5-3 LMF方法与Dogleg方法

LM方法

Gauss-Newton方法在迭代过程中会出现$J_k^T J_k$为奇异的情形。为了克服这个困难，Levenberg在1944年提出了求解如下方程获得$\boldsymbol{d}_k$:

$$(J_k^T J_k + \nu_k I)\boldsymbol{d} = -J_k^T \boldsymbol{r}_k \tag{5.3.1}$$

其中：$\nu_k \geq 0$。1964年，这个方法经Marquardt的努力获得广泛应用，故称为LM(Levenberg-Marquardt)方法。(5.3.1)式也称为LM方程。

LM方程中，对任意$\nu_k > 0$, $J_k^T J_k + \nu_k I$正定，从而保证了(5.3.1)方程中得到的方向是下降方向。从计算的角度出发，ν_k的值可需要取得合适的大，保证$J_k^T J_k + \nu_k I$充分正定。

LM方法是一种信赖域方法，ν_k的值可以用信赖域方法的思想在迭代中修正得到。只要找出LM方程与信赖域问题的关系，就可以根据修正信赖域半径的方法来修正ν_k的值。

定理5.3.1 LM方程与信赖域问题的关系.

$\boldsymbol{d}_k$为信赖域子问题：

$$\min_{\boldsymbol{d}} \quad \frac{1}{2}\|J_k\boldsymbol{d}+\boldsymbol{r}_k\|^2, \tag{5.3.2a}$$

$$\text{s.t.} \quad \|\boldsymbol{d}\|^2 \le \Delta_k^2, \Delta_k > 0 \tag{5.3.2b}$$

的全局极小解的充要条件是，对满足(5.3.2b)的$\boldsymbol{d}_k$，存在$\nu_k \ge 0$, 使得：

$$(J_k^T J_k + \nu_k I)\boldsymbol{d}_k = -J_k^T \boldsymbol{r}_k \tag{5.3.3a}$$

$$\nu_k(\Delta_k^2 - \|\boldsymbol{d}_k\|^2) = 0. \tag{5.3.3b}$$

证明：**必要性** 由约束优化问题的最优性条件，Lagrange函数：

$$L(\boldsymbol{d},\nu) = q_k(\boldsymbol{d}) - \frac{1}{2}\nu(\Delta_k^2 - \|\boldsymbol{d}\|^2)$$

定理5.3.1证明

证明续：必要性 存在$\nu_k \geq 0$，使得$(\boldsymbol{d}_k, \nu_k)$是如下Lagrange函数$L(\boldsymbol{d}, \nu)$的KKT对。这里：$q_k(\boldsymbol{d}) = \frac{1}{2}\|J_k\boldsymbol{d} + \boldsymbol{r}_k\|^2$。

$(\boldsymbol{d}_k, \nu_k)$ 满足：

$$\nabla_{\boldsymbol{d}} L(\boldsymbol{d}_k, \nu_k) = 0 \Rightarrow J_k^T \boldsymbol{r}_k + (J_k^T J_k + \nu_k I)\boldsymbol{d}_k = 0$$

即得到(5.3.3a)式。由互补条件得: $\nu_k(\Delta_k^2 - \|\boldsymbol{d}_k\|^2) = 0$，即(5.3.3b)式。

充分性 因为$J_k^T J_k + \nu_k I$半正定，所以方程(5.3.3a)的解$\boldsymbol{d}_k$是如下二次函数全局极小点：

$$\tilde{q}_k(\boldsymbol{d}) = \frac{1}{2}\boldsymbol{d}^T(J_k^T J_k + \nu_k I)\boldsymbol{d} + \boldsymbol{d}^T(J_k^T \boldsymbol{r}_k) + \frac{1}{2}\boldsymbol{r}_k^T \boldsymbol{r}_k$$

定理5.3.1证明

证明续：由(5.2.6)式中$q_k(\boldsymbol{d})$计算公式知：

$$\tilde{q}_k(\boldsymbol{d}) = q_k(\boldsymbol{d}) + \frac{1}{2}\nu_k\|\boldsymbol{d}\|^2$$

对任意$\boldsymbol{d} \in \mathbb{R}^n$，都有$\tilde{q}_k(\boldsymbol{d}) \geq \tilde{q}_k(\boldsymbol{d}_k)$，所以：

$$q_k(\boldsymbol{d}) \geq q_k(\boldsymbol{d}_k) + \frac{1}{2}\nu_k(\|\boldsymbol{d}_k\|^2 - \|\boldsymbol{d}\|^2).$$

