2 \le 0$



第3章 非线性规划-图解

◆例，非线性规划问题 $min\ f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 2)^2$;

◆$s.t. \begin{cases} x_1 + x_2^2 - 5x_2 = 0 \\ x_1 + x_2 - 5 \geq 0 \\ x_1,\ x_2 \geq 0 \end{cases}$



✓如图

第3章 非线性规划-直观理解

- **局部最优解(local optimal solution)与全局最优解(Global optimal solution)**
  - 局部最优解：x^*是局部最优点，是指存在可行域中以x^*为中心的领域$N(x^*,\delta)$,使得$\forall x\in N(x^*,\delta)\cap\aleph, f(x^*)\le f(x)$,$\delta>0$,如果没有等号，则成为严格局部最优解；如果$\forall x\in\aleph, f(x^*)\le f(x)$,则为全局最优解，没有等号，则为严格全局最优解
  - 线性规划(LP)问题有最优解，则其最优解必可在可行域的极点上达到
  - 非线性规划(NLP)的最优解则可能是可行域上的任一点
  - LP问题的最优解一定是全局最优解
  - NLP则有局部最优解和全局最优解之分，并且一般的非线性规划算法往往求出的是局部最优解
  - 凸规划：对于具有凸性特点的函数来说，其极值点只有一个，因而该点既是局部最优点亦为全局最优点

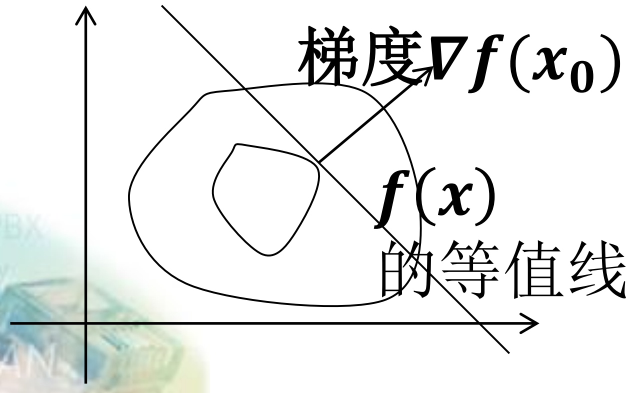
第3章 非线性规划(Nonlinear programming)

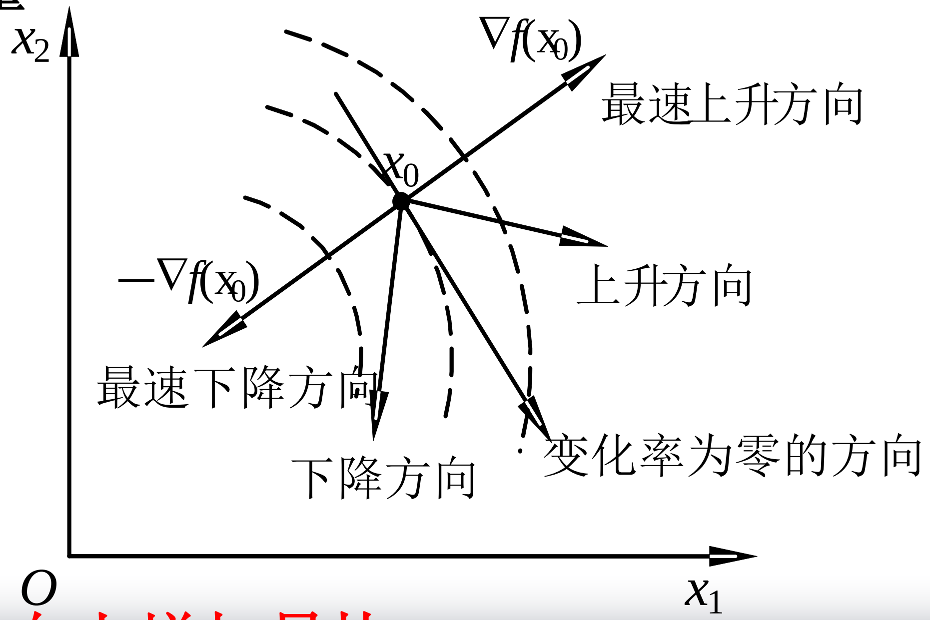
- ◆非线性规划的数学模型
- ◆非线性规划的直观理解
- ◆无约束问题的最优性条件
- ◆凸无约束问题的最优性条件
- ◆解非线性规划问题的基本思路
- ◆有约束问题的最优性条件简介
- ◆总结

第3章 非线性规划-无约束问题的最优性条件

◆ 多元函数的导数与极值

- 多元函数的一阶导数-梯度(Gradient):向量
- 微分与梯度的关系：函数$f(x)$在x_0处可微，则该点梯度也存在，并且$df(x_0)=\nabla f(x_0)^T\Delta x$
- 若$u=f(x)$是二元函数，则f在点x_0处的梯度的几何意义是：$\nabla f(x_0)$是过点x_0 的$f(x)$等值线在x_0点处的法向量
- 若$u=f(x)$是三元函数，则f在点x_0处的梯度的几何意义是：$\nabla f(x_0)$是过点x_0 的$f(x)$等值面在x_0点处的法向量





梯度方向表明函数值在该方向上增加最快

◆例，试求目标函数$f(x_1,x_2)=3x_1^2-4x_1x_2+x_2^2$在点$x^{(0)}=(0,1)^T$处的最速下降方向，如果沿着这个方向移动一个单位长度，新的目标函数值如何？

◆$Grad(f)=(6x_1-4x_2,-4x_1+2x_2)^T$,因此最速下降方向$p=-Grad(f)|_{x^{(0)}}=(4\ \ -2)^T$;该点函数值$f\left(x^{(0)}\right)=1$

◆单位向量$e=\left(\frac{2}{5}\sqrt{5}\ \ \frac{1}{5}\sqrt{5}\right)^T$,新点$x^{(1)}=x^{(0)}+e$,容易计算该点函数值为$f\left(x^{(1)}\right)=\frac{26}{5}-2\sqrt{5}$

◆多元函数的导数与极值

- ➢已知$f(x)=\sum_{i=1}^{n} b_i x_i = b^T x$,$b=(b_1,b_2,\cdots,b_n)^T$,$\nabla f(x)$=?
- ➢A为对称方阵，$f(x)=x^T A x$, $\nabla f(x)$=?

$b, 2Ax, Ax+b$

- ➢$f(x)=\frac{1}{2}x^T A x + b^T x + C, A^T = A, \nabla f(x)$=?

◆多元函数的二阶导数与Hessian矩阵H

- ➢$f(x)=\frac{1}{2}x^T A x + b^T x + C, A^T = A$, $\mathrm{H}(x)$=?
- ➢若$u=f(x)$二阶可导，记$\varphi(t)=f(x+t\Delta \mathrm{x})$,则 $\varphi'(t), \varphi''(t)$ =?

$\varphi'(t)=\nabla f(x+t\Delta x)^T \Delta x, \varphi''(t)=(\Delta x)^T H(x+t\Delta x)\Delta x$

◆多元函数的泰勒展开（Taylor Expansion)

- $f(x_0+\Delta x)=f(x_0)+\nabla f(x_0)^T\Delta x+\frac{1}{2}\Delta x^T\nabla^2 f(x_0)\Delta x+O(||\Delta x||^2)$
- 函数在x_0点的线性逼近: $f(x_0+\Delta x)\approx f(x_0)+\nabla f(x_0)^T\Delta x$
- 函数在x_0点的二次逼近$f(x_0+\Delta x)\approx f(x_0)+\nabla f(x_0)^T\Delta x+\frac{1}{2}\Delta x^T\nabla^2 f(x_0)\Delta x$

◆**(一阶必要条件)多元函数$f(x)$在x^*点可微，则x^*是$f(x)$的局部极值点$\Rightarrow \nabla f(x^*)=0$**

◆**(二阶必要条件)多元函数$f(x)$在x^*点二次可微，则x^*是$f(x)$的局部极小点$\Rightarrow \nabla f(x^*)=0$，且$\nabla^2 f(x^*)$半正定**

- (二阶充分条件1)多元函数$f(x)$在x^*点二次可微，则$\nabla f(x^*)=0$，且$\nabla^2 f(x^*)$正定 $\Rightarrow$ x^*是$f(x)$的严格局部极小点
- (二阶充分条件2)多元函数$f(x)$在x^*点的一个邻域内二次可微,$\nabla f(x^*)=0$,且对该邻域内的任意点x ,$\nabla^2 f(x)$半正定 $\Rightarrow$ x^*是$f(x)$的局部极小点
- 注意这两个充分条件的区别

- **定理:** 设实值函数$f(x) \in R^1, x \in R^n$, 在点x^*处可微.若存在向量$p = x - x^* \in R^n$，使$\nabla f(x^*)^T p < 0$,则存在数$\delta > 0$,使对每个$\lambda \in (0, \delta)$,有$f(x^* + \lambda p) < f(x^*)$
- **无约束非线性规划问题的最优解要满足的一阶必要条件：** 设实值函数$f(x) \in R^1, x \in R^n$, 在点x^*处可微.若x^*是无约束问题$min\, f(x)$的局部最优解，则$\nabla f(x^*) = 0$
- 证明： 反证法，若$\nabla f(x^*) \neq 0$,令$p = -\nabla f(x^*)$,则根据上面的定理，$\nabla f(x^*)^T p < 0$，从而有$f(x^* + \lambda p) < f(x^*)$，表明$x^*$不是无约束问题$min\, f(x)$的局部最优解，矛盾，证毕.

