4-5 信赖域方法

信赖域方法

不同于线搜索方法先确定搜索方向$\boldsymbol{d}_k$后确定步长α_k, 信赖域方法无需单独确定步长而直接得到调整量$\boldsymbol{d}_k$, 产生下一个迭代点:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_k \tag{4.5.1}$$

这里$\boldsymbol{d}_k$通常可由极小化$\min_{\boldsymbol{d}} f(\boldsymbol{x}_k+\boldsymbol{d})$来获得。通过优化$f(\boldsymbol{x}_k+\boldsymbol{d})$在$\boldsymbol{x}_k$的一阶或者二阶近似函数可较简单获得$\boldsymbol{d}_k$。这里我们介绍基于二次函数近似的信赖域方法，即考虑：

$$\min_{\boldsymbol{d}} q_k(\boldsymbol{d}) = f(\boldsymbol{x}_k) + \boldsymbol{g}_k^T \boldsymbol{d} + \frac{1}{2}\boldsymbol{d} G_k \boldsymbol{d} \tag{4.5.2}$$

注意$f(\boldsymbol{x}_k + \boldsymbol{d})$与$q_k(\boldsymbol{d})$的误差为$o(\|\boldsymbol{d}\|^2)$。若$\boldsymbol{x}_k$的邻域过大，则会出现较大的逼近误差。若$\boldsymbol{x}_k$的邻域过小，移动过慢导致增加迭代次数。

信赖域方法

在每步迭代选取$\boldsymbol{x}_k$的一个合适邻域，称之为**信赖域**(trust region)，求解如下问题：

$$\min\ q_k(\boldsymbol{d}) \tag{4.5.3a}$$

$$\text{s.t. } \|\boldsymbol{d}\| \leq \Delta_k, \Delta_k > 0. \tag{4.5.3b}$$

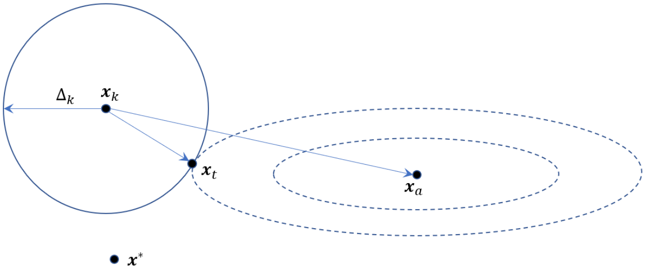
以上问题也称为**信赖域子问题**。

求得问题(4.5.3)的最优解$\boldsymbol{d}_k$后，再判断Δ_k是否是下一步的合适的信赖域半径。若不合适则修正Δ_k得到下一个信赖域半径Δ_{k+1}。

信赖域方法先限定了步长的范围，再同时决定迭代方向和步长。

信赖域方法

下图中$\boldsymbol{x}^*$表示问题最优解，Δ_k为当前信赖域半径，虚线为$q_k(\boldsymbol{d})$的等高线。



分别求解问题(4.5.2)和(4.5.3)得到$\boldsymbol{x}_a$与$\boldsymbol{x}_t$，容易发现$\boldsymbol{x}_t$比$\boldsymbol{x}_a$更接近问题最优解$\boldsymbol{x}^*$。信赖域半径的大小确定了不同的迭代方向。

信赖域半径调整

调整原则主要根据模型函数$q_k(\boldsymbol{d})$对目标函数$f(\boldsymbol{x})$的拟合程度来调整信赖域半径Δ_k，分别计算：

- 目标函数的实际下降量：

$$\Delta f_k = f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \boldsymbol{d}_k) \tag{4.5.4}$$

- 模型函数的预测下降量:

$$\Delta q_k = q_k(0) - q_k(\boldsymbol{d}_k) \tag{4.5.5}$$

这里$q_k(0) = f(\boldsymbol{x}_k)$。

信赖域半径调整

定义比值

$$\gamma_k = \frac{\Delta f_k}{\Delta q_k}, \tag{4.5.6}$$

该比值衡量模型函数$q_k(\boldsymbol{d})$与目标函数$f(\boldsymbol{x})$的一致性程度。

- γ_k越接近1, 表明模型函数q_k与目标函数f的一致性程度越好, 下一步应增大Δ_k以扩大信赖域。
- γ_k接近零或取负值，表明q_k与f之间的一致性程度不理想，下一步应减小Δ_k以缩小信赖域。
- $\gamma_k > 0$但不接近1，保持信赖域半径Δ_k不变。

信赖域算法步骤

算法4.5.1 信赖域算法

步1 给出初始点$\boldsymbol{x}_0$, $\varepsilon > 0$, $\Delta_0 > 0$, $k := 0$.

步2 若停机准则满足，输出$\boldsymbol{x}_k$, 停止迭代。

步3 求解子问题(4.5.3)得到$\boldsymbol{d}_k$。

步4 根据(4.5.6)计算比值γ_k，校正信赖域半径.令

$$
\begin{array}{ll}
\Delta_{k+1} \in (0, \tau_1 \Delta_k], & \text{如果 } \gamma_k < \eta_1; \\
\Delta_{k+1} \in [\tau_1 \Delta_k, \Delta_k], & \text{如果 } \gamma_k \in [\eta_1, \eta_2); \\
\Delta_{k+1} \in [\Delta_k, \min\{\tau_2 \Delta_k, \bar{\Delta}\}], & \text{如果 } \gamma_k \geq \eta_2.
\end{array}
$$

步5 若$\gamma_k \leq \eta_1$, 则$\boldsymbol{x}_{k+1} := \boldsymbol{x}_k$; 否则，$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_k$。转步2。

