3.8 最小二乘法

在很多实际问题中, 例如曲线拟合问题等, 经常会遇到求解若干个函数平方和的极小化问题, 即目标函数为

$$R(x)=\sum_{i=1}^{m} f_i^2(x), \quad \forall x \in \mathbb{R}^n$$

的极小化问题，称此类问题为最小二乘问题. 如果目标函数中所涉及到的函数都是线性的，那么对应的问题被称为线性最小二乘问题; 否则，即目标函数中涉及到的函数至少有一个是非线性的函数，则对应的问题称为非线性最小二乘问题. 本节介绍最小二乘问题的基本求解方法.

3.8 最小二乘法 — 线性最小二乘问题

§3.8.1 线性最小二乘问题

线性最小二乘问题的模型为

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|^2 \tag{3.8.1}$$

其中$A\in\mathbb{R}^{m\times n}$且$b\in\mathbb{R}^m$. 一般地，假设$n\le m$且矩阵$A$列满秩.

显然，问题(3.8.1)是一类特殊的无约束优化问题. 令其目标函数的梯度等于零，由无约束优化问题的最优性条件，于是有

$$A^\top Ax=A^\top b.$$

由于矩阵A列满秩，所以矩阵$A^\top A$是正定的. 因而可得线性最小二乘问题(3.8.1)的稳定点为

$$x^*=(A^\top A)^{-1}A^\top b.$$

3.8 最小二乘法 — 线性最小二乘问题

此外，易证：问题(3.8.1)的目标函数是一个凸函数，所以以上求得的稳定点是线性最小二乘问题(3.8.1)的全局极小值点. 这个极小值点也称为最小二乘解. 以上讨论可归结为如下定理.

定理 3.8.1 对于线性最小二乘问题*(3.8.1)*, 如果矩阵$A \in \mathbb{R}^{m\times n}$是列满秩的矩阵, 那么$x^* = (A^\top A)^{-1}A^\top b$ 为线性最小二乘问题*(3.8.1)*的全局极小值点.

对于线性最小二乘问题(3.8.1)，只要矩阵$A^\top A$非奇异，就可以用定理3.8.1来求解.

对于矩阵$A^\top A$是奇异的情况, 可采用一些改进的最小二乘方法来处理，如Golub的改进方法等，本书不作介绍.

3.8 最小二乘法 — 线性最小二乘问题

例 3.8.1 求函数

$$f(x) = \|Ax - b\|^2, \quad 其中 A = \begin{pmatrix} 1 & 1 \\ 2 & -3 \\ -1 & 4 \end{pmatrix}, \ b = \begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix}$$

的最小二乘解.

解 简单计算可得

$$A^\top A = \begin{pmatrix} 6 & -9 \\ -9 & 26 \end{pmatrix}, \quad (A^\top A)^{-1} = \frac{1}{75}\begin{pmatrix} 26 & 9 \\ 9 & 6 \end{pmatrix}.$$

因此,

$$x^* = \frac{1}{75}\begin{pmatrix} 26 & 9 \\ 9 & 6 \end{pmatrix}\begin{pmatrix} 1 & 2 & -1 \\ 1 & -3 & 4 \end{pmatrix}\begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 13 \\ 7 \end{pmatrix}$$

即为所求的最小二乘解. □

3.8 最小二乘法 — 非线性最小二乘问题

§3.8.2　非线性最小二乘问题

非线性最小二乘问题的模型为

$$\min_{x\in\mathbb{R}^n} R(x) = \sum_{i=1}^{m} f_i^2(x) \tag{3.8.2}$$

其中函数f_i $(i \in \{1, 2, \ldots, m\})$中至少有一个是非线性的且函数$R$具有一阶连续可微性.

求解非线性最小二乘问题(3.8.2)的基本方法如下：通过解一系列的线性最小二乘问题来逐步逼近非线性最小二乘问题的解. 具体地，假设x^k是最小二乘解的第k次的近似解, 将每个函数f_i在x^k处进行线性展开，以至于把原来的问题转化为线性最小二乘问题，将其最优解作为最小二乘解的第$k+1$次的近似解x^{k+1}，不断重复以上过程，直到得到原问题的解为止.

