第4章 非线性规划

基本概念

非线性规划

1. 基本概念
2. 凸函数和凸规划
3. 一维搜索方法
4. 无约束最优化方法
5. 约束最优化方法

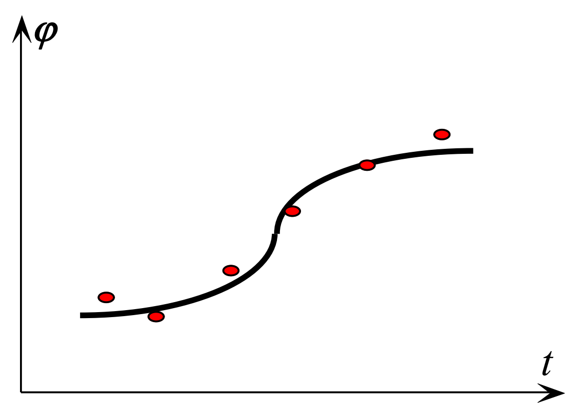
4.1 基本概念

例1，曲线的最优拟合

已知某物体的温度φ与时间 t 之间有如下形式的经验函数关系：

$$\varphi = c_1 + c_2 t + e^{c_3 t} \qquad (*)$$

其中 c_1，c_2，c_3 是待定参数。现通过测试获得 n 组φ与 t 之间的实验数据 (t_i, φ_i)。试确定参数 c_1，c_2，c_3，使理论曲线(*)尽可能地与 n 个测试点 (t_i, φ_i) 拟合。



例1，曲线的最优拟合

解：根据最小二乘法原理，计算（优化）

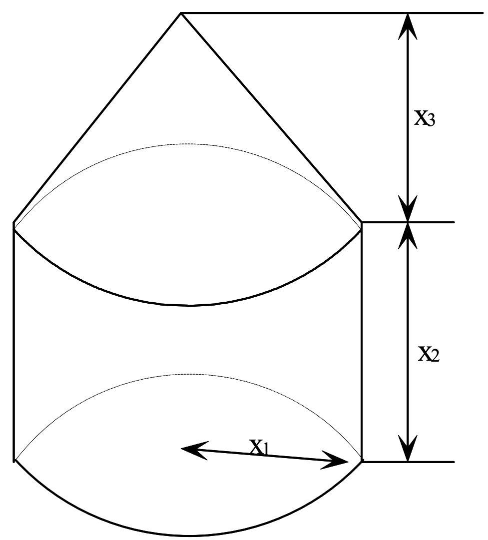
$$\min \sum_{i=1}^{n}\left[\varphi_i-\left(c_1+c_2 t+e^{c_3 t_i}\right)\right]^2,$$

其中 c_1，c_2，c_3 为变量。

例2，构件容积

有图所示的由圆锥和圆柱面围成的构件，构件的表面积为 S，圆锥部分的高 h 和圆柱部分的高 x_2 之比为 a。

给定 S 大小的原材料和参数 a，要求确定构件其他部分的尺寸，使其容积最大。



例2，构件容积

- **解：圆锥体的高 $h = ax_2$，其表面积为** $\pi x_1\sqrt{x_1^2 + a^2x_2^2}$ **。圆柱体的侧面面积为 $2\pi x_1x_2$，底面的面积为** πx_1^2 **。**
- **圆锥体的体积为** $\frac{1}{3}\pi x_1^2 ax_2$ **，圆柱体的体积为** $\pi x_1^2 x_2$ **。**
- **本题即为给定表面积 S，使体积 V 最大，即计算如下的优化问题。**

$$
\begin{aligned}
\max \quad & \frac{1}{3}\pi x_1^2 ax_2 + \pi x_1^2 x_2 \\
\text{s.t.} \quad & \pi x_1\sqrt{x_1^2 + a^2x_2^2} + 2\pi x_1x_2 + \pi x_1^2 = S \\
& x_1 \ge 0, x_2 \ge 0
\end{aligned}
$$

数学规划

- 设 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in R^n$，$\boldsymbol{f(x)}$，$\boldsymbol{g_i(x), i=1..p}$ 和 $\boldsymbol{h_j(x), j=1..q}$ 是 $R^n \mapsto R$ 的函数。如下的数学模型称为数学规划 **(Mathematical Programming, MP)**：

