5-2 非线性最小二乘

非线性最小二乘问题

考虑非线性最小二乘问题

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x}) = \frac{1}{2}r(\boldsymbol{x})^T r(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{m}[r_i(\boldsymbol{x})]^2, \qquad m \geq n, \tag{5.2.1}$$

其中: 剩余函数$\boldsymbol{r}(\boldsymbol{x}) = (r_1(\boldsymbol{x}), r_2(\boldsymbol{x}), \cdots, r_m(\boldsymbol{x}))^T$为$\boldsymbol{x} \in \mathbb{R}^n$的非线性函数。

记$\boldsymbol{r}(\boldsymbol{x})$的$m \times n$阶Jacobian矩阵如下：

$$J(\boldsymbol{x}) = \begin{bmatrix} \nabla r_1^T(\boldsymbol{x}) \\ \vdots \\ \nabla r_m^T(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial r_1}{\partial x_1}(\boldsymbol{x}) & \frac{\partial r_1}{\partial x_2}(\boldsymbol{x}) & \cdots \frac{\partial r_1}{\partial x_n}(\boldsymbol{x}) \\ \frac{\partial r_2}{\partial x_1}(\boldsymbol{x}) & \frac{\partial r_2}{\partial x_2}(\boldsymbol{x}) & \cdots \frac{\partial r_2}{\partial x_n}(\boldsymbol{x}) \\ \cdots & \cdots & \cdots \\ \frac{\partial r_m}{\partial x_1}(\boldsymbol{x}) & \frac{\partial r_m}{\partial x_2}(\boldsymbol{x}) & \cdots \frac{\partial r_m}{\partial x_n}(\boldsymbol{x}) \end{bmatrix} \in \mathbb{R}^{m\times n}$$

非线性最小二乘问题

$f(\boldsymbol{x})$的一阶和二阶导数(Hesse矩阵)分别为:

$$\boldsymbol{g}(\boldsymbol{x})=\nabla f(\boldsymbol{x})=\sum_{i=1}^{m} r_i(\boldsymbol{x})\nabla r_i(\boldsymbol{x})=J(\boldsymbol{x})^T\boldsymbol{r}(\boldsymbol{x}) \tag{5.2.2}$$

$$\begin{aligned} G(\boldsymbol{x})=\nabla^2 f(\boldsymbol{x})&=\sum_{i=1}^{m}\nabla r_i(\boldsymbol{x})\nabla r_i(\boldsymbol{x})^T+\sum_{i=1}^{m} r_i(\boldsymbol{x})\nabla^2 r_i(\boldsymbol{x}) \\ &=J(\boldsymbol{x})^T J(\boldsymbol{x})+S(\boldsymbol{x}) \end{aligned} \tag{5.2.3}$$

其中:

$$S(\boldsymbol{x})=\sum_{i=1}^{m} r_i(\boldsymbol{x})\nabla^2 r_i(\boldsymbol{x}).$$

这里: $r(\boldsymbol{x})$二阶连续可微。

非线性最小二乘问题

若$\boldsymbol{x}^*$为最小二乘问题(5.2.1)的最优解，记：

$$J^* = J(\boldsymbol{x}^*), J_k = J(\boldsymbol{x}_k),$$
$$S^* = S(\boldsymbol{x}^*), S_k = S(\boldsymbol{x}_k).$$

在$\boldsymbol{x}^*$处，$\|S^*\|$的大小取决于剩余量和问题的非线性性:

- 对零剩余量或线性最小二乘问题, $\|S^*\| = 0$;
- 随着剩余量的增大或$r_i(\boldsymbol{x})$的非线性性增强，$\|S^*\|$的值变大。