- **无约束非线性规划问题的最优解要满足的二阶必要条件：设实值函数$f(x) \in R^1, x \in R^n$, 在点x^*处二次可微.若x^*是无约束问题$min\ f(x)$的局部最优解，则$\nabla f(x^*) = 0$，且$\nabla^2 f(x^*) = H(x^*)$半正定**
- **证明，将$f(x)$在$x^* + \lambda p$点做泰勒展开，然后直接可证.**
- **无约束非线性规划问题的最优解的充分条件1：设实值函数$f(x) \in R^1, x \in R^n$, 在点x^*处二次可微.若$\nabla f(x^*) = 0$，且$\nabla^2 f(x^*) = H(x^*)$正定，则x^*是无约束问题$min\ f(x)$的严格局部最优解**

第3章 非线性规划-无约束问题的最优性条件

- **无约束非线性规划问题的最优解的充分条件2：函数$f(x)$在x^*点的一个δ邻域内二次可微, $\nabla f(x^*)=0$,且对该邻域内的任意点x, $\nabla^2 f(x)$半正定,则x^*是无约束问题$min\ f(x)$的局部最优解**
- **对于正定二次函数$f(x)=\frac{1}{2}x^TAx+b^Tx+C$($A$为$n$阶对称正定阵)有唯一极小点:$x^*=-A^{-1}b$，并且其等值面是同心椭球面族，且中心为其极小点$x^*$**
- **结论表明：若一个非二次函数$f(x)$在极小点x^*处的Hessian矩阵$\nabla^2 f(x^*)$是正定的，则这个函数在极小点附近近似于一个正定二次函数，这个结论以后经常用到**
- **大多数优化方法都是从二次函数模型导出。二次终止性是指算法应用于一个二次函数时，只要经过有限步迭代就能达到函数的极小值点x^*，因此具有很好的收敛性质**

◆例:利用最优性条件，求解下列问题

$$min\ f(x) = \frac{1}{3}x_1^3 + \frac{1}{3}x_2^3 - x_2^2 - x_1$$

解：求其梯度为0，可得四个平稳点$(x^{(1)} = (1,0)^T, x^{(2)} = (1,2)^T, x^{(3)} = (-1,0)^T, x^{(4)} = (-1,2)^T$

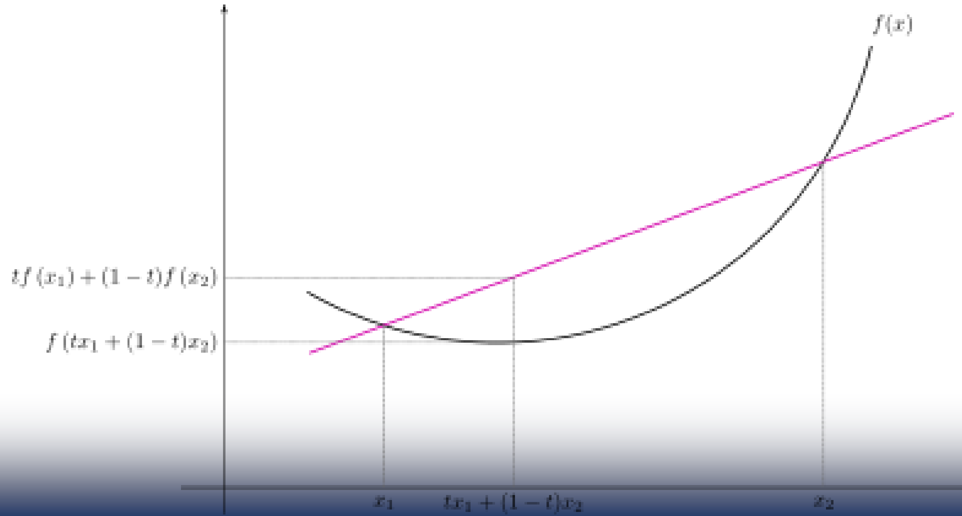
代入Hessian矩阵分别得到四点的Hessian矩阵为:

$\nabla^2 f\left(x^{(1)}\right) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}, \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}$,显然，1,4不定 $x^{(1)}, x^{(4)}$不是极值点，3为负定:$x^{(3)}$是极大点，2正定：$x^{(2)}$为严格局部最小值

- 例：利用最优性条件，求解问题$min\ f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$.
- 解：$\nabla f(x) = \begin{pmatrix} 4(x_1-2)^3+2(x_1-2x_2) \\ -4(x_1-2x_2) \end{pmatrix} \Rightarrow x^* = (2,1)^T$,
- $\nabla^2 f(x) = H(x) = \begin{pmatrix} 12(x_1-2)^2+2 & -4 \\ -4 & 8 \end{pmatrix}$,因此，$\nabla^2 \mathrm{f}(\mathrm{x}^*) = \begin{pmatrix} 2 & -4 \\ -4 & 8 \end{pmatrix}$,是半正定矩阵，选取点$x^*$的邻域，再计算其Hessian矩阵，也是半正定的，因此该点为其局部最优解

第3章 非线性规划-无约束问题的最优性条件

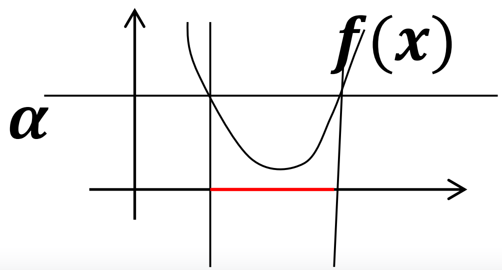
- 上述介绍的都是无约束问题极值点的最优性条件，要么为必要条件，要么为充分条件，但都不是充要条件
- 这些都只能用来研究规划问题的局部最优解
- 对一般的非线性函数，要给出极值点的充要条件的一般表达式非常困难
- 但对于一类特殊的函数，凸函数，有非线性凸规划问题极值点的充要条件



第3章 非线性规划(Nonlinear programming)

- ◆非线性规划的数学模型
- ◆非线性规划的直观理解
- ◆一般无约束问题的最优性条件
- ◆凸无约束问题的最优性条件
- ◆解非线性规划问题的基本思路
- ◆有约束问题的最优性条件简介
- ◆总结

- **(凸函数的一阶充要条件)多元函数f定义在非空开凸集$S \subset R^n$上的可微函数，则f为凸函数的充要条件是$f(y) \geq f(x) + \nabla f(x)^T(y - x), \forall x, y \in S$，如果>,则为严格凸函数**
- **(凸函数的二阶充要条件)多元函数f定义在非空开凸集$S \subset R^n$上的二阶可微函数，则f为凸函数的充要条件是f在S中的每一点的Hessian矩阵半正定。如果每点的Hessian矩阵正定，则严格凸函数。反之，严格凸函数，只能导致Hessian半正定。**
- **函数f定义在非空凸集$S \subset R^n$上的凸函数，则$f(x)$的任意一个极小点就是它在S上的全局极小点，而且所有极小点形成一个凸集.**