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

- 称 $X=\left\{x\in R^n \,\middle|\, \begin{array}{l} g_i(x)\le 0, i=1,\cdots,p \\ h_j(x)=0, j=1,\cdots,q \end{array}\right\}$ 为约束集或可行域。
- X 中的点 x 称为可行点或可行解。

无约束最优化问题和约束最优化问题

- 当 $p = 0, q = 0$ 时，MP 退化为 $\min f(x)$，称为无约束最优化问题。

- 对应地，一般的 MP
$$\begin{cases} \min \quad f(x) \\ \text{s.t.} \ \ g_i(x) \le 0, i = 1, \dots, p \\ \qquad\quad h_j(x) = 0, j = 1, \dots, q \end{cases}$$
称为约束最优化问题。

整体（全局）最优解

定义 4.1.1 对于非线性规划(MP)，若 $x^* \in X$，并且有

$$f(x^*) \le f(x), \quad \forall x \in X$$

则称 x^*是(MP)的整体最优解（或整体极小点），称 $f(x^*)$是(MP)的整体最优值（或整体极小值）。

如果有

$$f(x^*) < f(x), \quad \forall x \in X, x \ne x^*$$

则称 x^*是(MP)的严格整体最优解（或严格整体极小点），称 $f(x^*)$是(MP)的严格整体最优值（或严格整体极小值）。

局部最优解

定义 4.1.2 对于非线性规划 MP，若 $x^* \in X$，并且存在 x^* 的一个领域 $N_\delta(x^*) = \{x \in R^n \mid \|x - x^*\| < \delta\}$，$\delta > 0$，$\delta \in R$，使得 $\forall\ x \in N_\delta(x^*) \cap X$，都有

$$f(x^*) \le f(x),$$

则称 x^*是 MP 的局部最优解（或局部极小点），称 $f(x^*)$ 是 MP 的局部最优值（或局部极小值）。

如果 $\forall\ x \in N_\delta(x^*) \cap X$，$x \neq x^*$，都有

$$f(x^*) < f(x),$$

则称 x^*是 MP 的严格局部最优解（或严格局部极小点），称 $f(x^*)$是 MP 的严格局部最优值（或严格局部极小值）。

非线性规划方法，基本概念

定义 4.1.3 设 $f:\mathrm{R}^n \mapsto \mathrm{R}$， $\bar{x}\in\mathrm{R}^n$， $p\in\mathrm{R}^n, p\neq 0$。若存在δ > 0，使得

$$\boldsymbol{f(\bar{x}+tp)<f(\bar{x}),\quad \forall t\in(0,\delta)},$$

则称向量 $\boldsymbol{p}$ 是函数 $\boldsymbol{f(x)}$ 在点 $\bar{\boldsymbol{x}}$ 处的下降方向。

定义 4.1.4 设 $X\subset\mathrm{R}^n$， $\bar{x}\in X$， $p\in\mathrm{R}^n, p\neq 0$。若存在 $\boldsymbol{t>0}$，使得

$$\boldsymbol{\bar{x}+tp\in X},$$

则称向量 $\boldsymbol{p}$ 是函数 $\boldsymbol{f(x)}$ 在点 $\bar{\boldsymbol{x}}$ 处关于 $\boldsymbol{X}$ 的可行方向。

非线性规划方法的基本迭代格式

1 选取初始点 x_0，$k \leftarrow 0$。

2 构造搜索方向 p^k。

3 根据 p^k，确定步长 t_k。

4 $x^{k+1} \leftarrow x^k + t_k p^k$。

5 若 x^{k+1} 已满足某种终止条件，则停止迭代，输出 x^{k+1} 作为近似解。否则 $k \leftarrow k+1$，转回第 2 步。

4.2 凸函数和凸规划

凸函数（convex function）

定义 4.2.1 设 $S \subset \mathbf{R}^n$ 是非空凸集，$f: S \mapsto \mathbf{R}$，如果对任意的α $\in (0, 1)$，$\forall x^1, x^2 \in S$ 有

