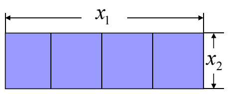
第三章　非线性规划

§1　问题的提出

例1. 某单位拟建一排厂房，厂房建筑平面如图所示。由于资金及材料的限制，围墙及隔墙的总长度不能超过80米。为使建筑面积最大，应如何选择长宽尺寸？



例2. 设某物理过程具有如下规律

$$\varphi(t) = x_1 + x_2 e^{-x_3 t}$$

用试验法求得 t_i 时的值 $\varphi(t_i)$，$i = 1, \ldots, m$。现要确定参数 x_1, x_2, x_3，使所得试验点构成的曲线与理论曲线误差平方和为最小，且满足

$$\begin{cases} x_1 + x_2 = 1 \\ x_3 \geq 0 \end{cases}$$

非线性规划的数学模型

$$\min_{x \in X} f(x)$$

$X = \mathrm{R}^n$，称为无约束问题；

$X \subseteq \mathrm{R}^n$，称为有约束问题，X 称为可行域。

标准型：

$$\begin{aligned} &\min f(x) \\ &g_i(x) \geq 0,\ i = 1,2,\cdots,m \\ &h_j(x) = 0,\ \ j = 1,2,\cdots,l \end{aligned}$$

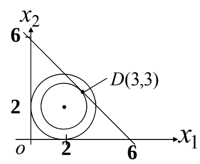
图解法

例3. 求解如下非线性规划问题

$$\begin{aligned} &\min f(x) = (x_1 - 2)^2 + (x_2 - 2)^2 \\ &\text{s.t}\quad x_1 + x_2 - 6 = 0 \end{aligned}$$

最优解　$x_1{}^* = x_2{}^* = 3$

$\min f(x) = f(x^*) = 2$



局部最优解与全局最优解

定义

(1) 若 $x^* \in X$, 且 $\min_{x \in X} f(x) = f(x^*)$, 称 x^* 为全局最优解。

(2) 若 $x^* \in X$, 存在 $\delta > 0$，使 $\min_{x \in N(x^*, \delta) \cap X} f(x) = f(x^*)$,

称 x^* 为局部最优解。

§2 无约束问题的最优性条件

设 $u = f(x),\ x = (x_1, \ldots, x_n)^T \in S \subseteq \mathrm{R}^n$,

梯度　$\nabla f(x) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n})^T$

Hesse矩阵

$$\nabla^2 f(x) = H(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

例1. 设 $f(x)=b^T x=\sum_{i=1}^{n} b_i x_i$, 求 $\nabla f(x)$.

例2. 设 A为 n 阶对称方阵，$f(x)=x^T Ax$，求 $\nabla f(x), \nabla^2 f(x)$.

多元函数的Talor公式

$$f(x_0+\Delta x)=f(x_0)+\nabla f(x_0)^T \Delta x+o(\|\Delta x\|)$$

$$f(x_0+\Delta x)=f(x_0)+\nabla f(x_0)^T \Delta x+\frac{1}{2}(\Delta x)^T \nabla^2 f(x_0+\theta x)\Delta x \quad 0<\theta<1$$

$$f(x_0+\Delta x)=f(x_0)+\nabla f(x_0)^T \Delta x+\frac{1}{2}(\Delta x)^T \nabla^2 f(x_0)\Delta x+o(\|\Delta x\|^2)$$

§3 凸函数与凸规划

定义. 设 $f(x)$定义在 R^n 的某个凸集 S上，若 $\forall 0<\alpha<1$, $x^{(1)}, x^{(2)}\in S\ (x^{(1)}\neq x^{(2)})$，有

$$f(\alpha x^{(1)}+(1-\alpha)x^{(2)})\le \alpha f(x^{(1)})+(1-\alpha)f(x^{(2)})$$