- 函数f定义在非空凸集$S \subset R^n$上的凸函数，则$f(x)$的任意一个极小点就是它在S上的全局极小点，而且所有极小点形成一个凸集
- 证明：假设x^*为局部极小点，则存在δ邻域$N(x^*, \delta)$内所有点都满足：$f(x) \geq f(x^*)$,在S中任取一点$\overline{x}$,则连接$\overline{x}, x^*$,显然存在一个$\lambda > 0$,使得$(1-\lambda)x^* + \lambda\,\overline{x} = x_0 \in N(x^*, \delta)$,此时有$f(x_0) \geq f(x^*)$,由于$f$是凸函数，因此$(1-\lambda)f(x^*) + \lambda f(\overline{x}) \geq f(x^*)$,从而可得$f(\overline{x}) \geq f(x^*)$,由该点的任意性，得$x^*$为全局极小点
- 关于极小点连成凸集，可采用水平集来证明，设x^*是一个极小点，极小值为β_0,则$H_S(f, \beta_0)$为凸集，因此该集合中的点$f(x) \leq \beta_0$,但这不可能<成立，因为x^*是全局极小点，因此得证。
- 函数f定义在非空凸集$S \subset R^n$上的可微凸函数，若存在点$x^* \in S$,使对于所有的点$x \in S$,都有$\nabla f(x^*)^T(x - x^*) \geq 0$.（注：这可由一阶凸函数的充要条件得到)，则$x^*$是$f(x)$在凸集$S$上的全局极小点。注意：这表明从$x^*$指向凸集$S$上的任意其它点的向量都与该点的梯度方向向量的夹角小于等于90度。

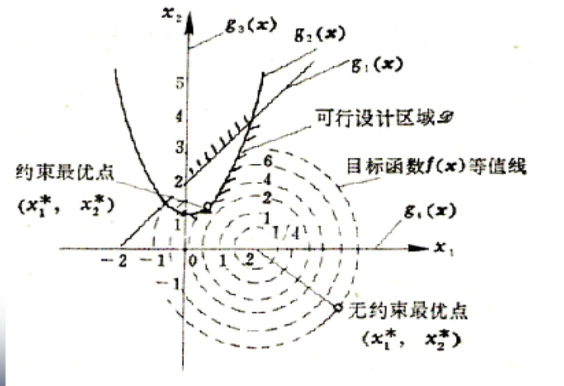
- 由前述可知，定义在凸集上的凸函数其极值点有很好的性质.把它应用到非线性规划问题上，相当于目标函数和约束都是凸函数，其可行域是凸集的规划问题，称为凸规划问题
- 上述介绍了如何判别函数为凸函数，可行域如何判断呢？
  - 根据凸集的交集仍然是凸集
  - 例如$g_i(x) \le 0, i = 1,2,\cdots,m, h_j(x) = 0, \mathrm{j} = 1,2,\cdots,l$是凸集，目标函数也是凸集，那就是凸规划
- 凸规划的局部最优解就是全局最优解，且为凸集。若目标函数是严格凸函数，又存在极小点，则此时凸规划的全局最优解唯一.

第3章 非线性规划(Nonlinear programming)

- ◆非线性规划的数学模型
- ◆非线性规划的直观理解
- ◆一般无约束问题的最优性条件
- ◆凸无约束问题的最优性条件
- ◆解非线性规划问题的基本思路
- ◆有约束问题的最优性条件简介
- ◆总结

第3章 非线性规划-无约束问题的最优性条件

- 无约束最优化问题具有最优性条件，感觉很简单呐！
- 实际问题直接应用最优性条件很困难，例如导数不存在，Hessian函数不存在，或者计算特别麻烦，例如一般情况下$\nabla f(x) = 0$是大型的非线性方程，求解困难，或者根本不可能有解析解
- 不论是无约束或有约束的优化问题，在实际应用中，要证明一个优化问题是否为凸规划，一般比较困难，有时甚至比求解优化问题本身还要麻烦。尤其对一些工程问题，由于其数学模型的型态都比较复杂，更难实现
- 因此，在优化设计的求解中，就不必花精力进行求证，而通常是从几个初始点出发，找出几个局部最优解，从中选择目标函数值最好的解
- 因此一般非线性规划采用数值计算中的迭代方法



◆最优化方法的结构

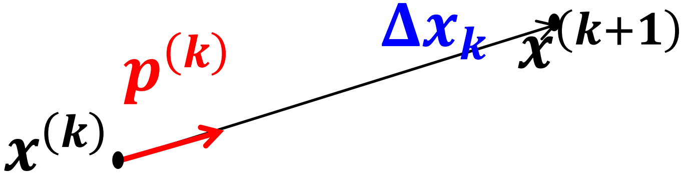
- 最优化方法通常采用迭代方法(iterative)求解
- 基本思想：给定一个初始点$x^{(0)}\in R^n$,按照某种迭代规则(一般称为算法)产生一个点$\{x^{(k)}\}$,使得当$\{x^{(k)}\}$是有穷点列时，其最后一个点是最优化模型问题的最优解；当$\{x^{(k)}\}$是无穷点列时，它有极限点，且其极限点是最优化模型问题的最优解
- 好的算法判断标准
  - ✓迭代点列能稳定地接近局部极小点x^*的邻域，然后迅速收敛于x^*. 当给定的某种收敛准则满足时，迭代终止。一般地，我们要证明迭代点列$\{x^{(k)}\}$的聚点(即子序列的极限点)为一局部极小点

◆下降方向:Descent directions, 步长：stepsize,迭代下降：iterative descent

第3章 非线性规划-基本思路

- 设$x^{(k)} \in R^n$是某迭代算法的第k轮迭代点，$x^{(k+1)} \in R^n$是第$k+1$轮迭代点，令Δx_k=$x^{(k+1)} - x^{(k)}$,此时有

$$x^{(k+1)} = x^{(k)} + \Delta x_k,$$



- 由于Δx_k是一个向量，且表示以$x^{(k)}$为起点，$x^{(k+1)}$为终点的n维向量，如果令$p^{(k)} = x^{(k+1)}/|| x^{(k+1)} ||$,则$p^{(k)}$为$\Delta x_k$方向的单位向量，此时$\Delta x_k = \lambda_k p^{(k)}$,从而可以获得求解非线性规划的基本迭代格式

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

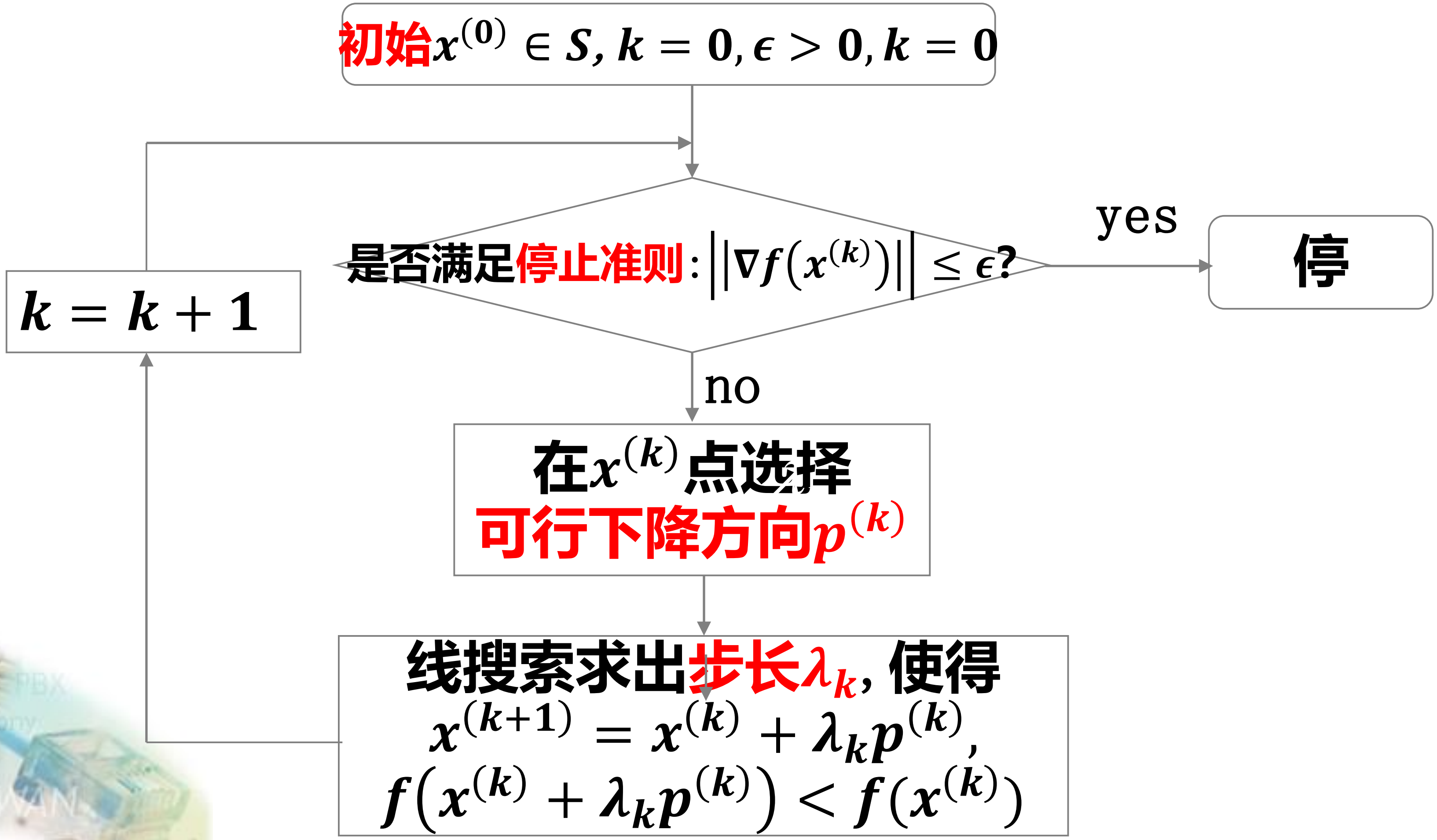
- 这里λ_k称为第k轮的步长，$p^{(k)}$称为第k轮的搜索方向。由此可知，不同的步长因子λ_k和不同的搜索方向$p^{(k)}$构成了不同的方法，后续章节介绍如何确定步长因子和搜索方向

