最优化方法

授课教师：徐海涛

2024年秋

第三章 无约束优化问题

3.1 什么是无约束优化问题?

1. 许多工程实际中都存在一些无约束情况下的多变量函数极小化问题
2. 比如，通过使得系统能量极小化来确定系统的平衡状态，利用最小二乘法构造一个方程对一组数据进行拟合，确定某个概率密度分布函数的参数来拟合一组数据。
3. 所有的无约束优化问题都可以写为如下形式：

$$minimize\ f(x_1, x_2, \cdots, x_n)$$

$$\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^T$$

本章将讨论最优性的必要条件和充分条件、凸函数、基于梯度的极小点求解方法

3.1 什么是无约束优化问题?

例 3.1 考虑如图 E3.1 所示的非线性弹簧系统。在施加了负载之后，弹簧的位移 Q_1 和 Q_2 可通过使得系统势能 Π

$$\Pi = \frac{1}{2}k_1(\Delta L_1)^2 + \frac{1}{2}k_2(\Delta L_2)^2 - F_1 Q_1 - F_2 Q_2$$

极小化确定，其中，两条弹簧的拉伸长度 ΔL_1 和 ΔL_2 分别与位移 Q_1 和 Q_2 有关：

$$\Delta L_1 = \sqrt{(Q_1+10)^2 + (Q_2-10)^2} - 10\sqrt{2}$$

$$\Delta L_2 = \sqrt{(Q_1-10)^2 + (Q_2-10)^2} - 10\sqrt{2}$$

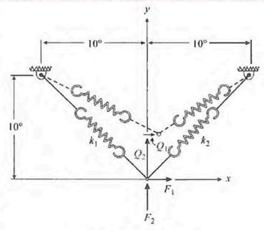


图 E3.1 非线性弹簧系统

因此，问题可构造为

$$\text{minimize} \quad \Pi(Q_1, Q_2)$$

当 $k_1 = k_2 = 1$ lb/in，$F_1 = 0$，$F_2 = 2$ 时，利用数值求解方法（接下来将讨论这些方法）可得最优解为 $Q_1 = 0$ in 和 $Q_2 = 2.55$ in。如果增加 F_2，最优解将出现非常有意思的现象（见习题）。

3.2 最优性

极小点的定义与分类

对于某个点而言，如果其邻域内的所有点对应的函数值都比该点对应的函数值要大，则该点为局部极小点。下面给出更为准确的描述：对于点 $\mathbf{x}^* = [x_1^*, x_2^*, \cdots, x_n^*]^{\mathrm{T}}$，如果存在一个 $\delta > 0$，使得对于所有满足 $\|\mathbf{x} - \mathbf{x}^*\| < \delta$ 的点 $\mathbf{x}$，都有 $f(\mathbf{x}^*) \leqslant f(\mathbf{x})$，则称点 $\mathbf{x}^*$ 是一个弱局部极小点；如果对于所有满足 $\|\mathbf{x} - \mathbf{x}^*\| < \delta$ 的点 $\mathbf{x}$，都有 $f(\mathbf{x}^*) < f(\mathbf{x})$，则称点 $\mathbf{x}^*$ 是一个强局部极小点。进一步，如果对于所有的 $\mathbf{x} \in \mathcal{R}^n$，都有 $f(\mathbf{x}^*) \leqslant f(\mathbf{x})$，则称 $\mathbf{x}^*$ 是一个全局极小点。

单选题 1分

假设我们有一个函数 $f(x) = (x-3)2$ ，并且我们考虑点 x^*=3。根据给定的定义，点 x^*=3是：

- A 一个弱局部极小点
- B 一个强局部极小点
- C 一个全局极小点
- D 既不是局部极小点也不是全局极小点

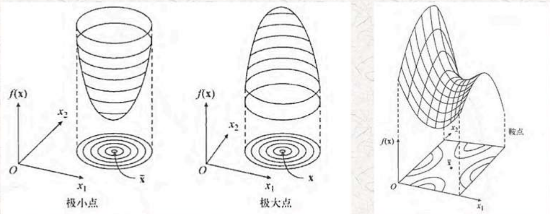
明理，精工，笃学，致远 6

3.2 最优性的必要条件

如果函数 $f \in C^1$，则 $\mathbf{x}^*$ 是局部极小点的必要条件为

$$\nabla f(\mathbf{x}^*) = \mathbf{0} \tag{3.1}$$

式（3.1）的含义为函数梯度等于0，即 $\dfrac{\partial f(\mathbf{x}^*)}{\partial x_i} = 0, i = 1, 2, \cdots, n$。满足式（3.1）的点 $\mathbf{x}^*$ 称为平稳点，可以是极大或极小点，也可以是鞍点，如图3.1所示。