根据问题的这种特点, 最小二乘算法可分为：

- 小剩余算法：处理$\|S^*\|$为零或者不大的问题；
- 大剩余算法：处理$\|S^*\|$较大的问题。

最小二乘问题牛顿法

解最小二乘问题牛顿方程为：

$$(J_k^T J_k + S_k)\boldsymbol{d}_k = -J_k^T \boldsymbol{r}_k \tag{5.2.4}$$

牛顿法的缺点: 每次迭代都需要计算S_k, 涉及到m个$n \times n$对称矩阵，这是一个计算上的很大负担。解决这个问题的办法：

- 或者在牛顿方程中忽略S_k,
- 或者用一阶导数信息近似S_k。

小剩余算法主要思想: 在$r_i(\boldsymbol{x})$接近于0或者接近于线性时忽略S_k。

(5.2.4)式中忽略S_k即可得到Gauss-Newton方法。

高斯-牛顿法

从另一个角度理解，在点$\boldsymbol{x}_k$处**线性化**剩余函数$r_i(\boldsymbol{x}_k+\boldsymbol{d}) \approx r_i(\boldsymbol{x}_k)+\nabla r_i(\boldsymbol{x}_k)^T\boldsymbol{d}$. 继而得到关于$\boldsymbol{d}$的线性最小二乘问题：

$$\min_{\boldsymbol{d}\in\mathbb{R}^n} q_k(\boldsymbol{d}) = \frac{1}{2}\|J_k\boldsymbol{d}+\boldsymbol{r}_k\|_2^2 \tag{5.2.5}$$

其中：

$$q_k(\boldsymbol{d}) = \frac{1}{2}(J_k\boldsymbol{d}+\boldsymbol{r}_k)^T(J_k\boldsymbol{d}+\boldsymbol{r}_k) = \frac{1}{2}\boldsymbol{d}^T J_k^T J_k\boldsymbol{d} + \boldsymbol{d}^T(J_k^T\boldsymbol{r}_k) + \frac{1}{2}\boldsymbol{r}_k^T\boldsymbol{r}_k. \tag{5.2.6}$$

这里$q_k(\boldsymbol{d})$是$f(\boldsymbol{x}_k+\boldsymbol{d})$的一种二次近似，与$f(\boldsymbol{x}_k+\boldsymbol{d})$的二次Taylor近似的差别在于二次项少了$S_k$。

高斯-牛顿法

问题(5.2.5)的极小点$\boldsymbol{d}_k$满足：

$$J_k^T J_k \boldsymbol{d} = -J_k^T \boldsymbol{r}_k \tag{5.2.7}$$

上式称为Gauss-Newton方程。

算法5.2.1（Gauss-Newton法）

步1 给定$\boldsymbol{x}_0, \varepsilon > 0$, $k = 0$;

步2 若终止条件满足，停止迭代;

步3 解方程组$J_k^T J_k \boldsymbol{d} = -J_k^T \boldsymbol{r}_k$得$\boldsymbol{d}_k$;

步4 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$, $k := k+1$, 转步2。

当$\alpha_k = 1$时，以上算法对应基本Gauss-Newton方法。带线搜索的Gauss-Newton方法称为阻尼Gauss-Newton方法。

高斯-牛顿法

Gauss-Newton法充分利用了最小二乘问题的结构特点，仅仅利用函数的一阶导数($J(\boldsymbol{x})$)的信息直接获得Hesse 矩阵的近似，而无需同拟牛顿方法那样利用函数一阶导数信息的逐步积累来获取Hesse矩阵的近似。

Gauss-Newton方法的优点:

- 无需计算$\boldsymbol{r}(\boldsymbol{x})$的二阶导数；
- $\boldsymbol{d}_k$是下降方向。由(5.2.2)式和(5.2.7)式知：

$$\boldsymbol{d}_k^T \boldsymbol{g}_k = \boldsymbol{d}_k^T J_k^T \boldsymbol{r}_k = -\boldsymbol{d}_k^T J_k^T J_k \boldsymbol{d}_k = -\|J_k \boldsymbol{d}_k\|^2$$

这说明, 当J_k满秩, $\boldsymbol{g}_k$非零时，$\boldsymbol{d}_k$是下降方向。

定理5.2.1 Gauss-Newton方法局部收敛性.