则称$f(x)$为 S上的凸函数。

严格凸函数

$$f(\alpha x^{(1)}+(1-\alpha)x^{(2)})<\alpha f(x^{(1)})+(1-\alpha)f(x^{(2)})$$

性质1. 设 $f(x)$是凸集 S上的凸函数，则 $\forall \alpha\ge 0, \alpha f(x)$ 也是 S上的凸函数。

性质2. 设 $f_1(x), f_2(x)$是凸集 S上的凸函数，则 $f_1(x)+f_2(x)$ 也是 S上的凸函数。

定义. 称 $H_S(f,\beta)=\{x\mid x\in S, f(x)\le\beta\}$ 为 $f(x)$在S上关于β 的水平集。

性质3. 设$f(x)$是凸集 S上的凸函数，则 $\forall\beta\in R$, $H_S(f,\beta)$ 也是凸集 。

性质4. 设 $f(x)$是凸集 S上的凸函数，则$f(x)$的任一极小点就是S上的全局极小点，且所有极小点形成一个凸集。

凸函数的判别准则

定理1. 设 $f(x)$在凸集 S上有一阶连续偏导数，则 $f(x)$为S上的凸函数 $\Leftrightarrow \forall\ x^{(1)}, x^{(2)}\in S$，有

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

注. 将 $\ge$ 改为 $>$ 就是严格凸函数的充要条件。

几何意义. 任一点处的切线增量不超过函数的增量。

定理2. 设 $f(x)$在开凸集 S上有二阶连续偏导数，则 $f(x)$为S上的凸函数 $\Leftrightarrow \nabla^2 f(x)$ 处处半正定。

注. $\nabla^2 f(x)$处处正定，则 $f(x)$是严格凸函数。(非必要)

例1. $f(x)=x_1^2+x_2^2$ 在R^2上是凸函数。

定理3. 设 $f(x)$在凸集 S上的可微凸函数，若 $\exists x^*\in S, \forall x\in S$, $\nabla f(x^*)^T(x-x^*)\ge 0$, 则$x^*$是 f 在S上的全局极小点。

几何意义. 向量 $x-x^*$与 $\nabla f(x^*)$ 夹角不超过90° .

定理4. 设 $f(x)$在凸集 S上的可微凸函数，若 $\nabla f(x^*)=0$, 则 x^*为 $f(x)$ 在 S 上的全局极小点。

凸规划

目标函数是凸函数，可行域是凸集的规划问题。

例如：

$$\begin{aligned} &\min f(x) \\ &\text{s.t } g_i(x) \le 0,\ i = 1,2,\cdots,m \\ &\quad h_j(x) = 0,\ j = 1,2,\cdots,l \end{aligned}$$

其中，f, g_i为凸函数，h_j为线性函数。

§4 解非线性规划的基本思路

对大多数实际问题，用最优性条件求解困难。

数值计算：

$$x^{(k)} \to x^{(k+1)},\ \{x^{(k)}\}$$

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

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

其中，$\lambda_k > 0$ 称为步长

$\| p^{(k)} \| = 1$ 称为搜索方向

定义 设$f: \mathrm{R}^n \to \mathrm{R}$, $\bar{x} \in \mathrm{R}^n, 0 \ne p \in \mathrm{R}^n$, 若$\exists \delta > 0$,

$\forall \lambda \in (0, \delta)$有

$$f(\bar{x} + \lambda p) < f(\bar{x})$$

则称 p 是f在 $\bar{x}$ 的下降方向。

问题: 下降方向不止一个, 如何求下降最快的方向?

定义 设 $X \subseteq \mathrm{R}^n$，$\bar{x} \in X, 0 \ne p \in \mathrm{R}^n$，若$\exists \delta > 0$,

$\forall \lambda \in [0, \delta]$有 $\bar{x} + \lambda p \in X$

则称 p 是$\bar{x}$ 处关于X的可行方向。

迭代算法一般步骤：

(1) 选取初始点 $x^{(0)}$.

(2) 构造搜索方向 $p^{(k)}$.

(3) 确定步长 λ_k.

(4) 求出下一个点 $x^{(k+1)}$.

满足 $f(x^{(k)} + \lambda_k p^{(k)}) = \min\limits_{\lambda} f(x^{(k)} + \lambda p^{(k)})$

的步长 λ_k 称为**最优步长**。

求最优步长的过程，称为**一维搜索**（线性搜索）。

定理. 设 $f(x)$有一阶连续偏导数, $x^{(k+1)}$ 满足

$$\begin{cases} f(x^{(k)} + \lambda_k p^{(k)}) = \min\limits_{\lambda} f(x^{(k)} + \lambda p^{(k)}) \\ x^{(k+1)} = x^{(k)} + \lambda_k p^{(k)} \end{cases}$$

则 $\nabla f(x^{(k+1)})^T p^{(k)} = 0$.