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

则称 f 是 S 上的凸函数。

如果对于任意的α $\in (0, 1)$，$\forall x^1, x^2 \in S$，$x^1 \neq x^2$ 有

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

则称 f 是 S 上的严格凸函数。

若 $-f$ 是 S 上的（严格）凸函数，则称 f 是 S 上的（严格）凹函数。

数乘和加之后，仍是凸函数

定理 4.2.1 设 $S \subset \mathbf{R}^n$ 是非空凸集。

（1） 若 $f: \mathbf{R}^n \mapsto \mathbf{R}$ 是 S 上的凸函数，$\alpha \geq 0$，则 αf 是 S 上的凸函数；

（2） 若 $f_1, f_2: \mathbf{R}^n \mapsto \mathbf{R}$ 都是 S 上的凸函数，则 $f_1 + f_2$ 是 S 上的凸函数。

“截”凸函数得到凸集

定理 4.2.2 设 $S \subset \mathbf{R}^n$ 是非空凸集，$f: \mathbf{R}^n \mapsto \mathbf{R}$ 是凸函数，$c \in \mathbf{R}$，则集合 $H_S(f,c) = \{x \mid x \in S, f(x) \le c\}$ 是凸集。

- 证：任取 $x^1, x^2 \in H$，则 $x^1, x^2 \in S$。
- 任取 $\alpha \in (0, 1)$，因为 S 是凸集，故 $\alpha x^1 + (1-\alpha)x^2 \in S$。 (1)
- 而 $f(\alpha x^1 + (1-\alpha)x^2) \le \alpha f(x^1) + (1-\alpha)f(x^2)$

$$\le \alpha c + (1-\alpha)c \le c。 \tag{2}$$

- (1), (2) $\Rightarrow \alpha x^1 + (1-\alpha)x^2 \in H$，即，$H$ 是凸集。

凸函数的判别

定理 4.2.3 设 $S \subset \mathbf{R}^n$ 是非空开凸集，$f: S \mapsto \mathbf{R}$ 可微，则

（1） f 是 S 上的凸函数，当且仅当

$$\nabla f(x^1)^{\mathrm{T}}(x^2 - x^1) \le f(x^2) - f(x^1), \quad \forall x^1, x^2 \in S;$$

（2） f 是 S 上的严格凸函数，当且仅当

$$\nabla f(x^1)^{\mathrm{T}}(x^2 - x^1) < f(x^2) - f(x^1), \quad \forall x^1, x^2 \in S, x^1 \ne x^2,$$

其中 $\nabla f(x^1) = \left(\dfrac{\partial f(x^1)}{\partial x_1}, \cdots, \dfrac{\partial f(x^1)}{\partial x_n} \right)^{\mathrm{T}}$ 是函数 f 在点 x^1 处的一阶导数（即，梯度）。

凸函数的判别

- 证：(2)的证明与(1)类似，下面仅证(1)。
- (⇒)。f 是 S 上的凸函数。由定义，$\forall \alpha \in (0, 1)$，$\forall x^1, x^2 \in S$，有：$f(\alpha x^2 + (1-\alpha)x^1) \le \alpha f(x^2) + (1-\alpha)f(x^1)$。
- 所以，$\dfrac{f\left(x^1+\alpha\left(x^2-x^1\right)\right)-f\left(x^1\right)}{\alpha} \le f\left(x^2\right)-f\left(x^1\right)$。
- 由多元函数的 Taylor 展开式，有：

$$f\left(x^1+\alpha\left(x^2-x^1\right)\right)=f\left(x^1\right)+\nabla f\left(x^1\right)\left(\alpha\left(x^2-x^1\right)\right)+o\left(\left\|\alpha\left(x^2-x^1\right)\right\|\right)。$$

- 因此，$f\left(x^1+\alpha\left(x^2-x^1\right)\right)-f\left(x^1\right)=\alpha\nabla f\left(x^1\right)\left(x^2-x^1\right)+R_1$，其中 $R_1=o\left(\left\|\alpha\left(x^2-x^1\right)\right\|\right)$。
- 于是，$\nabla f\left(x^1\right)^{\mathrm{T}}\left(x^2-x^1\right)+\dfrac{R_1}{\alpha} \le f\left(x^2\right)-f\left(x^1\right)$。