- 在最优化中，搜索方向$p^{(k)}$是目标函数f在$x^{(k)}$处的下降方向(Descent Directions)，即满足：$\nabla f\left(x^{(k)}\right)^T p^{(k)} < 0$,或者$f\left(x^{(k)} + \lambda_k p^{(k)}\right) < f(x_k)$

第3章 非线性规划-基本思路

- **最优化方法的基本结构，给定初始点$x^{(0)}$**
  - 1.确定搜索方向$p^{(k)}$:即依照一定的准则，构造f在$x^{(k)}$点处的下降方向作为搜索方向
  - 2.确定步长因子λ_k，使目标函数值有某种意义的下降
  - 3.令$x^{(k+1)} = x^{(k)} + \lambda_k p^{(k)}$,若$x^{(k+1)}$满足某种终止条件，则停止迭代，得到近似最优解$x^{(k+1)}$,否则，重复以上步骤
- **这里面重点**
  - 步长因子
  - 下降方向
  - 也导致收敛速度的差别：衡量最优化方法有效性的重要方面

◆优化方法的基本结构



初始$x^{(0)} \in S, k=0, \epsilon>0, k=0$
是否满足停止准则: $\left\|\nabla f\left(x^{(k)}\right)\right\| \le \epsilon$?
yes
停
no
在$x^{(k)}$点选择
可行下降方向$p^{(k)}$
线搜索求出步长λ_k, 使得
$x^{(k+1)} = x^{(k)} + \lambda_k p^{(k)}$,
$f\left(x^{(k)} + \lambda_k p^{(k)}\right) < f\left(x^{(k)}\right)$
$k = k+1$

- 下降方向与可行下降方向(Descent direction and feasible descent direction)
- 根据前面迭代算法的要求$x^{(k+1)} = x^{(k)} + \lambda_k p^{(k)}$,要求新点的目标函数值要比上一次迭代点的目标函数值小，因此

$$f(x^{(k+1)}) = f\left(x^{(k)} + \lambda_k p^{(k)}\right) < f(x^{(k)}),$$

表明搜索方向$p^{(k)}$必须指向目标函数值下降的方向

- **定义(下降方向):** 是指对目标函数$f: R^n \to R^1, \bar{x} \in R^n$, 向量$p \in R^n (p \neq 0)$,若存在$\delta > 0, \forall \lambda \in (0, \delta)$,都有$f(\bar{x} + \lambda p) < f(\bar{x})$,则称向量$p$为函数在$\bar{x}$的下降方向
- 凡满足这种迭代性质的最优化方法都可称为**下降方法**(Descent methods)

第3章 非线性规划-基本思路

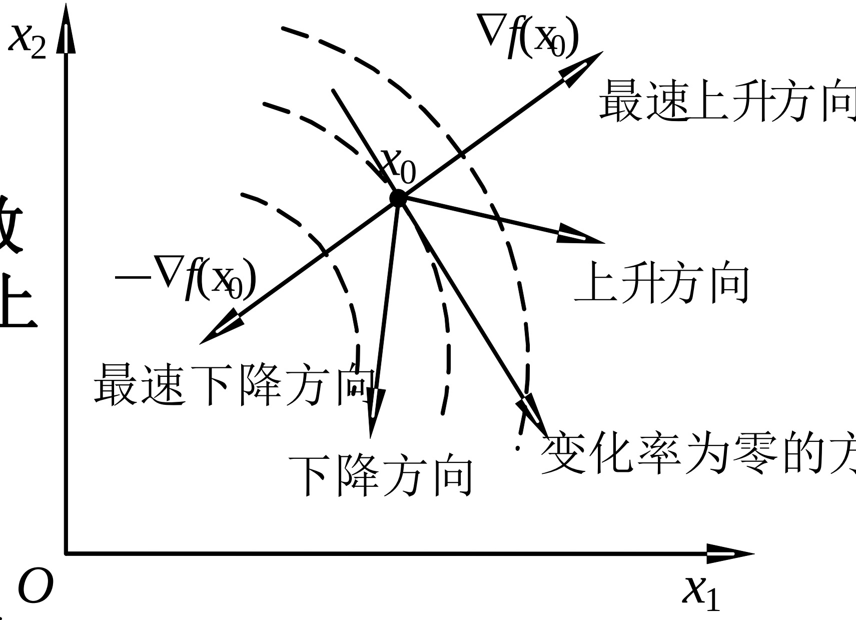
◆可行下降方向

- 对于有约束的非线性规划问题, $min\ f(x), x \in \aleph$,不仅要求下降，而且$x^{(k+1)}$必须仍然在原问题的可行域内，此时称为可行下降方向

◆步长因子的确定

- 显然步长λ是$f(x^{(k)}+\lambda p^{(k)})$的函数，因此求步长因子$\lambda_k$,就相当于求上述函数的极小点

$$f\left(x^{(k)}+\lambda_k p^{(k)}\right)$$
$$=min_\lambda f(x^{(k)}+\lambda p^{(k)})$$



- 上述求最优步长的过程称为一维搜索，或者线性搜索，具体怎么求?

- 这个方向有什么特殊性质?

◆一维搜索的最优步长λ_k,所对应的点$x^{(k+1)}$处的目标函数的梯度$\nabla f(x^{(k+1)})$与搜索方向$p^{(k)}$正交.

◆定理：设目标函数$f(x)$具有连续的一阶偏导数，$x^{(k+1)}$由下列规则产生：

$$\begin{cases} f\left(x^{(k)}+\lambda_k p^{(k)}\right)=min_\lambda f\left(x^{(k)}+\lambda p^{(k)}\right) \\ x^{(k+1)}=x^{(k)}+\lambda_k p^{(k)} \end{cases}$$

则有：$\nabla f\left(x^{(k+1)}\right)^{\mathrm{T}} p^{(k)}=0$

证明：设$\varphi(t)=f(x^{(k)}+tp^{(k)})$，因为$\lambda_k$是$\varphi(t)$的极值点，因此必为其临界点，从而必有$\frac{\partial\varphi}{\partial t}|_{\lambda_k}=0 \Rightarrow$

$\nabla f\left(x^{(k)}+\lambda_k p^{(k)}\right)^T p^{(k)}=0$ 证毕.