明理，精工，笃学，致远 7

3.2 最优性的必要条件

式（3.1）称为一阶必要条件，下面给出推导过程。令 $\mathbf{u}_i = [0, \cdots, 1, \cdots, 0]^{\mathrm{T}}$ 表示单位向量，第 i 个元素为1，其余元素为0。因此，$\mathbf{x}^* + h\mathbf{u}_i (h > 0)$ 就可以认为是对 $\mathbf{x}^*$ 的第 i 个元素进行大小为 h 的扰动。由于 $\mathbf{x}^*$ 是局部极小点，故当 h 足够小时，必有

$$f(\mathbf{x}^* + h\mathbf{u}_i) - f(\mathbf{x}^*) \geqslant 0 \tag{3.2a}$$

$$f(\mathbf{x}^* - h\mathbf{u}_i) - f(\mathbf{x}^*) \geqslant 0 \tag{3.2b}$$

对 $f(\mathbf{x}^* + h\mathbf{u}^i)$ 和 $f(\mathbf{x}^* - h\mathbf{u}^i)$ 分别在点 $\mathbf{x}^*$ 进行一阶泰勒展开：

$$f(\mathbf{x}^* + h\mathbf{u}^i) = f(\mathbf{x}^*) + h\,\partial f(\mathbf{x}^*)/\partial x_i + O(h^2) \tag{3.3a}$$

$$f(\mathbf{x}^* - h\mathbf{u}^i) = f(\mathbf{x}^*) - h\,\partial f(\mathbf{x}^*)/\partial x_i + O(h^2) \tag{3.3b}$$

对于足够小的 h，$O(h^2)$ 是 $h\partial f(\mathbf{x}^*)/\partial x_i$ 的高阶无穷小量。因此，在这种情况下，可推出 $\partial f(\mathbf{x}^*)/\partial x_i \geqslant 0$ 和 $-\partial f(\mathbf{x}^*)/\partial x_i \geqslant 0$，这意味着必有 $\partial f(\mathbf{x}^*)/\partial x_i = 0$，且对于 $i = 1, 2, \cdots, n$ 都成立，即

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

明理，精工，笃学，致远 8

3.2 最优性的充分条件

点 $\mathbf{x}^*$ 是函数 $f(\mathbf{x})$ 严格局部极小点的充分条件为

$$\nabla f(\mathbf{x}^*)=\mathbf{0} \tag{3.4}$$

且

$$\nabla^2 f(\mathbf{x}^*)\text{正定} \tag{3.5}$$

下面给出该条件的推导过程。令 $\mathbf{y}$ 为 $\mathcal{R}^n$ 中的任意向量，为了保证 $\mathbf{x}^*$ 是严格局部极小点，故要求对于充分小的 h，有

$$f(\mathbf{x}^*+h\mathbf{y})-f(\mathbf{x}^*)>0 \tag{3.6}$$

在 $\mathbf{x}^*$ 处对 $f(\mathbf{x}^*+h\mathbf{y})$ 进行泰勒展开，可得

$$f(\mathbf{x}^*+h\mathbf{y})=f(\mathbf{x}^*)+h\nabla f(\mathbf{x}^*)^{\mathrm{T}}\mathbf{y}+\tfrac{1}{2}h^2\mathbf{y}^{\mathrm{T}}\nabla^2 f(\mathbf{x}^*)\mathbf{y}+O(h^3) \tag{3.7}$$

注意到，$\nabla f(\mathbf{x}^*)=\mathbf{0}$ 当 h 足够小的时候，余项 $O(h^3)$ 是 $\mathbf{y}^{\mathrm{T}}\nabla^2 f(\mathbf{x}^*)\mathbf{y}$ 的高阶无穷小量，因此有

$$\mathbf{y}^{\mathrm{T}}\nabla^2 f(\mathbf{x}^*)\mathbf{y}>0$$

上式对于任意 $\mathbf{y}$ 都成立，因此，$\nabla^2 f(\mathbf{x}^*)$ 一定是正定矩阵。式（3.5）的几何意义与凸性的有关概念密切相关，接下来将讨论这一点。

3.2 最优性的充分条件（例题）

例 3.2 利用形如 $y=a+\dfrac{b}{x}$ 的方程对一组点（1,6），（3,10），（6,2）进行最小二乘意义上的“最好的”拟合。试利用极小点的必要条件和充分条件求 a 和 b 的值。

构造最小二乘问题：

$$\text{minimize } f=\sum_{1}^{3}\left(a+\frac{b}{x_i}-y_i\right)^2$$

有两个待求变量 a 和 b，将各点 (x_i,y_i) 代入上式之后，可得

$$\text{minimize } f=3a^2+41/36b^2+3ab-58/3b-36a+140$$

极小点的必要条件要求

$$\partial f/\partial a=6a+3b-36=0$$
$$\partial f/\partial b=41/18b+3a-58/3=0$$

解之可得 a 和 b 及其对应的函数值：

$$a^*=5.1428,\quad b^*=1.7143,\quad f^*=30.86$$

3.2 最优性的充分条件（例题）

例 3.2 利用形如 $y=a+\dfrac{b}{x}$ 的方程对一组点（1,6），（3,10），（6,2）进行最小二乘意义上的“最好的”拟合。试利用极小点的必要条件和充分条件求 a 和 b 的值。