设$r_i(\boldsymbol{x}) \in C^2(i = 1, \ldots, m)$，$\boldsymbol{x}^*$是最小二乘问题(5.1.1)的最优解，且$J^{*T} J^*$正定。假定基本Gauss-Newton方法产生的迭代点列$\{\boldsymbol{x}_k\}$收敛于$\boldsymbol{x}^*$，则当$G(\boldsymbol{x})$和$J(\boldsymbol{x})^T J(\boldsymbol{x})$在$\boldsymbol{x}^*$的邻域内Lipschitz连续时，有：

$$\|\boldsymbol{h}_{k+1}\| \leq \|(J^{*T} J^*)^{-1}\| \|S^*\| \|\boldsymbol{h}_k\| + O(\|\boldsymbol{h}_k\|^2) \tag{5.2.8}$$

其中：$\boldsymbol{h}_k = \boldsymbol{x}_k - \boldsymbol{x}^*$。

证明：因为$f \in C^2$,且$G(\boldsymbol{x})$在$\boldsymbol{x}^*$邻域内Lipschitz连续，当$\boldsymbol{x}_k$充分接近$\boldsymbol{x}^*$时，由**牛顿法收敛性定理**（教材50页，定理3.5）证明知：

$$\boldsymbol{g}(\boldsymbol{x}_k + \boldsymbol{d}) = \boldsymbol{g}_k + G_k \boldsymbol{d} + O(\|\boldsymbol{d}\|^2) \tag{a}$$

定理5.2.1证明

证明续：令$\boldsymbol{d} = -\boldsymbol{h}_k$, 得

$$0 = \boldsymbol{g}^* = \boldsymbol{g}_k - G_k\boldsymbol{h}_k + O(\|\boldsymbol{h}_k\|^2) \quad \text{(b)}$$

由(5.2.2)和(5.2.3)两式计算$\boldsymbol{g}_k$和G_k代入上式得：

$$J_k^T\boldsymbol{r}_k - (J_k^TJ_k + S_k)\boldsymbol{h}_k + O(\|\boldsymbol{h}_k\|^2) = 0. \quad \text{(c)}$$

由于$J^{*T}J^*$正定，当$\boldsymbol{x}_k$充分靠近$\boldsymbol{x}^*$时，$J_k^TJ_k$亦正定。用$(J_k^TJ_k)^{-1}$左乘(c)式
由(5.2.7)式得：

$$-\boldsymbol{d}_k - \boldsymbol{h}_k - (J_k^TJ_k)^{-1}S_k\boldsymbol{h}_k + O(\|\boldsymbol{h}_k\|^2) = 0.$$

又因为：

$$\boldsymbol{d}_k + \boldsymbol{h}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k + \boldsymbol{x}_k - \boldsymbol{x}^* = \boldsymbol{h}_{k+1}$$

定理5.2.1证明

证明续：所以：

$$
\begin{aligned}
\|\boldsymbol{h}_{k+1}\| &\le \|(J_k^T J_k)^{-1} S_k\|\|\boldsymbol{h}_k\| + O(\|\boldsymbol{h}_k\|^2) \\
&\le \|(J_k^T J_k)^{-1} S_k - ({J^*}^T J^*)^{-1} S^*\|\|\boldsymbol{h}_k\| \\
&\quad + \|({J^*}^T J^*)^{-1}\|\|S^*\|\|\boldsymbol{h}_k\| + O(\|\boldsymbol{h}_k\|^2) \qquad \text{(d)}
\end{aligned}
$$