- 显然，这种方法是一类迭代方法，令$T: R^n \to R^n$

$$x^{k+1} = T(x^{(k)})$$

- 就是要找到映射的一个不动点(a fixed point) x^*,使得$x^* = T(x^*)$，这只需要映射T关于某种范数是收缩映射(Contraction mapping)：$||T(x) - T(y)|| \le \beta||x - y||, \forall x, y \in R^n$,$\beta < 1$,则意味着映射$T$有唯一的不动点，从任意$x^{(0)}$点，迭代后一定收敛于$x^*$点
- 如果$\beta = 1$,且范数为欧氏范数，且有至少一个不动点，则迭代$x^{(k+1)} = (1 - \alpha_k)x^{(k)} + \alpha_k T\left(x^{(k)}\right), \alpha_k \in [0, 1]$,且$\sum_{k=0}^{\infty} \alpha_k(1 - \alpha_k) = \infty$,从任意$x^{(0)}$点出发，迭代后一定收敛于$x^*$点

◆计算的终止条件(Termination criterion)与收敛速度(Convergence rate)

- **TC1**.最自然的停止准则：$\left\|\nabla f\left(x^{(k)}\right)\right\| < \epsilon_1$, ϵ_1为某个事先定义的精度，这表明梯度向量趋近于0，并且迭代序列$\{x^{(k)}\}$收敛到某个平稳点
- **TC2**. 相邻两次迭代点关系： $\left\|x^{(k+1)} - x^{(k)}\right\| < \epsilon_2$,或者 $\frac{\left\|x^{(k+1)} - x^{(k)}\right\|}{\left\|x^{(k)}\right\|} < \epsilon_3$,
- **TC3**.相邻两次迭代函数值的关系:$f\left(x^{(k)}\right) - f\left(x^{(k+1)}\right) < \epsilon_4$,或者$\frac{f\left(x^{(k)}\right) - f\left(x^{(k+1)}\right)}{\left|f\left(x^{(k)}\right)\right|} < \epsilon_5$ 为什么不用$\left|f\left(x^{(k)}\right) - f(x^*)\right| < \epsilon$

在不同情况下，可以分别采用这些准则来终止迭代，这里这些ϵ_i都是事先定义的精度值

◆计算的终止条件(Termination criterion)与收敛速度(Convergence rate)

- 收敛速率是算法的局部刻画，度量了优化算法的有效性，令迭代序列$\{\boldsymbol{x}^{(k)}\}$依某范数收敛到$\boldsymbol{x}^*$:

$\lim_{k\to\infty}\left\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^*\right\|=\mathbf{0}$,若存在实数$\boldsymbol{\alpha}\geq\mathbf{1}$,和与迭代次数$\boldsymbol{k}$无关的正数$\boldsymbol{\beta}>\mathbf{0}$，使得

$$\lim_{k\to\infty}\frac{\left\|\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^*\right\|}{\left\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^*\right\|^{\boldsymbol{\alpha}}}=\boldsymbol{\beta}$$

则称算法产生的点列具有$\boldsymbol{Q}-\boldsymbol{\alpha}$阶收敛速度，$\boldsymbol{Q}$: $\boldsymbol{Quotient}$

- 当$\boldsymbol{\alpha}=\mathbf{1},\boldsymbol{\beta}\in(\mathbf{0},\mathbf{1})$,迭代点列$\{\boldsymbol{x}^{(k)}\}$叫做具有Q-线性收敛
- 当$\boldsymbol{\alpha}\in(\mathbf{1},\mathbf{2}),\boldsymbol{\beta}>\mathbf{0}$或$\boldsymbol{\alpha}=\mathbf{1},\boldsymbol{\beta}=\mathbf{0}$时，$\{\boldsymbol{x}^{(k)}\}$叫做具有Q-超线性收敛
- 当$\boldsymbol{\alpha}=\mathbf{2}$，$\{\boldsymbol{x}^{(k)}\}$叫做具有Q-二阶收敛(Q-quadratic)

第3章 非线性规划-基本思路

- **定理**：如果迭代点列$\{x^{(k)}\}$具有Q-超线性收敛到x^*, 则$\lim\limits_{k\to\infty}\frac{\|x^{(k+1)}-x^{(k)}\|}{\|x^{(k)}-x^*\|}=1$,反之一般不成立
- 表明可以用相邻两次迭代点关系: $\|x^{(k+1)}-x^{(k)}\|<\epsilon_2$,来作为终止条件,这就是前面提到的终止条件TC2
- 当然也可以用TC3
- 问题：TC2和TC3有时候矛盾，就是一个大，另一个很小，此时Himmeblau提出使用TC2和TC3同时作为终止准则，最终准则如下
  - TC4:当$\|x^{(k)}\|>\epsilon_3$和$|f(x^{(k)})|>\epsilon_3$时，采用$\frac{\|x^{(k+1)}-x^{(k)}\|}{\|x^{(k)}\|}\le\epsilon_2$, $\frac{|f(x^{(k)})-f(x^{(k+1)})|}{|f(x^{(k)})|}<\epsilon_2$; 否则采用$\|x^{(k+1)}-x^{(k)}\|<\epsilon_2$，$|f(x^{(k)})-f(x^{(k+1)})|\le\epsilon_2$
- 对于有一阶导数信息，且收敛速度不太快的算法，例如共轭梯度法，可以采用TC1的终止条件：$\|\nabla f(x^{(k)})\|<\epsilon_1$，但由于临界点可能是鞍点，因此建议可以结合上述的TC4和TC1结合使用，一般可取$\epsilon_1=10^{-4}$, $\epsilon_2=\epsilon_3=10^{-5}$
- 再次提及：大多数最优化方法是从二次函数模型导出，因为在极小点附近可用二次函数很好的近似，此时如果算法具有二次终止性，则会很快收敛

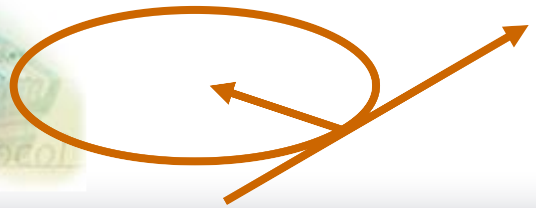
第3章 非线性规划-基本思路

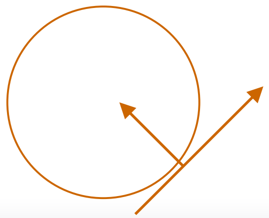
◆二次终止性

- 一个算法用于解正定二次函数的无约束极小时，有限步迭代可达最优解，则称该算法具有二次终结性。
- 二次终结性=共轭方向+精确一维搜索

◆共轭方向(conjugate direction)

- 定义:设 $A_{n\times n}$为对称正定矩阵,若$\boldsymbol{p}^{(1)},\boldsymbol{p}^{(2)}\in \boldsymbol{R}^n,\boldsymbol{p}^{(1)}\neq \mathbf{0},\boldsymbol{p}^{(2)}\neq \mathbf{0}$,满足$\boldsymbol{p}^{(1)^T}\boldsymbol{A}\boldsymbol{p}^{(2)}=\mathbf{0}$，则称$\boldsymbol{p}^{(1)},\boldsymbol{p}^{(2)}$关于矩阵$A$共轭.若$\boldsymbol{m}$个不同非零向量 $\boldsymbol{p}^{(1)},\boldsymbol{p}^{(2)},\cdots,\boldsymbol{p}^{(m)}\in \boldsymbol{R}^n,\boldsymbol{m}<\boldsymbol{n}$, 满足$\boldsymbol{p}^{(i)^T}\boldsymbol{A}\boldsymbol{p}^{(j)}=\mathbf{0}, \boldsymbol{if}\ \boldsymbol{i}\neq \boldsymbol{j}$,则称这$\boldsymbol{m}$个向量为共轭向量组，或$\boldsymbol{A}$共轭,如果$\boldsymbol{A}=\boldsymbol{I}$(单位矩阵),则共轭等价于正交





正定

第3章 非线性规划-基本思路

◆ **一般的共轭方向法(General conjugate direction)**

- **Step1**:给定初始点$x^{(0)}$, $\epsilon > 0, k = 0$, 计算$\nabla f(x)|_{x^{(0)}} = g(x^{(0)})$,再计算方向$p^{(0)}$,使得${p^{(0)}}^T g^{(0)} < 0$;
- **Step2**: 若$\|g^{(k)}\| \le \epsilon$, 终止
- **Step3**: 计算步长因子λ_k，令$x^{(k+1)} = x^{(k)} + \lambda_k p^{(k)}$,
$$\begin{cases} f\left(x^{(k)} + \lambda_k p^{(k)}\right) = min_\lambda f\left(x^{(k)} + \lambda p^{(k)}\right) \\ x^{(k+1)} = x^{(k)} + \lambda_k p^{(k)} \end{cases};$$
- **Step4**:通过某种共轭方向法计算$p^{(k+1)}$,使得${p^{(k+1)}}^T G p^{(j)} = 0, j = 0, 1, \cdots, k$
- **Step5**: 令$k = k + 1$, Goto Step2