由(5.3.3b)知，若$\nu_k = 0$，有：$q_k(\boldsymbol{d}) \geq q_k(\boldsymbol{d}_k)$；若$\nu_k \neq 0$，有：$\|\boldsymbol{d}_k\|^2 = \Delta_k^2$，继而有：

$$q_k(\boldsymbol{d}) \geq q_k(\boldsymbol{d}_k) + \frac{1}{2}(\Delta_k - \|\boldsymbol{d}\|^2)$$

这说明，对任意$\nu_k \geq 0$和任意满足$\|\boldsymbol{d}\|^2 \leq \Delta_k^2$的$\boldsymbol{d}$，$\boldsymbol{d}_k$都是问题(5.3.2)的全局最优解。 ■

LMF方法

LM方程与信赖域问题的关系是Fletcher于1981年提出的，故由此建立起来的方法称为LMF(Levenberge-Marquardt-Fletcher)方法。

ν_k的修正方法与信赖域半径Δ_k修正是相关的。在信赖域方法中，从$\boldsymbol{x}_k$到$\boldsymbol{x}_k + \boldsymbol{d}_k$的实际减少量为：

$$\Delta f_k = f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \boldsymbol{d}_k)$$

$f(\boldsymbol{x}_k + \boldsymbol{d})$的二次近似函数$q_k(\boldsymbol{d})$的减少量为：

$$\Delta q_k = q_k(0) - q_k(\boldsymbol{d}_k)$$

这里$q_k(0) = f_k$。

LMF方法

另外，由LM方程以及$\boldsymbol{d}_k^T \boldsymbol{g}_k < 0$知：

$$\begin{aligned}
\Delta q_k &= q_k(0) - q_k(\boldsymbol{d}_k) \\
&= -\frac{1}{2}\boldsymbol{d}_k^T J_k^T J_k \boldsymbol{d}_k - \boldsymbol{d}_k^T (J_k^T \boldsymbol{r}_k) \\
&= \frac{1}{2}\boldsymbol{d}_k^T (-J_k^T J_k \boldsymbol{d}_k - \nu_k \boldsymbol{d}_k + \nu_k \boldsymbol{d}_k - 2J_k^T \boldsymbol{r}_k) \\
&= \frac{1}{2}\boldsymbol{d}_k^T (-(J_k^T J_k + \nu_k)\boldsymbol{d}_k + \nu_k \boldsymbol{d}_k - 2J_k^T \boldsymbol{r}_k) \\
&= \frac{1}{2}\boldsymbol{d}_k^T (\nu_k \boldsymbol{d}_k - \boldsymbol{g}_k) > 0 \qquad (5.3.4)
\end{aligned}$$

其中：$\boldsymbol{g}_k = J_k^T \boldsymbol{r}_k$. (公式(5.2.2) or 教材120页-公式(5.3))

LMF方法

定义:

$$\gamma_k = \frac{\Delta f_k}{\Delta q_k} \tag{5.3.5}$$

比值γ_k可以反映出$q_k(\boldsymbol{d}_k)$近似$f(\boldsymbol{x}_k + \boldsymbol{d}_k)$的好坏，在前一章中信赖域方法已经讨论过。

由LM方程知, ν_k可以控制$\|\boldsymbol{d}_k\|$的大小，从而控制信赖域的大小。若ν_k变大的化，$\|\boldsymbol{d}_k\|$会变小，反之亦然，所以对ν_k的大小修正，应该与信赖域方法中对于信赖域半径Δ_k的修正相反。

算法5.3.1 LMF方法

步1 给出$\boldsymbol{x} \in \mathbb{R}^n$, $\nu_0 > 0, \varepsilon > 0, k := 0$.

步2 若终止条件满足，则输出结果，停止迭代.

步3 求解LM方程(5.3.1)得$\boldsymbol{d}_k$.

步4 由(5.3.5)计算比值γ_k.

步5 若$\gamma_k < 0.25$，则$\nu_{k+1} := 4\nu_k$；若$\gamma_k > 0.75$，则$\nu_{k+1} = \frac{1}{2}\nu_k$；否则，$\nu_{k+1} := \nu_k$.

步6 若$\gamma_k \leq 0$，则$\boldsymbol{x}_{k+1} := \boldsymbol{x}_k$；否则，$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_k, k := k+1$，转步2.