故“最好的”拟合函数为 $y(x)=5.1428+1.7143/x$。接下来利用充分条件判断 (a^*,b^*) 是否是极小点，函数的黑塞矩阵为

$$\nabla^2 f=\begin{bmatrix}\dfrac{\partial^2 f}{\partial a^2} & \dfrac{\partial^2 f}{\partial a\partial b}\\ \dfrac{\partial^2 f}{\partial b\partial a} & \dfrac{\partial^2 f}{\partial b^2}\end{bmatrix}_{(a^*,b^*)}=\begin{bmatrix}6 & 3\\ 3 & \frac{41}{18}\end{bmatrix}$$

由于有 $6>0$ 和 $(6)(41/18)-3^2>0$，根据第 1 章给出的西尔维斯特判别准则，可知 $\nabla^2 f$ 是正定的，解 (a^*,b^*) 为强局部极小点。在 3.3 节中还将证明 f 是严格凸的，故 (a^*,b^*) 还是强全局极小点。

3.2 最优性的充分条件（例题）

例 3.3 某函数为 $f=2x_1^2+8x_2^2+8x_1x_2-5x_1$，利用必要条件求该函数的局部极小点，有

$$\partial f/\partial x_1=4x_1+8x_2-5=0 \quad \text{或} \quad 4x_1+8x_2=5$$
$$\partial f/\partial x_2=8x_1+16x_2=0 \quad \text{或} \quad 4x_1+8x_2=0$$

显然这个方程组无解，这意味着在 $\mathcal{R}^2$ 内该函数不存在局部极小点。该函数的曲面如图 E3.3 所示。

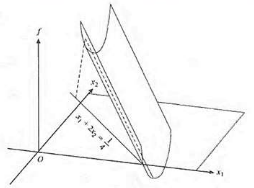


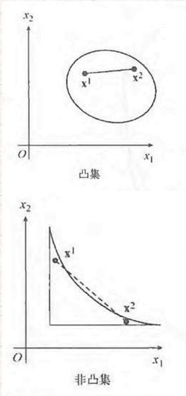
图 E3.3 函数 $f=2x_1^2+8x_2^2+8x_1x_2-5x_1$ 的曲面

3.3 凸性

黑塞矩阵$\nabla^2 f$与函数凸性也有着密切关联

凸集

对R^n中的集合S，如果对于任意$\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$和任意实数$\alpha, 0 < \alpha < 1$，都有$\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2 \in S$成立，则$S$是一个凸集。也就是说，凸集$S$中的任意两点所构成线段上的所有点仍然位于$S$中，



凸集

非凸集

3.3 凸性

凸函数

引申应用到n变量函数中，对于定义域为凸集S的函数$f(x)$，如果对于任意$\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$和所有$\alpha, 0 < \alpha < 1$，都有

$$f(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) \leq \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2)$$

则称$f(x)$为S上的凸函数

3.3 凸函数的性质

（1）对于定义在凸集S上的函数$f \in C^1$（即函数f的导函数连续），当且仅当对于所有的$\boldsymbol{x}, \boldsymbol{y} \in S$，有

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x})$$

则f为S上的凸函数

一阶条件：给定的条件是凸函数的一阶条件，也称为凸性的条件。它表明，对于凸函数f，任意点 y 处的函数值至少与通过点 x 处的切平面预测的值一样大。

（2）对于定义在凸集S上的函数$f \in C^2$（即函数f的二阶导函数连续），S至少存在一个内点，当且仅当在整个S上$\nabla^2 f$都半正定，则f为S上的凸函数。

充分条件实际上就是说如果f是某个凸集上的严格凸函数，则$\boldsymbol{x}^*$是严格局部极小点

3.3 凸函数的性质

（3）如果f是凸函数，则f的任意局部极小点都是全局极小点

（4）如果f是凸函数，$\boldsymbol{x}^*$满足必要条件$\nabla f(\boldsymbol{x}^*) = 0$，那么$\boldsymbol{x}^*$是$f$的全局极小点。将$\boldsymbol{x}^* = \boldsymbol{x}$代入式(3. 9) 中，就可以得到这一结论。

这意味着对于凸函数，全局最优性的必要条件同时也是充分条件

3.3 凸函数的性质

例 3.4　函数

$$f = x_1 x_2$$

定义在凸集 $S = \{(x_1, x_2) \in \mathcal{R}^2 : x_1 > 0, x_2 > 0\}$ 上。试问 f 是不是 S 上的凸函数？

由于函数 f 二阶连续可导，故可利用黑塞矩阵作为凸函数的判据：

$$\nabla^2 f = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

黑塞矩阵的两个特征值分别为 -1 和 1，并不是半正定的，因此，f 不是凸函数。从图 E3.4 所示的函数图像中，可以明显看出这一点。

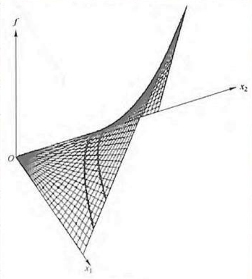


图 E3.4

明理，精工，笃学，致远　17

本节结束，谢谢

Thank you for listening

课程名称：最优化方法

明理，精工，笃学，致远

习题

作业题：P3.1 P3.3 P3.4 P3.10

思考题：P3.5 P3.6

明理，精工，笃学，致远　19