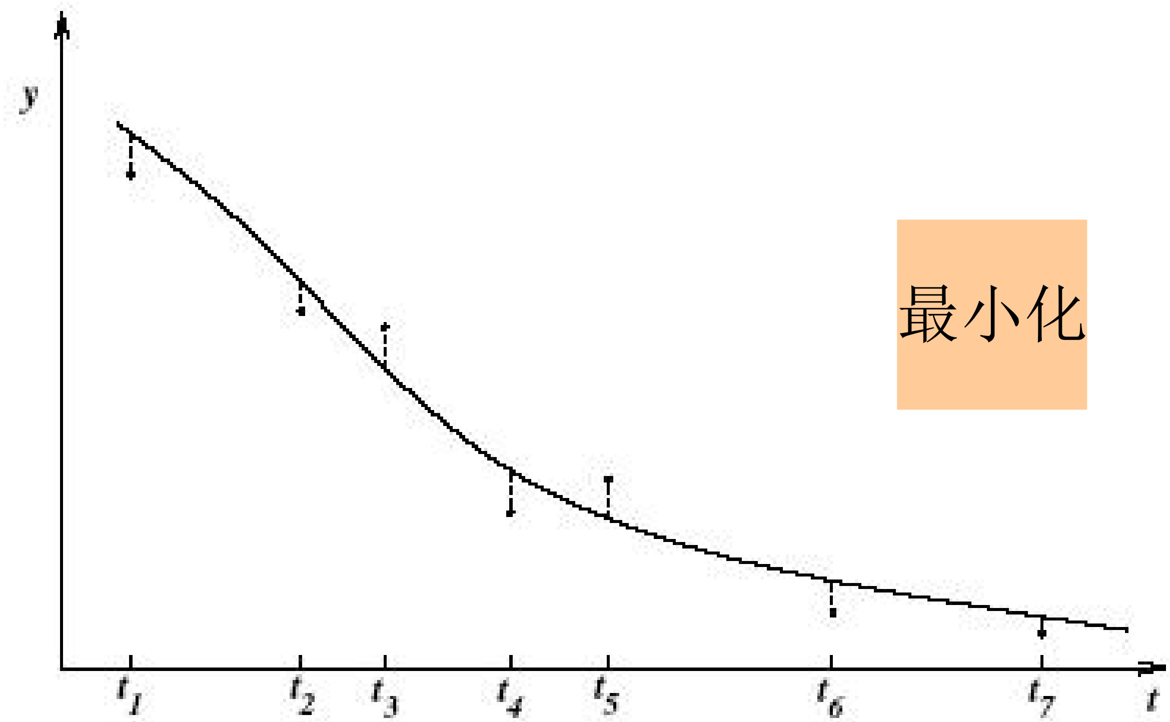
第十五讲 无约束优化问题的应用

- 最小二乘法
- 非线性方程组

例:某种药对病人的影响研究

- 病人服药后,在时间t,抽取血液样本测量出药物浓度为y,预测浓度参数为 $x=(x_1,x_2,x_3,x_4,x_5)$:

$$\phi(x;t)=x_1+tx_2+t^2x_3+x_4e^{-x_5t}.$$



最小化

$$\frac{1}{2}\sum_{j=1}^{m}[\phi(x;t_j)-y_j]^2.$$

最小二乘问题

$$\min f(x) = \frac{1}{2}\sum_{i=1}^{m} F_i(x)^2,$$

其中， $F_i : R^n \to R,\ i = 1, 2, \ldots, m$ 连续可微.

线性最小二乘问题

$$\min f(x) = \frac{1}{2}\sum_{i=1}^{m} (a_i^T x - b_i)^2$$

非线性最小二乘问题

$$\min f(x) = \frac{1}{2}\sum_{i=1}^{m} F_i(x)^2$$

应用：数据拟合、参数估计、函数逼近

线性最小二乘问题

$$\min f(x) = \frac{1}{2}\sum_{i=1}^{m}(a_i^T x - b_i)^2 \tag{8.15}$$

最优性条件 $\nabla f(x) = \sum_{i=1}^{m}(a_i^T x - b_i)a_i = 0.$

$A = (a_1, \cdots, a_m)^T$

$b = (b_1, \cdots, b_m)^T$

$$A^T A x - A^T b = 0. \tag{8.16}$$

线性方程组 (8.16) 称为问题 (8.15) 的正规方程组. 由于正规方程组的系数矩阵的秩与其增广矩阵的秩相等, 因此, 方程组有解. 特别, 当 A 是满秩矩阵时, 系数矩阵对称正定, 方程组有惟一解. 此时, 可利用附录中求解线性方程组的算法进行求解.