凸函数的判别

- 两端取极限 $\lim\limits_{\alpha\to 0^+}$，就得到 $\nabla f\left(x^1\right)^{\mathrm{T}}\left(x^2-x^1\right)\le f\left(x^2\right)-f\left(x^1\right)$。
- （⇐）。由凸函数的定义，需要证明 $\forall\alpha\in(0,1)$，
  $\forall x^1,x^2\in S$， $f\left(\alpha x^1+(1-\alpha)x^2\right)\le\alpha f\left(x^1\right)+(1-\alpha)f\left(x^2\right)$。
- 任取 $\alpha\in(0,1)$， $x^1,x^2\in S$。令 $x=\alpha x^1+(1-\alpha)x^2$。因为 $\boldsymbol{S}$ 是凸集，故 $\boldsymbol{x}\in\boldsymbol{S}$。
- 对 $\boldsymbol{x}^1,\boldsymbol{x}$ 以及 $\boldsymbol{x}^2,\boldsymbol{x}$ 应用已知条件，可得
  $\nabla f(x)^{\mathrm{T}}\left(x^1-x\right)\le f\left(x^1\right)-f(x)$，及 **(1)**
  $\nabla f(x)^{\mathrm{T}}\left(x^2-x\right)\le f\left(x^2\right)-f(x)$。 **(2)**
- **α × (1) + (1–α) × (2)**，得：
  $\nabla f(x)^{\mathrm{T}}\left(\alpha x^1+(1-\alpha)x^2\right)-\nabla f(x)^{\mathrm{T}}x\le\alpha f\left(x^1\right)+(1-\alpha)f\left(x^2\right)-f(x)$。

凸函数的判别

- $\Rightarrow 0 \le \alpha f\left(x^1\right)+\left(1-\alpha\right)f\left(x^2\right)-f\left(x\right)$，

  $\Rightarrow f\left(x\right) \le \alpha f\left(x^1\right)+\left(1-\alpha\right)f\left(x^2\right)$，

  $\Rightarrow f\left(\alpha x^1+\left(1-\alpha\right)x^2\right) \le \alpha f\left(x^1\right)+\left(1-\alpha\right)f\left(x^2\right)$。

附：多元函数的Taylor展开式

设 n 元函数 $f(x^1, x^2, \cdots, x^n)$在点 $\left(x_1^0, x_2^0, \cdots, x_n^0\right)$附近（δ邻域）有 2 阶连续偏导数，则在这点附近有：

$$
\begin{aligned}
& f\left(x_1^0+\Delta x_1, x_2^0+\Delta x_2, \cdots, x_n^0+\Delta x_n\right) \\
= \quad & f\left(x_1^0, x_2^0, \cdots, x_n^0\right)+\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\left(x_1^0, x_2^0, \cdots, x_n^0\right) \cdot \Delta x_i \\
& +o\left(\left\|\left(\Delta x_1, \Delta x_2, \cdots, \Delta x_n\right)\right\|\right)
\end{aligned}
$$

。

附：计算极限

$$R_1 = o\left(\left\|\alpha\left(x^2 - x^1\right)\right\|\right) = o\left(\sqrt{\alpha^2 \Delta x_1 + \alpha^2 \Delta x_2 + \cdots + \alpha^2 \Delta x_n}\right)$$表明

$$\lim_{\substack{\alpha\Delta x_1 \to 0 \\ \cdots \\ \alpha\Delta x_n \to 0}} \frac{R_1}{\sqrt{\alpha^2 \Delta x_1 + \alpha^2 \Delta x_2 + \cdots + \alpha^2 \Delta x_n}} = 0$$。

因此，

$$\lim_{\alpha \to 0} \frac{R_1}{\alpha} = \lim_{\alpha \to 0} \frac{R_1 \sqrt{\sum_{i=1}^{n} \alpha^2 \Delta x_i}}{\alpha \sqrt{\sum_{i=1}^{n} \alpha^2 \Delta x_i}} = \sqrt{\sum_{i=1}^{n} \Delta x_i} \lim_{\alpha \to 0} \frac{R_1}{\sqrt{\sum_{i=1}^{n} \alpha^2 \Delta x_i}} = 0$$。