下证$S(\boldsymbol{x})$与$(J(\boldsymbol{x})^T J(\boldsymbol{x}))^{-1}$在$\boldsymbol{x}^*$的邻域内Lipschitz连续。对于矩阵$A(\boldsymbol{x})$, 采用记号: $A(\boldsymbol{x}) = A_{\boldsymbol{x}}$。因为$G_{\boldsymbol{x}}$和$J_{\boldsymbol{x}}^T J_{\boldsymbol{x}}$在$\boldsymbol{x}^*$的邻域内Lipschitz连续，所以存在$\beta, \gamma > 0$, 使得对$\boldsymbol{x}^*$的邻域内任意两点$\boldsymbol{x}, \boldsymbol{y}$, 有:

$$
\|G_{\boldsymbol{x}} - G_{\boldsymbol{y}}\| \le \beta\|\boldsymbol{x} - \boldsymbol{y}\|, \quad \|J_{\boldsymbol{x}}^T J_{\boldsymbol{x}} - J_{\boldsymbol{y}}^T J_{\boldsymbol{y}}\| \le \gamma\|\boldsymbol{x} - \boldsymbol{y}\|
$$

从而：

$$
\begin{aligned}
\|S_{\boldsymbol{x}} - S_{\boldsymbol{y}}\| = \|G_{\boldsymbol{x}} - G_{\boldsymbol{y}} - J_{\boldsymbol{x}}^T J_{\boldsymbol{x}} + J_{\boldsymbol{y}}^T J_{\boldsymbol{y}}\| &\le \|G_{\boldsymbol{x}} - G_{\boldsymbol{y}}\| + \|J_{\boldsymbol{x}}^T J_{\boldsymbol{x}} - J_{\boldsymbol{y}}^T J_{\boldsymbol{y}}\| \\
&\le (\beta + \gamma)\|\boldsymbol{x} - \boldsymbol{y}\|
\end{aligned}
$$

定理5.2.1证明

证明续：对$\boldsymbol{x}^*$邻域内的任意点$\boldsymbol{x}$，由于$J^{*T}J^*$的正定性知，存在$\xi > 0$, 使得$\|({J_{\boldsymbol{x}}}^T J_{\boldsymbol{x}})^{-1}\| \le \xi$, 从而：

$$\begin{aligned}\|(J_{\boldsymbol{x}}^T J_{\boldsymbol{x}})^{-1} - (J_{\boldsymbol{y}}^T J_{\boldsymbol{y}})^{-1}\| &= \|(J_{\boldsymbol{x}}^T J_{\boldsymbol{x}})^{-1}(J_{\boldsymbol{y}}^T J_{\boldsymbol{y}} - J_{\boldsymbol{x}}^T J_{\boldsymbol{x}})(J_{\boldsymbol{y}}^T J_{\boldsymbol{y}})^{-1}\| \\ &\le \|(J_{\boldsymbol{x}}^T J_{\boldsymbol{x}})^{-1}\|\|(J_{\boldsymbol{y}}^T J_{\boldsymbol{y}})^{-1}\|\|(J_{\boldsymbol{y}}^T J_{\boldsymbol{y}} - J_{\boldsymbol{x}}^T J_{\boldsymbol{x}})\| \\ &\le \gamma\xi^2(\|\boldsymbol{x} - \boldsymbol{y}\|)\end{aligned}$$

所以$S_{\boldsymbol{x}}$与$(J_{\boldsymbol{x}}^T J_{\boldsymbol{x}})^{-1}$在$\boldsymbol{x}^*$的邻域内Lipschitz连续。
当$\boldsymbol{x}_k$充分靠近$\boldsymbol{x}^*$时，有：

$$\begin{aligned}\|(J_k^T J_k)^{-1}S_k - (J^{*T}J^*)^{-1}S^*\| &\le \|(J_k^T J_k)^{-1}S_k - (J_k^T J_k)^{-1}S^*\| \\ &\quad + \|(J_k^T J_k)^{-1}S^* - (J^{*T}J^*)^{-1}S^*\| \\ &\le (\beta+\gamma)\|(J_k^T J_k)^{-1}\|\|\boldsymbol{h}_k\| + \gamma\xi^2\|S^*\|\|\boldsymbol{h}_k\| \\ &\le ((\beta+\gamma)\xi + \gamma\xi^2\|S^*\|)\|\boldsymbol{h}_k\|\end{aligned}$$