LMF方法与修正牛顿方法

LMF方法可以用于求解一般无约束优化问题。在修正牛顿法中，修正牛顿方程和信赖域问题的关系如下：

定理5.3.2 修正牛顿方程与信赖域问题的关系.

$\boldsymbol{d}_k$为信赖域子问题：

$$\min_{\boldsymbol{d}} \quad q_k(\boldsymbol{d}) \tag{5.3.6a}$$

$$\text{s.t.} \quad \|\boldsymbol{d}\|^2 \leq \Delta_k^2, \Delta_k > 0 \tag{5.3.6b}$$

的全局极小解的充分必要条件是，对满足(5.3.6b)的$\boldsymbol{d}_k$，存在$\nu_k \geq 0$，使得：

$$(G_k + \nu_k I)\boldsymbol{d}_k = -\boldsymbol{g}_k \tag{5.3.7a}$$

$$\nu_k(\Delta_k^2 - \|\boldsymbol{d}_k\|^2) = 0 \tag{5.3.7b}$$

$$G_k + \nu_k I\text{半正定} \tag{5.3.7c}$$

Dogleg方法

在LM方法中，由LM方程得到的方向$\boldsymbol{d}_k^{\mathrm{LM}}$受到$\nu_k$取值的影响。当$\nu_k$很大时，$\boldsymbol{d}_k^{\mathrm{LM}}$偏向于负梯度方向；若$\nu_k$很小，则$\boldsymbol{d}_k^{\mathrm{LM}}$接近Gauss-Newton方向；否则，$\boldsymbol{d}_k^{\mathrm{LM}}$介于负梯度方向和Gauss-Newton方向之间。受此影响的启示，1970年Powell提出解如下信赖域子问题的Dogleg方法

$$\min \quad \frac{1}{2}\|J_k\boldsymbol{d}+\boldsymbol{r}_k\|^2 \tag{5.3.8a}$$

$$\text{s.t.} \quad \|\boldsymbol{d}\| \le \Delta_k, \Delta_k > 0 \tag{5.3.8b}$$

$\boldsymbol{d}_k^{\mathrm{GN}}$与$\boldsymbol{d}_k^{\mathrm{SD}}$分别由Gauss-Newton方程或者最速下降方向给出。最速下降法的步长为：

$$\alpha_k = \arg\min q_k(\alpha\boldsymbol{d}_k^{\mathrm{SD}}) = \frac{\|\boldsymbol{d}_k^{\mathrm{SD}}\|^2}{\|J_k\boldsymbol{d}_k^{\mathrm{SD}}\|^2}$$

其中：$q_k(\alpha\boldsymbol{d}_k^{\mathrm{SD}}) \triangleq \frac{1}{2}\|\alpha J_k\boldsymbol{d}_k^{\mathrm{SD}}+\boldsymbol{r}_k\|^2 = \frac{1}{2}\|J_k\boldsymbol{d}_k^{\mathrm{SD}}\|^2\alpha^2 - \|\boldsymbol{d}_k^{\mathrm{SD}}\|^2\alpha + f_k$.

Dogleg方法

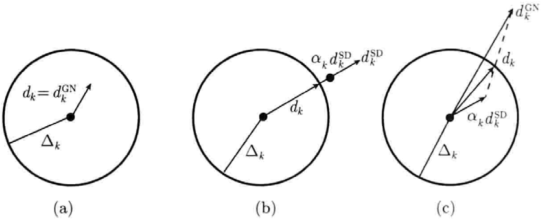
算法5.3.2 Dogleg方法求解最小二乘信赖域子问题

步1 给出$\Delta_k > 0, J_k, \boldsymbol{r}_k$;

步2 若$\|\boldsymbol{d}_k^{\mathrm{GN}}\| \le \Delta_k$, 则$\boldsymbol{d}_k = \boldsymbol{d}_k^{\mathrm{GN}}$, 输出$\boldsymbol{d}_k$, 迭代停止;

步3 若$\alpha_k\|\boldsymbol{d}_k^{\mathrm{SD}}\| \ge \Delta_k$, 则$\boldsymbol{d}_k = \frac{\Delta_k}{\|\boldsymbol{d}_k^{\mathrm{SD}}\|}\boldsymbol{d}_k^{\mathrm{SD}}$, 输出$\boldsymbol{d}_k$, 迭代停止;

步4 计算$\boldsymbol{d}_k = (1-\beta)\alpha_k\boldsymbol{d}_k^{\mathrm{SD}} + \beta\boldsymbol{d}_k^{\mathrm{GN}}$, 其中要确定$\beta$使得$\|\boldsymbol{d}_k\| = \Delta_k$, 输出$\boldsymbol{d}_k$, 迭代停止。