二阶连续可导凸函数的判别

定理 4.2.4 设 $S \subset \mathbf{R}^n$ 是非空开凸集，$f: S \mapsto \mathbf{R}$ 二阶连续可导，则

（1）f 是 S 上的凸函数 $\Leftrightarrow f$ 的 Hesse 矩阵 $\nabla^2 f(x)$ 在 S 上是半正定的。

（2）若 $\nabla^2 f(x)$ 在 S 上是正定的，则 f 是 S 上的严格凸函数。（注意：该逆命题不成立。）

Hesse（黑塞）矩阵

函数 $f: \mathbf{R}^n \mapsto \mathbf{R}$ 在点 $x \in \mathbf{R}^n$ 处的 Hesse 矩阵定义为：

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

若$\forall x \in S$，$\nabla^2 f(x)$都是半正定的，则称$\nabla^2 f(x)$在 S 上是半正定的。（$\nabla^2 f(x)$在 S 上是正定的定义类似。）

凸规划

回忆数学规划 **MP** 及其约束集 X 的定义为：

$$\begin{cases} \min \quad f(x) \\ s.t. \quad g_i(x) \le 0, i = 1, \dots, p \\ \qquad\quad h_j(x) = 0, j = 1, \dots q \end{cases} \qquad \textbf{(MP)}$$

$$X = \left\{ x \in \mathrm{R}^n \,\middle|\, \begin{array}{l} g_i(x) \le 0, i = 1, \dots, p \\ h_j(x) = 0, j = 1, \dots, q \end{array} \right\}。$$

如果 **MP** 的约束集 X 是凸集，目标函数 f 是 X 上的凸函数，则 **MP** 叫做凸规划。

一种凸规划

定理 4.2.5 对于非线性规划(MP)，若 $g_i(x)$，$i = 1, \cdots, p$ 皆为 R^n 上的凸函数，$h_j(x)$，$j = 1, \cdots, q$ 皆为线性函数，并且目标函数 f 是 X 上的凸函数，则(MP)是凸规划。

- 证：只需要证 X 是凸集。
- 令 $S_i = \{x \mid x \in \mathrm{R}^n, g_i(x) \le 0\}$，$i = 1, \cdots, p$，

  $$T_j = \{ x \mid x \in \mathrm{R}^n, h_j(x) = 0\},\ j = 1, \cdots, q,$$

  则 X = 诸 S_i 和 T_j 的交。
- 由定理 4.2.2，每个 S_i 都是凸集。
- 由于 $h_j(x)$为线性函数，易知 T_j 为凸集。
- 因为若干凸集的交还是凸集，于是 X 是凸集。○

凸规划：局部最优导致全局最优

定理 4.2.6 凸规划的任一局部最优解都是它的全局最优解。

- 证：设 x^*是凸规划 MP 的一个局部最优解。由定义，存在 x^*的邻域 $N_\delta(x^*)$，$\forall x \in X \cap N_\delta(x^*)$，都有 $f(x^*) \le f(x)$。(1)
- 假设 x^*不是 MP 的全局最优解。即，$\exists \bar{x} \in X$，$f(\bar{x}) < f(x^*)$。
- 因为 f 是凸函数，所以

$$
\begin{aligned}
f\left(\alpha\bar{x} + (1-\alpha)x^*\right) &\le \alpha f(\bar{x}) + (1-\alpha) f(x^*) \\
&< \alpha f(x^*) + (1-\alpha) f(x^*) \\
&= f(x^*)
\end{aligned}
$$

。

- 取$\alpha > 0$ 充分小，可使 $\alpha\bar{x} + (1-\alpha)x^*$ 落在 $X \cap N_\delta(x^*)$内。此时亦有 $f\left(\alpha\bar{x} + (1-\alpha)x^*\right) < f(x^*)$，与(1)矛盾。

谢谢大家