定理5.2.1证明

证明续：代入(d)式得：

$$\|\boldsymbol{h}_{k+1}\| \leq \|({J^*}^T J^*)^{-1}\| \|S^*\| \|\boldsymbol{h}_k\| + O(\|\boldsymbol{h}_k\|^2)$$

故结论成立。 ■

该定理说明：若$\boldsymbol{x}_k \to \boldsymbol{x}^*$，基本Gauss-Newton方法有如下两种情形得收敛速度：

- 二阶收敛速度. 若$\|S(\boldsymbol{x}^*)\| = 0$，即在零剩余问题或者线性最小二乘情形，则方法在$\boldsymbol{x}^*$附近具有Newton方法得收敛速度。
- 线性收敛速度. 若$\|S(\boldsymbol{x}^*)\| \neq 0$, 则方法的收敛速度是线性的，收敛速度随着$\|S^*\|$的增大而变慢。

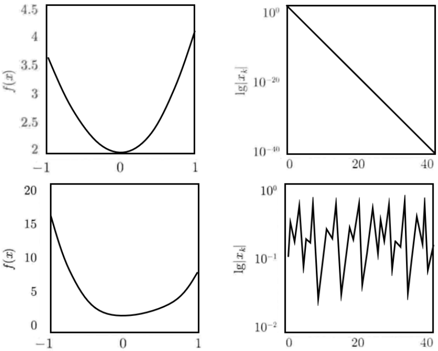
注意： 对于剩余量很大或者剩余函数非线性程度很强的问题，Gauss-Newton方法可能不收敛。

例题收敛性分析

例5.2.1 (教材125页) 考虑最小二乘问题：

$$\min f(\boldsymbol{x}) = \frac{1}{2}[(x+1)^2 + (\lambda x^2 + x - 1)^2], x \in \mathbb{R}$$

该问题当$\lambda < 1$时，有全局解$x^* = 0$.



在$\lambda = 0.1$(上)与$\lambda = -1$(下)两种情形下，$f(\boldsymbol{x})$的图像(左)与$\{\log|x_k|\}$的轨迹(右)

定理5.2.2 阻尼高斯-牛顿法全局收敛性.

设有界水平集$L(\boldsymbol{x}_0)=\{\boldsymbol{x}|f(\boldsymbol{x})\le f(\boldsymbol{x}_0)\}$上，$r_i(\boldsymbol{x})(i=1,\ldots,m)$连续可微，$J(\boldsymbol{x})$列满秩，则对采用Wolfe准则的阻尼Gauss-Newton方法产生的$\{\boldsymbol{x}_k\}$，或者存在N，使得$\boldsymbol{g}_N=0$，或者$\boldsymbol{g}_k\to 0, k\to\infty$。

证明：由于$L(\boldsymbol{x}_0)$是有界闭集，故$\boldsymbol{g}(\boldsymbol{x})$在$L(\boldsymbol{x}_0)$上一致连续. 由非精确线性搜索收敛性定理(教材24页，定理2.8)，只需要证明$\boldsymbol{d}_k$和$-\boldsymbol{g}_k$之间的夹角一致有界.

由$r_i(\boldsymbol{x})$在$L(\boldsymbol{x}_0)$上连续可微知，存在$\xi>0$，使对任意$\boldsymbol{x}\in L(\boldsymbol{x}_0)$，有$\|\nabla r_i(\boldsymbol{x})\|\le\xi, i=1,\ldots,m$，从而：

$$\|J(\boldsymbol{x})\|_\infty=\max_{1\le i\le m}\sum_{j=1}^{n}\left|\frac{\partial r_i(\boldsymbol{x})}{\partial x_j}\right|\le\xi,\forall\boldsymbol{x}\in L(\boldsymbol{x}_0)$$