◆ **共轭方向法基本定理：对于正定次函数，GCD方法至多经n步精确线性搜索终止**

- 当$d^{(1)}, d^{(2)}, \ldots, d^{(m)}$关于正定矩阵$A$两两共轭时，$d^{(1)}, d^{(2)}, \ldots, d^{(m)}$线性无关
  - 证明:设$\boldsymbol{d} = \boldsymbol{\alpha_1 p^{(1)}} + \boldsymbol{\alpha_2 p^{(2)}} + \cdots + \boldsymbol{\alpha_m p^{(m)}} = 0$, $\forall \boldsymbol{j} = \mathbf{1}, \mathbf{2}, \cdots, \boldsymbol{m}, \boldsymbol{p^{(j)}}^{\boldsymbol{T}} \boldsymbol{A d} = \boldsymbol{\alpha_j p^{(j)}}^{\boldsymbol{T}} \boldsymbol{A p^{(j)}} = \mathbf{0}$
    从而$\boldsymbol{\alpha_j} = \mathbf{0}$,即证其线性无关
- 超线性收敛和二次终结性常用来讨论算法的优点

第3章 非线性规划-基本思路

◆基本的优化方法如下

初始$x^{(0)} \in S, k=0, \epsilon > 0, k=0$

是否满足停止准则：$\left\|\nabla f\left(x^{(k)}\right)\right\| \leq \epsilon$?

yes

停

no

在$x^{(k)}$点选择可行下降方向$p^{(k)}$

线搜索求出步长λ_k，使得

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

$$f\left(x^{(k)} + \lambda_k p^{(k)}\right) < f\left(x^{(k)}\right)$$

$k = k + 1$

第3章 非线性规划-总体概括

- **从前述可知，搜索方向和步长是其中两个关键的步骤**
- **最优步长(optimal step size)一般没有解析的形式，例如函数不能求导，此时采用近似求解的办法**
  - 首先，确定包含最小值的初始搜索区间
  - 其次，利用某些区间分割技术或者插值技术来迭代减小该区间，直到区间宽度小于预先设定的精度，即得其近似解
- **可分为试探法和插值法(interpolations)**
  - 进退法：初始点$\boldsymbol{x}^{(0)}$和初始步长$\boldsymbol{h}$,以及精度参数$\boldsymbol{\epsilon}$,首先取$\boldsymbol{x}^{(0)}+\boldsymbol{h}$作为新的试探点：若$\boldsymbol{f}\left(\boldsymbol{x}^{(0)}+\boldsymbol{h}\right)<\boldsymbol{f}\left(\boldsymbol{x}^{(0)}\right)$,搜索成功，下一步搜索增大步长为$\boldsymbol{\beta_1 h}, \boldsymbol{\beta_1}>\mathbf{1}$,继续搜索；否则，搜索失败，后退一小步,步长缩小为$\boldsymbol{\beta_2 h}, \boldsymbol{\beta_2}<\mathbf{1}$,继续搜索；最后直到步长参数下降到预先给定的精度时停止。

第3章 非线性规划-总体概括

- 进退法简单，但效率低，可用于求包含函数极小点的区间(称为区间搜索算法)

- **Golden Section(黄金分割法，0.618法),只对单谷函数(unimodal function)**
  - 对区间$[a,b]$,任意计算两个试探点x_1,x_2,按照对称原则:$x_1-a=b-x_2$;等比收缩原则:$\frac{x_2-a}{x_1-a}=\frac{b-a}{x_2-a}\Rightarrow$ 每次保留的区间是原区间的0.618倍

- **插值法，根据原函数在某些试探点的信息，构造一个与$f(x)$近似的函数$\hat{f}(x)$,在一定条件下可以期望$\hat{f}$的极小点会接近$f(x)$的极小点，然后取$\hat{f}$的极小点作为新的试探点，设法缩短搜索区间，一般去二次或三次函数近似**
  - 二次插值和试位法
  - 三次插值法(Davidon搜索法)

第3章 非线性规划-总体概括

- 一阶近似：$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$

$$x_{new} = x_k + \lambda_k p_k$$

- 若x_0是最小值位置的近似，二阶近似：$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2$
- 因此，在靠近x_0的局部区域内，可用$q(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2$来近似$f$
- $0 \approx 0 + f'(x_0) + f''(x_0)(x - x_0)$
- 从而有$x_{new} = x_k - \frac{f'(x_k)}{f''(x_k)}$
- 因此，x_{new}是一个比x_k更好的近似

◆无约束最优化算法分类

- 使用导数的方法，如一阶导数(梯度)，二阶导数(Hessian矩阵)
  - ✓最速下降法：搜索方向为负梯度方向，用一次函数做目标函数的近似，但本身收敛速度在出现锯齿现象时，很慢
  - ✓**Newton**法：以函数的二次近似为基础，具有二次终止性；但若初始点选择不好，远离极小点时，函数值可能会上升，因而不能保证收敛；此时可采用线搜索只搜索增量方向，称为阻尼牛顿法。但不管如何，这两种方法中搜索方向包含**Hessian**矩阵求逆，其为奇异时，无法求解；非奇异时，不能保证方向为下降方向。
  - ✓但如果**Hessian**矩阵正定，则不会出现上述情况，因此可强迫**H**为正定来避免不收敛
    - **Gill-Murray**方法: $\boldsymbol{G_k = H_k + E_k}$
    - **Fletcher-Freeman**方法:采用**Bunch-Parlett**分解，求出负曲率方向。$\boldsymbol{H_k}$通过交换若干行和列后$\boldsymbol{\widetilde{H}_k = LDL'}$,其中$\boldsymbol{L}$为单位下三角，$\boldsymbol{D}$为一阶或二阶对角阵，然后再根据$\boldsymbol{D}$的特征值全为正、有负数或没有负数但有零特征根三种情况分别由负曲率方向确定搜索方向
  - ✓**Newton**法是把求二次函数极小点的方法用于求一般函数的极小点

◆无约束最优化算法分类

- 使用导数的方法，如一阶导数(梯度)，二阶导数(Hessian矩阵)
  - ✓共轭方向法：介于最速下降法和Newton之间，克服了最速下降法的锯齿现象，提高收敛速度，迭代公式简单，不必计算二阶导数，减少了计算量和存储量
    - 共轭梯度法：利用$x^{(k)}$处的梯度来构造共轭方向的算法；但各搜索方向的共轭性依赖于初始负梯度方向，也不能精确计算λ_k和$p^{(k)}$,从而可能多次迭代后不共轭，需要迭代r次后重新初始点后再迭代，如Fletcher-Reeves共轭梯度法
  - ✓拟Newton法(变尺度法)：Newton法的搜索方向$p=-H^{-1}g$,g为$f(x)$的梯度方向，显然新方向需要求Hessian矩阵的逆矩阵，最速下降法中$p=-g$，简单但收敛慢。变尺度法可看作是最速下降法的推广。也可看作是Newton法的推广，因此成为拟Newton法.

- Quasi-Newton法的迭代公式：$x^{(k+1)} = x^{(k)} - H_k^{-1}g_k$,$g_k$, H_k分别表示目标函数$f(x)$在$x^{(k)}$处的梯度和Hessian矩阵. 为消除H_k^{-1},考虑用近似矩阵G_k来代替：$x^{(k+1)} = x^{(k)} - G_k g_k$, 此时$p^{(k)} = G_k g_k$为第$k$次迭代的搜索方向.如果考虑沿$p^{(k)}$方向进行直线搜索，即考虑一般的迭代公式：

$$x^{(k+1)} = x^{(k)} - \lambda_k G_k g_k,$$

这里λ_k是步长因子，此时要求G_k与H_k^{-1}近似且容易计算

  - ✓为保证搜索方向是下降方向，要求G_k正定；
  - ✓要求G_k之间迭代具有简单形式，最简单的显然为$G_{k+1} = G_k + E_k$, E_k称为校正矩阵；
  - ✓G_k必须满足拟Newton性质：$G_{k+1}\Delta g_k = \Delta x_k$ (称为拟牛顿条件或拟牛顿方程)

- 根据拟牛顿方程，可以构造出各种G_k,最有代表性的是DFP

第3章 非线性规划-总体概括

- ➢ DFP变尺度法：要求$G_{k+1} = G_k + \Delta G_k$满足拟Newton条件，得$(G_k + \Delta G_k)\Delta g_k = \Delta x_k$,若能选取向量$u_k$,$v_k$,使其满足规范化条件$u_k^T g_k = v_k^T g_k = 1$,则可以选取$\Delta G_k = \Delta x_k v_k^T - G_k \Delta g_k u_k^T$,故$G_{k+1} = G_k + \Delta x_k v_k^T - G_k \Delta g_k u_k^T$，如选取$u_k^T = (\Delta g_k^T G_k)/(\Delta g_k^T G_k \Delta g_k)$, $v_k^T = (\Delta x_k^T)/(\Delta x_k^T \Delta g_k)$,代入得DFP的迭代公式