非线性最小二乘问题

$$\min f(x) = \frac{1}{2}\sum_{i=1}^{m} F_i(x)^2,$$

至少有一个 F_i 是非线性函数

设 F 二次连续可微.

F 在 x 处的 Jacobian 矩阵

$$F'(x) = (\nabla F_1(x), \nabla F_2(x), \cdots, \nabla F_m(x))^T$$

$$\Rightarrow \quad \nabla f(x) = \sum_{i=1}^{m} F_i(x)\nabla F_i(x) = F'(x)^T F(x),$$

$$\nabla^2 f(x) = F'(x)^T F'(x) + \sum_{i=1}^{m} F_i(x)\nabla^2 F_i(x),$$

注意到该问题的特殊结构，我们可以针对该问题设计特别的算法.

Gauss-Newton 法和 Levenberg-Marquardt 算法

Gauss-Newton法

$d^{(k)}$ 是线性最小二乘问题 $\min \frac{1}{2}\|F(x^{(k)})+F'(x^{(k)})d\|^2$ 的解.

$$\Longleftrightarrow \quad \boxed{F'(x^{(k)})^T F'(x^{(k)})d + F'(x^{(k)})^T F(x^{(k)}) = 0.} \tag{8.17}$$

注意到 $\nabla^2 f(x)$ 的表达式, 线性方程组 (8.17) 的系数矩阵实际上是在 $\nabla^2 f(x^{(k)})$ 中去掉了高阶导数项 $\sum_{i=1}^{m} F_i(x)\nabla^2 F_i(x)$ 后的余下部分. 若 $\|F(x)\|$ 很小, 则 (8.17) 确定的 $d^{(k)}$ 是一种近似 Newton 方向或非精确 Newton 方向. 下面的定理表明, 若 $\nabla f(x^{(k)}) \neq 0$, 则 $d^{(k)}$ 是 f 在 $x^{(k)}$ 处的一个下降方向.

定理 8.4.1 设函数 $F=(F_1,F_2,\ldots,F_m)^T : R^n \to R^m$ 连续可微. 函数 $f: R^n \to R$ 由 (8.14) 定义. 若 $\nabla f(x^{(k)}) \neq 0$, 则由 (8.17) 确定的 $d^{(k)}$ 满足:

$$\nabla f(x^{(k)})^T d^{(k)} < 0.$$

$$F'(x^{(k)})^T F'(x^{(k)})d + F'(x^{(k)})^T F(x^{(k)}) = 0.$$

算法 8.4 *(***非线性最小二乘问题的** *Gauss-Newton* **法***)*

步 *1* 给定初始点 $x^{(0)} \in R^n$, 常数 $\rho \in (0,1)$, $\sigma_1 \in (0,1/2)$, 精度 $\epsilon > 0$. 令 $k := 0$.

步 *2* 若 $\|\nabla f(x^{(k)})\| \leq \epsilon$, 则终止算法, 得解 $x^{(k)}$. 否则, 转步 *3*.

步 *3* 解线性方程组 (8.17) 得方向 $d^{(k)}$.

步 *4* 确定步长 α_k 为集合 $\{\rho^i \mid i = 0,1,2,\ldots\}$ 中使得下面的不等式成立的最大者:

$$f(x^{(k)} + \alpha_k d^{(k)}) \leq f(x^{(k)}) + \sigma_1 \alpha_k \nabla f(x^{(k)})^T d^{(k)}.$$

步 *5* 令 $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$, $k := k+1$. 转步 *2*.

定理 8.4.2 设函数 $F=(F_1,F_2,\ldots,F_m)^T: R^n\to R^m$ 连续可微，函数 $f: R^n\to R$ 由 (8.14) 定义且存在常数 $\gamma>0$ 使得

$$\|F'(x)d\|\geq\gamma\|d\|,\qquad \forall x,d\in R^n.$$

设 $\{x^{(k)}\}$ 是由 *Gauss-Newton* 法产生的点列且水平集

$$\Omega=\{x\in R^n\mid f(x)\leq f(x^{(0)})\}$$

有界. 则

$$\lim_{k\to\infty}\|\nabla f(x^{(k)})\|=0. \tag{8.19}$$

定理 8.4.3 设定理 *8.4.2* 的条件成立. 再设由 *Gauss-Newton* 法产生的点列 $\{x^{(k)}\}$ 收敛于 x^* 且 $F(x^*) = 0$. 则 $\{x^{(k)}\}$ 具有超线性收敛速度. 若进一步假设 F' *Lipschitz* 连续, 则 $\{x^{(k)}\}$ 二次收敛于 x^*.