迭代终止条件

(1) $\| x^{(k+1)} - x^{(k)} \| < \varepsilon_1$，或 $\dfrac{\| x^{(k+1)} - x^{(k)} \|}{\| x^{(k)} \|} < \varepsilon_2$

(2) $f(x^{(k)}) - f(x^{(k+1)}) < \varepsilon_3$,

或 $\dfrac{f(x^{(k)}) - f(x^{(k+1)})}{| f(x^{(k)}) |} < \varepsilon_4$

(3) $\| \nabla f(x^{(k+1)}) \| < \varepsilon_5$.

§5 一维搜索

若记 $\varphi(\lambda)=f(x^{(k)}+\lambda p^{(k)})$, 理论上 λ_k 应满足 $\varphi'(\lambda_k)=0$,
但对许多问题，求导并不容易，且不便于用计算机求解。

一般求解 λ_k 的近似值：

主要方法：
- 区间收缩法　黄金分割法、加步探索法
- 函数逼近法　牛顿法、抛物线法

1、黄金分割法

定义. 设函数 $\varphi(\lambda)$: $R \to R$, 闭区间 $[a_0, b_0]$, 若存在 $\lambda^* \in [a_0, b_0]$, 使 $\varphi(\lambda)$ 在 $[a_0, \lambda^*]$ 上严格递减，在 $[\lambda^*, b_0]$ 上严格递增，则称 $\varphi(\lambda)$ 为 $[a_0, b_0]$ 上的单谷函数，$[a_0, b_0]$ 为单谷区间。

性质. 设 $\varphi(\lambda)$ 为 $[a_0, b_0]$ 上的单谷函数，λ^* 为极小点，

$a_1, b_1 \in [a_0, b_0]$, $a_1 < b_1$

(1) 若 $\varphi(a_1) < \varphi(b_1)$, 则 $\lambda^* \in [a_0, b_1]$

(2) 若 $\varphi(a_1) > \varphi(b_1)$, 则 $\lambda^* \in [a_1, b_0]$

0.618法基本原理与步骤

(1) 在单谷区间 $[a_0, b_0]$ 中找出两个试点 x_1, x_1', 比较 $\varphi(x_1)$ 和 $\varphi(x_1')$ 的大小，将区间缩小为 $[a_1, b_1]$。

(2) 在 $[a_1, b_1]$ 继续找两个试点 x_2, x_2' 后, 比较 $\varphi(x_2)$ 和 $\varphi(x_2')$ 的大小，将区间缩小为 $[a_2, b_2]$。

(3) 重复上述过程直到 $[a_k, b_k]$ 的区间长度足够小，并达到了事先规定的精度，令 $\lambda^* = (a_k + b_k)/2$.

这一系列的点 x_k, x_k' 如何确定？

取黄金分割点及其对称点

$$x_1 = a_0 + 0.618(b_0 - a_0)$$

$$x_1' = a_0 + 0.382(b_0 - a_0)$$

算法步骤

(1) 确定初始搜索区间 $[a_0, b_0]$，最后区间精度 δ.

(2) 计算黄金分割点 x_{k+1} 及对称点 x_{k+1}' , 令 $k := 0$.

(3) 若 $\varphi(x_{k+1}') \le \varphi(x_{k+1})$, 转(4); 若 $\varphi(x_{k+1}') > \varphi(x_{k+1})$, 转(5).

(4) 令 $a_{k+1} = a_k$, $b_{k+1} = x_{k+1}$. 若 $\dfrac{b_{k+1}-a_{k+1}}{b_0-a_0} < \delta$, 结束. 否则
令 $x_{k+2} = x_{k+1}$, $x_{k+2}' = a_{k+1} + 0.382(b_{k+1} - a_{k+1})$, $k = k+1$. 转(3).

(5) 令 $a_{k+1} = x_{k+1}'$, $b_{k+1} = b_k$. 若 $\dfrac{b_{k+1}-a_{k+1}}{b_0-a_0} < \delta$, 结束. 否则
令 $x_{k+2}' = x_{k+1}$, $x_{k+2} = a_{k+1} + 0.618(b_{k+1} - a_{k+1})$, $k = k+1$. 转(3).

例. 求 $\min\limits_{\lambda \ge 0} \varphi(\lambda) = \lambda^3 - 2\lambda + 1$ 的近似最优解.