◆不使用导数信息，称为直接法

- ➢ 一般使用导数方法收敛速度快，但需计算梯度或Hessian阵；第二类方法不涉及导数，适用性强，但收敛较慢
- ➢ 例如变量轮换法

- 无约束最优化问题的最优性条件
  - 多变量函数$f(x)$,无约束最优化问题: $min\ f(x), f: R^n \to R$
- 极小点的一阶必要条件：$f(x)$连续可微，x^*为极小点，则$\nabla f(x^*) = 0$
- 极小点的二阶必要条件：$f(x)$二次连续可微，x^*为其局部极小点，则$\nabla f(x^*) = 0$，且$\nabla^2 f(x^*)$半正定
- 极小点的二阶充分条件: $f(x)$二次连续可微，$x^* \in R^n$, 若$\nabla f(x^*) = 0$，且$\nabla^2 f(x^*)$正定，则x^*为其严格局部极小点
- 设$f: R^n \to R$为可微凸函数，此时无约束问题为凸规划问题，如果x^*是驻点，则x^*为其最小点；若f严格凸，则驻点一定是其唯一全局最小点

- **有约束最优化问题的最优性条件**
  - 问题 $\boldsymbol{min}\, f(x), s.t.\, g(x) \le 0, h(x) = 0$,约束集 $\boldsymbol{S} = \{x|g(x) \le 0, h(x) = 0\}$
- **等式约束问题的最优性条件，建立等式约束的Lagrange乘子及一阶必要条件，优化问题的拉格朗日函数：**
- $\boldsymbol{L(x, v) = f(x) + v^T h(x), v \in R^l}$**为拉格朗日乘子**
  - 若$\boldsymbol{x^*}$为函数$\boldsymbol{f(x)}$的极小值点，则$\boldsymbol{\nabla f(x^*) = \sum_{i=1}^{l} (-v_i^*) \nabla h_i(x^*)}$
- **注意乘子的意义：表示**$\boldsymbol{f(x^*)}$**对各阶约束分量扰动的敏感程度**

思考：红色字体部分的几何意义？

第3章 非线性规划-最优性约束条件

◆一般约束问题的最优性条件

- 考虑只含不等式约束的最优化问题$min\ f(x), s.t.\ g(x) \le 0$；所有可行集中，令不等式约束中为等式约束的分量$g_i$的集合为$x^*$的起作用集或者紧约束集

◆**最优性定理：Kuhn-Tucker条件，设$x^* \in S, I$为x^*点的紧约束集。函数$f(x), g_i(x), i \in I$在x^*点可微，$g_i(x), i \notin I$在x^*点连续，向量组$\{\nabla g_i(x^*) | i \in I\}$线性无关。如果$x^*$为不等式约束最优化问题的局部极值点，则存在$u_i^* \ge 0, i \in I$，使$\nabla f(x^*) + \sum_{i\in I} u_i^* \nabla g_i(x^*) = 0$**

◆**该定理结合等式约束可推广到一般情况**

◆**K-T条件的几何意义是目标函数的负梯度可表示成紧约束集函数梯度的非负组合。K-T点相当于无约束中的驻点**

- **而对有约束凸优化问题的最优性条件而言：前述条件是其充要条件**
- **约束最优化问题的二阶条件**
  - 无约束问题中
    - ✓必要条件：要求$H(f(x^*))$半正定
    - ✓充分条件：要求$H(f(x^*))$正定
  - 有约束问题，只考虑$H(f)$是不够的
    - ✓考虑例子(Fiacco and MeCormick,1968)
    - ✓ $$\begin{cases} \min f(x) = \frac{1}{2}\left[(x_1-1)^2 + x_2^2\right] \\ s.t.\ h(x) = -x_1 + \beta x_2^2 = 0, \beta\text{为常数} \end{cases}$$
  - 请听后续分解！

第3章 非线性规划数学模型-总结

- ◆如何建立模型
- ◆最优性条件
  - ➢多元函数的极值条件
  - ➢无约束优化问题的最优性条件
  - ➢凸规划问题的最优性条件
  - ➢优化问题的解决思路：迭代方法之搜索方向和搜索步长
  - ➢有约束优化问题的最优性条件简介

◆ **Epigraph(上境图)**

$$epi f = \{(x,u): x \in R^n, \mu \in R, \mu \geq f(x)\}$$

◆ **函数$f(x)$为凸等价于其epigraph为凸**

◆ **Strict Epigraph(严格上镜图)**

$$epi_S f = \{(x,u): x \in R^n, \mu \in R, \mu > f(x)\} \subseteq R^{n+1}$$

◆ **Jensen不等式，对于凸函数$f(x) \in R, x \in R^n$, 如果 $a_i \geq 0$,则有**

$$f\left(\sum_i (a_i x_i) \backslash (\sum_i a_i)\right) \leq \sum_i a_i f(x_i) \backslash (\sum_i a_i)$$

◆ **推广: $f(Ex) \leq E(f(x))$**

- 函数亚图(hypograph,subgraph，下镜图)

$$hypf = \{(x, u): x \in R^n, \mu \in R, \mu \leq f(x)\}$$

- 函数$f(x)$为凹的等价于其hypograph为凸
- Strict hypograph(严格亚图，严格下镜图)

$$hyp_S f = \{(x, u): x \in R^n, \mu \in R, \mu < f(x)\} \subseteq R^{n+1}$$

- (水平集(level set))函数f定义在非空凸集$S \subset R^n$上的凸函数，α是一个实数，则水平集$L_\alpha = \{x | x \in S, f(x) \leq \alpha\}$是凸集，有时也用$H_S(f, \alpha)$表示，表示集合$S$上关于数$\alpha$的水平集

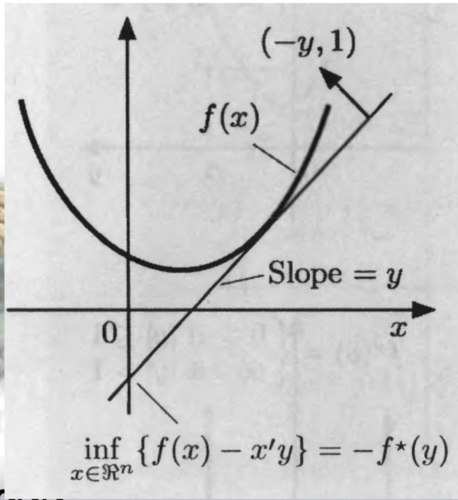
第3章非线性规划-附注几个基本概念

- (次梯度(subgradient))多元函数f定义在非空开凸集$S \subset R^n$上的凸函数，但f不一定处处可微。此时采用次梯度的概念来类比梯度的概念。函数f在x_0处的次梯度是一个向量v,如果对于S内的任意x，都有:$f(x) - f(x_0) \geq v \cdot (x - x_0)$成立，所有次梯度的集合称为次微分(subdifferential)，记为$\partial f(x_0)$
- 如果为一维，则可称为次导数(subderivative),例如凸函数$f(x) = |x|$,其在原点的次微分是区间$[-1, 1]$,而在$x_0 < 0$,$x_0 > 0$的次微分分别为单元素集合$\{-1\}, \{1\}$
- 又如$f(x) = max\{0, \frac{1}{2}(x^2 - 1)\}$，练习求出其次梯度
- 函数f在x_0以$\nabla f(x_0)$可微,当且仅当它有$\nabla f(x_0)$作为在x_0的唯一次梯度
- 注意，次微分是非空的，凸的，和紧的。