Levenberg-Marquardt算法

$d^{(k)}$

$$(F'(x^{(k)})^T F'(x^{(k)}) + \mu_k I)d + F'(x^{(k)})^T F(x^{(k)}) = 0, \quad (8.20)$$

系数矩阵对称正定，因而，其解存在惟一.

算法 8.5 (**非线性最小二乘问题的** *Levenberg-Marquardt* **算法**)

步 *1* 给定初始点 $x^{(0)} \in R^n$, 常数 $\rho \in (0,1)$, $\sigma_1 \in (0,1/2)$, 精度 $\epsilon > 0$. 令 $k := 0$.

步 *2* 若 $\|\nabla f(x^{(k)})\| \le \epsilon$, 则终止算法，得解 $x^{(k)}$. 否则，转步 *3*.

步 *3* 解线性方程组 (8.20) 得方向 $d^{(k)}$.

步 *4* 确定步长 α_k 为集合 $\{\rho^i \mid i = 0,1,2,\ldots\}$ 中使得下面的不等式成立的最大者：

$$f(x^{(k)} + \alpha_k d^{(k)}) \le f(x^{(k)}) + \sigma_1 \alpha_k \nabla f(x^{(k)})^T d^{(k)}.$$

步 *5* 令 $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$, $k := k+1$. 转步 *2*.

Levenberg-Marquardt 算法与 Gauss-Newton 算法的惟一区别在于确定下降方向的子问题.

LM法的收敛性

定理 8.4.5 设 F 连续可微，且水平集

$$\Omega = \{x \in R^n \mid f(x) \le f(x^{(0)})\}$$

有下界. $\{x^{(k)}\}$ 由算法 *8.5* 产生. 若 (x^*, μ^*) 是序列 $\{(x^{(k)}, \mu_k)\}$ 的一个极限点使得矩阵 $F'(x^*)^T F'(x^*) + \mu^* I$ 正定，则 x^* 满足 $\nabla f(x^*) = 0$.

类似于定理，关于 Levenberg-Marquardt 算法的收敛速度，我们有如下定理.

定理 8.4.6 设定理 *8.4.5* 的条件成立. 再设由 *Levenberg-Marquardt* 算法产生的点列 $\{x^{(k)}\}$ 收敛于 x^* 且 $F(x^*) = 0$, $F'(x^*)$ 满秩. 若 $\{\mu_k\} \to 0$, 则 $\{x^{(k)}\}$ 具有超线性收敛速度. 若进一步假设 F' *Lipschitz* 连续，且存在常数 $C > 0$ 使得 $\mu_k \le C\|F(x^{(k)})\|$, 则 $\{x^{(k)}\}$ 二次收敛于 x^*.

注记

- 对于非线性最小二乘问题的极小点x*,如果||F(x*)||不是很小，与0的距离较大，那么忽略高阶导数项的Guass-Newton法和LM法的计算效果不好，因此，应用拟牛顿校正思想来产生忽略掉的高解项的近似是一个很好的选择。

例：控制问题
分析飞行器对飞行员指令反应的稳定性

平衡方程

$$F(x) \equiv Ax + \phi(x) = 0,$$

$$A = \begin{bmatrix} -3.933 & 0.107 & 0.126 & 0 & -9.99 & 0 & -45.83 & -7.64 \\ 0 & -0.987 & 0 & -22.95 & 0 & -28.37 & 0 & 0 \\ 0.002 & 0 & -0.235 & 0 & 5.67 & 0 & -0.921 & -6.51 \\ 0 & 1.0 & 0 & -1.0 & 0 & -0.168 & 0 & 0 \\ 0 & 0 & -1.0 & 0 & -0.196 & 0 & -0.0071 & 0 \end{bmatrix}$$