Dogleg方法

在算法5.3.2中

- 步3中$\alpha_k\|\boldsymbol{d}_k\| \geq \Delta_k$意味着$\Delta_k$相对较小，即$\|\boldsymbol{d}_k^{\mathrm{LM}}\|$比较小，此时对应$\nu_k$比较大，因此选择下降方向来确定$\boldsymbol{d}_k$。
- 若J_k列满秩，则$J_k^TJ_k$正定，Gauss-Newton方程有解。另外，只要$\boldsymbol{r}_k \neq 0$, $\boldsymbol{d}_k^{\mathrm{SD}}$就是下降方向。因此，Dogleg方法适合解决$J_k$列满秩的信赖域子问题。
- 欲使得$\|\boldsymbol{d}_k\| = \Delta_k$，即求函数

$$\tilde{q}(\beta) = \|\alpha_k\boldsymbol{d}_k^{\mathrm{SD}} + (\boldsymbol{d}_k^{\mathrm{GN}} - \alpha_k\boldsymbol{d}_k^{\mathrm{SD}})\beta\|^2 - \Delta_k^2$$

的零点。

大剩余量问题

Gauss-Newton方法时在Newton方法的Hesse阵中忽略了二阶项$S(\boldsymbol{x}) = \sum_{i=1}^{m} r_i(\boldsymbol{x})\nabla^2 r_i(\boldsymbol{x})$得到的。对于剩余量大或者剩余函数的非线性程度很高的问题，忽略二阶项会影响算法的收敛性和收敛速度。如果不忽略而直接计算二阶导数矩阵，将涉及到很大计算复杂性。

与拟牛顿法类似，利用一阶梯度信息构造$S(\boldsymbol{x})$的近似矩阵$\hat{B}$. 假定点$\boldsymbol{x}_k$已得到S_k的近似矩阵$\hat{B}_k$, 在点$\boldsymbol{x}_{k+1}$处满足什么条件的$\hat{H}_{k+1}$可以作为S_{k+1}的近似矩阵，使得：

$$J_{k+1}^T J_{k+1} + \hat{B}_{k+1} \approx G_{k+1}$$

这里$\hat{B}_{k+1}$应该满足：

$$\hat{B}_{k+1}\boldsymbol{s}_k = \hat{\boldsymbol{y}}_k$$

其中: $\hat{\boldsymbol{y}}_k = (J_{k+1} - J_k)^T \boldsymbol{r}_{k+1}$。

大剩余量问题

类似于拟牛顿法，在所有对称、满足拟牛顿条件矩阵中，寻找在加权范数意义下与$\hat{B}_k$的差最小的矩阵，即考虑如下问题：

$$
\begin{aligned}
\min_{\hat{B}} \quad & \|W^{-T}(\hat{B} - \hat{B}_k)W^{-1}\|_F \\
\text{s.t.} \quad & \hat{B} = \hat{B}^T, \hat{B}\boldsymbol{s}_k = \hat{\boldsymbol{y}}_k
\end{aligned}
$$

其中$W \in \mathbb{R}^{n\times n}$，$W^TW$满足拟牛顿条件$W^{-T}W\boldsymbol{s}_k = \boldsymbol{y}_k$。该问题的解为：

$$
\hat{B}_{k+1} = \hat{B}_k + \frac{(\hat{\boldsymbol{y}}_k - \hat{B}_k\boldsymbol{s}_k)\boldsymbol{y}_k^T + \boldsymbol{y}_k(\hat{\boldsymbol{y}}_k - \hat{B}_k\boldsymbol{s}_k)^T}{\boldsymbol{y}_k^T\boldsymbol{s}_k} - \frac{(\hat{\boldsymbol{y}}_k - \hat{B}_k\boldsymbol{s}_k)^T\boldsymbol{s}_k}{(\boldsymbol{y}_k^T\boldsymbol{s}_k)^2}\boldsymbol{y}_k\boldsymbol{y}_k^T.
$$

其中：

$$
\begin{aligned}
\boldsymbol{s}_k &= \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \\
\boldsymbol{y}_k &= J_{k+1}^T\boldsymbol{r}_{k+1} - J_k^T\boldsymbol{r}_k \\
\hat{\boldsymbol{y}}_k &= (J_{k+1} - J_k)^T\boldsymbol{r}_{k+1}.
\end{aligned}
$$