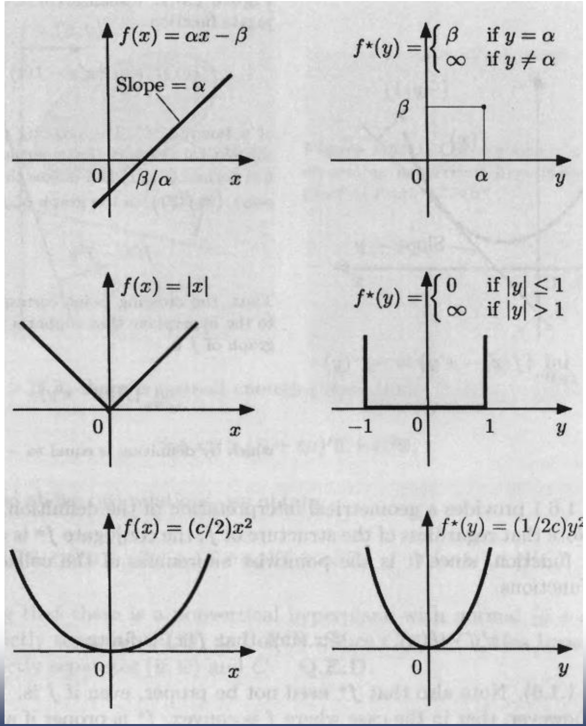
- **共轭函数(Conjugate function)**
  - 若函数$f: R^n \rightarrow [-\infty, \infty]$,其共轭函数$f^*: R^n \rightarrow [-\infty, \infty]$定义为：

$$f^*(y) = sup_{x \in R^n}\{x'y - f(x)\}$$

- **无论函数f是否为凸函数，其共轭函数f^*总是闭的凸函数**
- **下图为示意图**

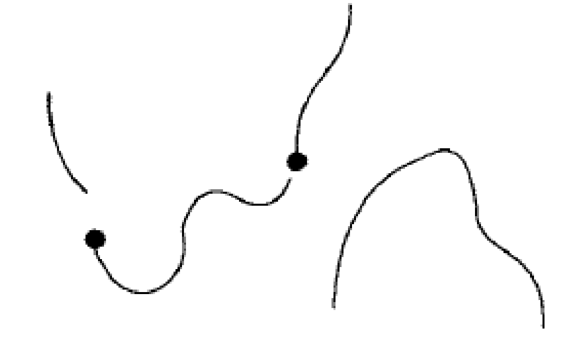


右图为原函数和对应的共轭函数



第3章 非线性规划-附注几个基本概念

- 凸函数的要求比较严格，在解决实际问题的时候，不一定要求函数是凸函数，一些具有跟凸、凹函数类似的性质的函数，可能具有凸函数的某些重要性质
- **Quasiconvex 函数(拟凸函数)**:函数f是定义在非空凸集$S \in R^n$上的拟凸函数，如果对每个$x_1, x_2 \in S$, 下列不等式成立：$0 < \lambda < 1, f(\lambda x_1 + (1-\lambda)x_2) \le max\{f(x_1), f(x_2)\}$,类似的，如果$-f$是拟凸函数，则$f$是拟凹函数。如果是**严格拟凸函数**，则要求$f(x_1) \neq f(x_2)$, 且不等式严格成立,如下图(a)是拟凸,(b)是拟凹
- **Pseudoconvex 函数(伪凸函数)**

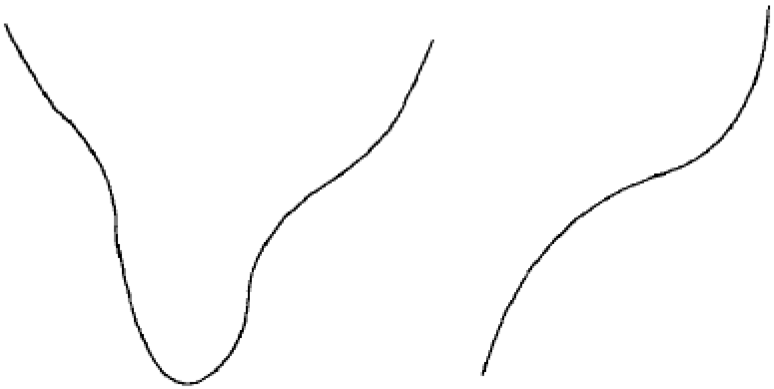


(a) (b)

◆Pseudoconvex 函数(伪凸函数)

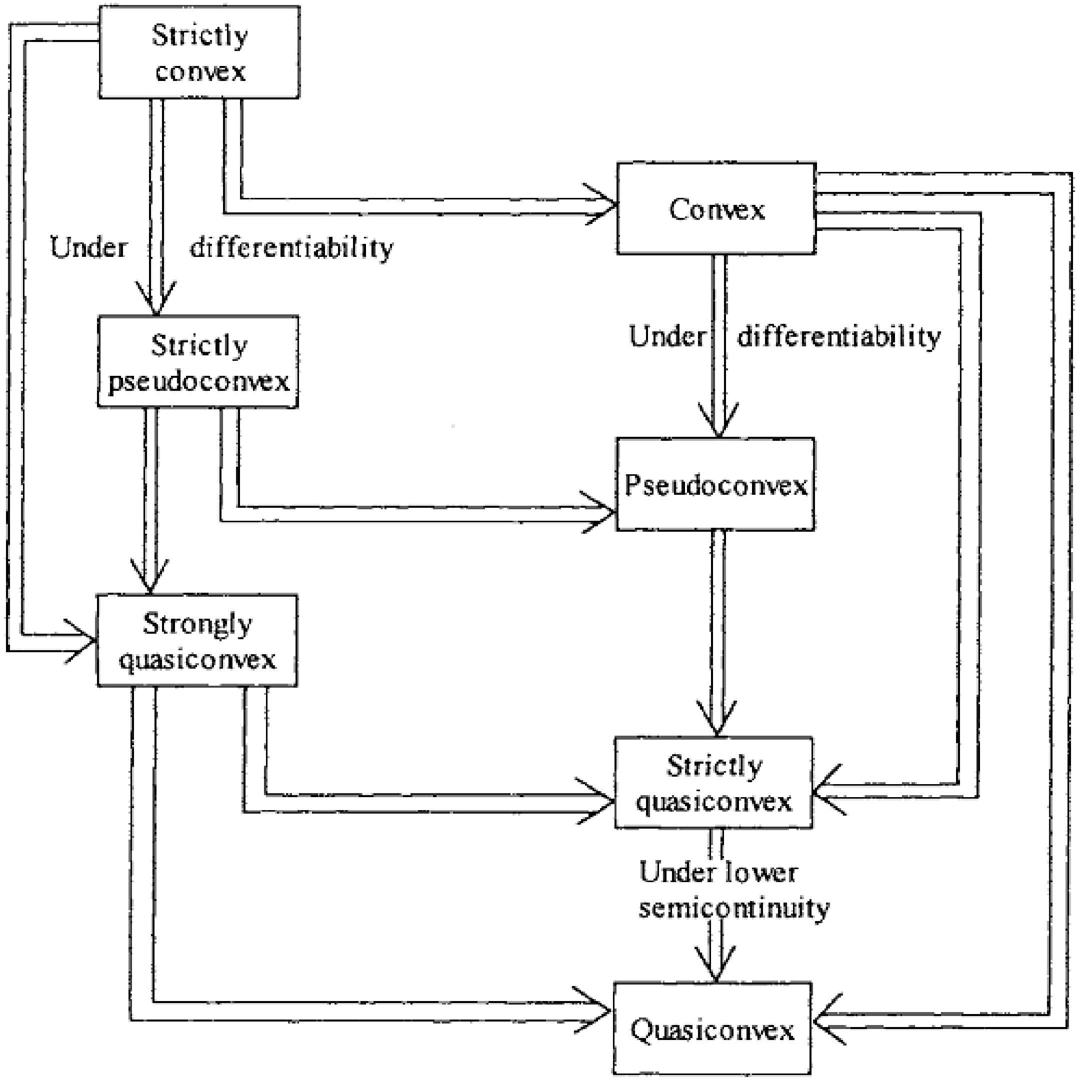
- 函数f是定义在非空开集$S\in R^n$上的可微函数，如果对每个$x_1,x_2\in S$, 下列关系成立：若$\nabla f(x_1)^T(x_2-x_1)\geq 0$,则$f(x_2)\geq f(x_1)$,或等价地，如果$f(x_2)<f(x_1)$,则$\nabla f(x_1)^T(x_2-x_1)<0$，则函数$f$称为是伪凸的。如果$-f$拟是伪凸的，则$f$是伪凹函数。如果是**严格伪凸函数**，则要求$x_1,x_2\in S$是不同的点,如下图(a)是伪凸,(b)是伪凹也是伪凹
- 由定义，可知若$\forall\bar{x},\nabla \mathrm{f}(\bar{x})=0$

则$f(x)\geq f(\bar{x})$,因此为全局极小点



(a) (b)

◆凸．拟凸．伪凸之间的关系



在R上考虑。在凸要求不高过xy连线的地方，拟凸只要求不高过xy中较大者，所以凸蕴含拟凸；而伪凸则要求，“切线”“指哪”函数“打哪”，即导数（或者更一般地，上Dini导数）非负则向右非减、非正则向左非减，也即所有的驻点都是极小值点，伪凸蕴含拟凸