定理5.2.2证明

证明续：由范数的等价性知，存在$\bar{\xi} > 0$, 使得：

$$\|J(\boldsymbol{x})\| = \|J(\boldsymbol{x})^T\| \leq \bar{\xi}$$

由Gauss-Newton方程可得：

$$\cos\theta_k = -\frac{\boldsymbol{g}_k^T\boldsymbol{d}_k}{\|\boldsymbol{g}_k\|\|\boldsymbol{d}_k\|} = -\frac{\boldsymbol{r}_k^T J_k\boldsymbol{d}_k}{\|J_k^T\boldsymbol{r}_k\|\|\boldsymbol{d}_k\|} = \frac{\|J_k\boldsymbol{d}_k\|^2}{\|J_k^T J_k\boldsymbol{d}_k\|\|\boldsymbol{d}_k\|} \geq \frac{\delta^2}{\xi^2} > 0$$

由非精确线性搜索收敛性定理(教材定理2.8)知结论成立。 ■

定理5.2.2要求$J(\boldsymbol{x})$列满秩，否则$J(\boldsymbol{x})^T J(\boldsymbol{x})$奇异，此时不能从Gauss-Newton方程(5.2.7)式中解出$\boldsymbol{d}_k$.

Gauss-Newton方程求解

线性最小二乘问题(5.2.6)求解和Gauss-Newton方程(5.2.7)求解是等价的。注意两者中相应矩阵J与J^TJ条件数有如下关系：

$$\text{cond}^2(J) = \text{cond}(J^TJ)$$

因此，Gauss-Newton方程(5.2.7)条件数是最小二乘问题(5.2.6)条件数的平方, 这使得采用Gauss-Newton方程求解过程**增加了对舍入误差的敏感性**, 直接求解问题(5.2.6)可避免这个问题。

求解的方法使用**正交化方法**，考虑2范数的正交不变性有：

$$\|J_k\boldsymbol{d} + \boldsymbol{r}_k\| = \|Q_k^T(J_k\boldsymbol{d} + \boldsymbol{r}_k)\|.$$

其中：$Q_k \in \mathbb{R}^{m\times m}$为任意正交矩阵。

Gauss-Newton方程求解

求解问题(5.2.6)等价于求解：

$$\min \frac{1}{2}\|Q_k^T(J_k\boldsymbol{d}+\boldsymbol{r}_k)\|^2.$$

首先对于J_k进行QR分解：

$$J_k = Q_k \begin{bmatrix} R_k \\ \mathbf{0} \end{bmatrix} \tag{5.2.9}$$

其中: $Q_k \in \mathbb{R}^{m\times m}$为正交阵，$R_k \in \mathbb{R}^{n\times n}$是具有非负对角元的上三角阵，$\mathbf{0} \in \mathbb{R}^{(m-n)\times m}$是零矩阵。进一步对$Q_k$进行分块:

$$Q_k = \begin{bmatrix} (Q_1^{(k)})_{m\times n} & (Q_2^{(k)})_{m\times(m-n)} \end{bmatrix}$$

Gauss-Newton方程求解

令：

$$Q_k^T \boldsymbol{r}_k = \begin{bmatrix} Q_1^{(k)^T} \\ Q_2^{(k)^T} \end{bmatrix} \boldsymbol{r}_k = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix} \begin{matrix} {\scriptstyle n} \\ {\scriptstyle m-n} \end{matrix}$$

则：

$$\|J_k \boldsymbol{d} + \boldsymbol{r}_k\|^2 = \|Q_k^T J_k \boldsymbol{d} + Q_k^T \boldsymbol{r}_k\|^2 = \|R_k \boldsymbol{d} + \boldsymbol{b}_1\|^2 + \|\boldsymbol{b}_2\|^2$$

由此可知，$\boldsymbol{d}_k$是最小二乘问题(5.2.6)的解当且仅当$\boldsymbol{d}_k$是$R_k \boldsymbol{d} = -\boldsymbol{b}_1$的解。综上所述，利用QR分解求最小二乘问题(5.2.6)的解的基本步骤：

- 计算J_k的QR分解(5.2.9);
- 计算$\boldsymbol{b}_1 = Q_1^{(k)^T} \boldsymbol{r}_k$;
- 求解上三角方程组$R_k \boldsymbol{d} = -\boldsymbol{b}_1$，即得到$\boldsymbol{d}_k$。