初始搜索区间 $[0, 3]$, 精度 $\delta = 0.15$.

2、加步探索法

基本思想：从一点出发，确定函数值“高—低—高”的3个点.

即找到 $x_1 < x_2 < x_3$, 满足 $f(x_1) > f(x_2), f(x_2) < f(x_3)$.

步骤

(1) 给定初始点 x_1, 初始步长 $h_0 > 0$.

(2) 找出初始区间　　令 $x_2 = x_1 + h_0$

① 若 $f(x_2) < f(x_1)$, 步长加倍, 令 $x_3 = x_2 + 2h_0$, 若仍有 $f(x_3) < f(x_2)$, 步长再加倍, 令 $x_4 = x_3 + 4h_0$, …, 直到点 x_k 的函数值刚刚变为增加为止。

② 若 $f(x_2) > f(x_1)$, 令 $x_3 = x_1 - h_0$, 若 $f(x_3) < f(x_2)$, 步长再加倍, 令 $x_4 = x_3 - 2h_0$, …, 直到点 x_k 的函数值刚刚变为增加为止。

(3) 缩小区间　　令 $x_{k+1} = \frac{1}{2}(x_{k-1} + x_k)$,

则 $x_{k-2}, x_{k-1}, x_{k+1}, x_k$ 是等距的4个点。

令 $f(x_{k-2}), f(x_{k-1}), f(x_{k+1}), f(x_k)$ 最小的点为 x_2, 左右邻点分别为 x_1, x_3, 则

$$x_1 < x_2 < x_3,\ \ f(x_1) > f(x_2), f(x_2) < f(x_3).$$

例. 求 $\min\limits_{x \ge 0} f(x) = x^3 - 2x + 1$ 的搜索区间，

取 $x_1 = 0$, $h_0 = 1$.

3、牛顿法

基本思想：在极小点附近用二阶泰勒多项式近似 $f(x)$

设已有 $f(x)$ 极小点的第 k 级估计，则

$$f(x) = f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) + \frac{1}{2} f''(x^{(k)})(x - x^{(k)})^2 + o((x - x^{(k)})^2)$$

$$\approx f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) + \frac{1}{2} f''(x^{(k)})(x - x^{(k)})^2$$

$$:= \varphi(x)$$

用 $\varphi(x)$ 的极小点来近似 $f(x)$ 的极小点。

令 $\varphi'(x) = f'(x^{(k)}) + f''(x^{(k)})(x - x^{(k)}) = 0$,

得

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}$$

可以证明，$\{x^{(k)}\}$ 在一定条件下收敛于 $f(x)$ 的极小点。

算法步骤

(1) 给定初始点 $x^{(1)}$，精度 $\varepsilon > 0$, 令 $k := 1$.

(2) 计算 $f'(x^{(k)})$ 及 $f''(x^{(k)})$，若 $|f'(x^{(k)})| < \varepsilon$, 停止迭代，得到近似极小点 $x^{(k)}$，否则转(3).

(3) 计算 $x^{(k+1)}$, 令 $k := k+1$, 转(2).

例. 求 $\min f(x)=\int_0^x \arctan t dt$,

取 $x^{(1)}=1$, $\varepsilon=0.01$.

4、抛物线法

基本思想：在极小点附近用二次三项式逼近 $f(x)$.

设 $x_1<x_2<x_3$, $f(x_1)>f(x_2), f(x_2)<f(x_3)$.

令 $\varphi(x)=ax^2+bx+c$, 且

$$\begin{cases} ax_1^2+bx_1+c=f(x_1) \\ ax_2^2+bx_2+c=f(x_2) \\ ax_3^2+bx_3+c=f(x_3) \end{cases}$$

解出 a , b, c, 得到$\varphi(x)$.

令 $\varphi'(x)=2ax+b=0$, $\quad x^*=-\dfrac{b}{2a}$

以 x^* 近似 $f(x)$的极小点 $x^{(k)}$，从 $x_1, x_2, x_3, x^{(k)}$ 中，选择函数值最小的点作为新的x_2, 左、右邻点为新的x_1, x_3, 继续求 $x^{(k+1)}$.

算法终止条件：

$|f(x^{k+1})-f(x^k)|<\varepsilon$ 或 $|x^{(k+1)}-x^{(k)}|<\delta$.