3.8 最小二乘法 — 非线性最小二乘问题

若记

$$\begin{aligned} f(x) &:= (f_1(x), \cdots, f_m(x))^\top, \\ A_k &:= (\nabla f_1(x^k) \ \cdots \ \nabla f_m(x^k)), \\ b &:= A_k x^k - f(x^k), \end{aligned}$$

则上述方法中的x^{k+1}为下面无约束优化问题的解:

$$\min_{x \in \mathbb{R}^n} \ \|A_k x - b\|^2.$$

假设每个A_k是列满秩的矩阵, 那么可得到迭代公式为

$$x^{k+1} = x^k + d^k, \quad \text{其中} d^k = -(A_k^\top A_k)^{-1} A_k^\top f(x^k). \tag{3.8.3}$$

(3.8.3)中的迭代格式称为Gauss-Newton公式，其中的d^k称为Gauss-Newton方向.

3.8 最小二乘法 — 非线性最小二乘问题

类似于带线搜索的牛顿法，可从x^k出发，沿方向d^k进行一维线搜索得到迭代步长λ_k，进而，可设计带线搜索的Gauss-Newton法.

算法 3.8.1 *(带线搜索的Gauss-Newton法)*
选择初始点$x^0 \in \mathbb{R}^n$. 置$k := 0$.

步1 若$\|A_k^\top f(x^k)\| = 0$, 则算法终止, 得到解为x^k. 否则, 转步2.

步2 由$A_k^\top A_k d = -A_k^\top f(x^k)$求得迭代方向$d^k$.

步3 由线搜索计算步长λ_k.

步4 置$x^{k+1} := x^k + \lambda_k d^k$和$k := k + 1$, 转步1.

3.8 最小二乘法 — 非线性最小二乘问题

在实际中，经常会出现矩阵$A_k^\top A_k$是奇异的情况. 在这种情况下，可采用下面的Levenberg-Marquart方法来求解.

算法 3.8.2 *(Levenberg-Marquart法)* 取初始点$x^0 \in \mathbb{R}^n$, 初始参数$\alpha > 0$, 增长因子$\beta > 1$. 置$k := 0$.

步*1* 置$\alpha := \alpha/\beta$.

步*2* 由$(A_k^\top A_k + \alpha I)d = -A_k^\top f(x^k)$求得迭代方向$d^k$. 置$x^{k+1} := x^k + d^k$.

步*3* 若$R(x^{k+1}) < R(x^k)$, 则转步5; 否则，转步4.

步*4* 若$\|A_k^\top f(x^k)\| = 0$, 算法终止，得到解为x^k; 否则，置$\alpha := \beta\alpha$, 转步2.

步*5* 若$\|A_k^\top f(x^k)\| = 0$, 算法终止，得到解为x^{k+1}; 否则，置$k := k+1$, 转步1.

3.8 最小二乘法 — 非线性最小二乘问题

在一定条件下，Gauss-Newton法与Levenberg-Marquart法均具有好的收敛性质. 下面的定理给出了Levenberg-Marquart法的收敛性, 其证明本书省略.

定理 3.8.2 设$\{x^k\}$是由算法3.8.2产生的迭代序列. 假设水平集$\mathscr{L}(R)=\{x\in\mathbb{R}^n \mid R(x)\le R(x^0)\}$是紧的, 且矩阵$A_k^\top A_k$在$\mathscr{L}(R)$上恒为正定矩阵. 那么算法3.7.1或者有限终止; 或者产生一个无穷序列. 在后一种情况, 迭代序列的极限点必为函数$R(\cdot)$的稳定点.

3.8 最小二乘法 — 作业

3.19 求解下面的线性最小二乘问题:

$$\min \quad f(x) = \|Ax - b\|^2,$$

其中

$$A = \begin{pmatrix} 1 & 1 \\ 2 & -3 \\ -1 & 4 \end{pmatrix}, \quad b = \begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix}.$$