$$\phi(x) = \begin{bmatrix} -0.727x_2x_3 + 8.39x_3x_4 - 684.4x_4x_5 + 63.5x_4x_2 \\ 0.949x_1x_3 + 0.173x_1x_5 \\ -0.716x_1x_2 - 1.578x_1x_4 + 1.132x_4x_2 \\ -x_1x_5 \\ x_1x_4 \end{bmatrix}.$$

x_1:转动率, x_2:倾斜率，x_3:偏航率, x_4:迎角增量，x_5:侧滑角，

x_6:升降舵偏差, x_7:副翼偏差，x_8:方向舵偏差。（控制变量）

非线性方程组 F(x)=0

- 局部牛顿法 $$x^{(k+1)} = x^{(k)} + d^{(k)}$$

$$\boxed{F'(x^{(k)})d + F(x^{(k)}) = 0}$$

定理 8.1.1 设 $F : F^n \to R^n$ 连续可微，$\bar{x}$ 是方程组 (8.1) 的一个解，$F'(\bar{x})$ 非奇异. 则存在 $\bar{x}$ 的一个邻域 $U(\bar{x})$, 使得当 $x^{(0)} \in U(\bar{x})$ 时，局部 *Newton* 法产生的点列 $\{x^{(k)}\}$ 包含于该邻域中，而且 $\{x^{(k)}\}$ 超线性收敛于 $\bar{x}$. 若再假设 F' *Lipschitz* 连续，则 $\{x^{(k)}\}$ 二次收敛于 $\bar{x}$.

阻尼Newton法求解F(x)=0

- 设 $F(x)=(F_1(x),F_2(x),\cdots,F_n(x))^T$. 引入函数

$$\theta(x)=\frac{1}{2}\|F(x)\|^2=\frac{1}{2}\sum_{i=1}^{n}F_i(x)^2.$$

$$\Downarrow$$

$$\nabla\theta(x)=\sum_{i=1}^{n}F_i(x)\nabla F_i(x)=F'(x)^TF(x).$$

$$\Downarrow$$

$$F'(x)d+F(x)=0. \tag{8.2}$$

$$\Downarrow$$

$$\nabla\theta(x)^T\bar{d}=-\|F(x)\|^2<0.$$

算法 8.1 *(非线性方程组的阻尼 Newton 法)*

步 *1* 给定初始点 $x^{(0)} \in R^n$, 常数 $\rho \in (0,1)$, $\sigma_1 \in (0,1/2)$, 精度 $\epsilon > 0$. 令 $k := 0$.

步 *2* 若 $\|F(x^{(k)})\| \le \epsilon$, 则终止算法, 得解 $x^{(k)}$. 否则, 转步 *3*.

步 *3* 解线性方程组 (8.2) 得方向 $d^{(k)}$.

步 *4* 确定步长 α_k 为集合 $\{\rho^i \mid i = 0, 1, 2, \ldots\}$ 中使得下面的不等式成立的最大者:

$$\theta(x^{(k)} + \alpha_k d^{(k)}) \le (1 - 2\sigma_1\alpha_k)\theta(x^{(k)}).$$

步 *5* 令 $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$, $k := k + 1$. 转步 *2*.

阻尼Newton法的全局收敛性和超线性收敛性

定理 8.3.1 设 $F: R^n \to R^n$ 连续可微且对每一个 $x \in R^n$, $F'(x)$ 非奇异, $\{x^{(k)}\}$ 是由阻尼 *Newton* 法产生的点列. 假设水平集

$$\Omega = \{x \in R^n \mid \theta(x) \leq \theta(x^{(0)})\}$$

有界. 则

(1) 序列 $\{x^{(k)}\}$ 收敛于非线性方程组 (8.1) 的惟一解, 而且收敛速度是超线性的.

(2) 若再假设 F' *Lipschitz* 连续, 则 $\{x^{(k)}\}$ 二次收敛.

非线性方程组求解其它方法

- 局部拟牛顿法
- 全局方法:

  线性搜索型拟牛顿算法

  信赖域算法

总结

- 最小二乘问题求解方法
- 非线性方程组求解方法

可采取无约束优化方法求解,但由于有特殊结构, 有更好的算法. 收敛性和收敛速度可应用第2章的结果证明.

作业：习题8： 1、5

练习

- 掌握最小二乘和非线性方程组的求解算法和收敛理论.

提高训练:

- 理论:应用第二章理论证明收敛性理论和收敛速度
- 程序: 实现所学算法,并与直接采用的无约束优化方法进行比较.