信赖域算法参数

在上面的算法中，

- 若$\gamma_k \geq \eta_2$，故$\Delta_{k+1} \geq \Delta_k$，这种情形称为**很成功迭代**；
- 类似地，$\gamma_k \in [\eta_1, \eta_2)$的情形称为**成功迭代**；
- $\gamma_k < \eta_1$的情形称为**不成功迭代**。

很成功迭代和成功迭代称为成功迭代。

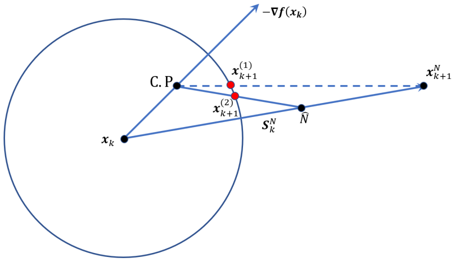
上面算法中的一些参数选择可以建议如下：

$$\eta_1 = 0.25,\ \eta_2 = 0.75,\ \tau_1 = 0.5,\ \tau_2 = 2,$$

$$\Delta_0 = 1 \quad \text{或} \quad \Delta_0 = \frac{1}{10}\|\boldsymbol{g}_0\|$$

信赖域算法4.5.1中最关键的第3步是信赖域子问题求解。接下来介绍解信赖域子问题的折线法(dog-leg method, Powell(1970))。

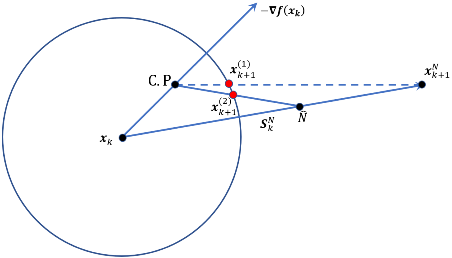
为近似求解子问题(4.5.3)，即求$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{s}_k$满足$\|\boldsymbol{s}_k\| \le \Delta_k$，Cauchy点(最速下降法产生的极小点C.P.)和牛顿点(牛顿法产生的极小点$\boldsymbol{x}_{k+1}^N$)



折线法

折线法做法是:

- 连接Cauchy点与牛顿点，其连线与信赖域边界的交点取为$\boldsymbol{x}_{k+1}$. 此时有$\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|=\Delta_k$.
- 当牛顿步$\boldsymbol{s}_k^N$满足$\|\boldsymbol{s}_k^N\|\le\Delta_k$时，则$\boldsymbol{x}_{k+1}$取为牛顿点$\boldsymbol{x}_{k+1}^N$。



解信赖域子问题

最速下降法极小化二次模型：

$$q_k(\boldsymbol{x}_k - \alpha \boldsymbol{g}_k) = f(\boldsymbol{x}_k) - \alpha \|\boldsymbol{g}_k\|_2^2 + \frac{1}{2}\alpha^2 \boldsymbol{g}_k^T G_k \boldsymbol{g}_k, \tag{4.5.7}$$

精确线性搜索因子α_k的表达式为:

$$\alpha_k = \frac{\|\boldsymbol{g}_k\|_2^2}{\boldsymbol{g}_k^T G_k \boldsymbol{g}_k}. \tag{4.5.8}$$

Cauchy步为:

$$\boldsymbol{s}_k^c = -\alpha_k \boldsymbol{g}_k = -\frac{\boldsymbol{g}_k^T \boldsymbol{g}_k}{\boldsymbol{g}_k^T G_k \boldsymbol{g}_k} \boldsymbol{g}_k. \tag{4.5.9}$$

解信赖域子问题

情形1： 若$\|\boldsymbol{s}_k^c\|_2 = \|\alpha_k \boldsymbol{g}_k\|_2 \geq \Delta_k$，取

$$\boldsymbol{s}_k = -\frac{\Delta_k}{\|\boldsymbol{g}_k\|_2}\boldsymbol{g}_k, \tag{4.5.10}$$

此时，

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \frac{\Delta_k}{\|\boldsymbol{g}_k\|_2}\boldsymbol{g}_k. \tag{4.5.11}$$

情形2： 若$\|\boldsymbol{s}_k^c\|_2 = \|\alpha_k \boldsymbol{g}_k\|_2 < \Delta_k$，进一步计算牛顿步$\boldsymbol{s}_k^N$，

$$\boldsymbol{s}_k^N = -G_k^{-1}\boldsymbol{g}_k. \tag{4.5.12}$$

解信赖域子问题

类似地，对牛顿步也分两种情况：

情形1：若$\|\boldsymbol{s}_k^N\|_2 \leq \Delta_k$, 则取

$$\boldsymbol{s}_k = \boldsymbol{s}_k^N = -G_k^{-1}\boldsymbol{g}_k, \tag{4.5.13}$$

情形2：否则，取

$$\boldsymbol{s}_k = \boldsymbol{s}_k^c + \lambda(\boldsymbol{s}_k^N - \boldsymbol{s}_k^c), \tag{4.5.14}$$

其中λ的值由解方程

$$\|\boldsymbol{s}_k^c + \lambda(\boldsymbol{s}_k^N - \boldsymbol{s}_k^c)\| = \Delta_k$$

得到。

单折线法

综上所述，得到单折线法迭代格式：

$$
\boldsymbol{x}_{k+1} = \begin{cases} \boldsymbol{x}_k - \frac{\Delta_k}{\|\boldsymbol{g}_k\|_2}\boldsymbol{g}_k, & \|\boldsymbol{s}_k^c\| \geq \Delta_k \\ \boldsymbol{x}_k + \boldsymbol{s}_k^c + \lambda(\boldsymbol{s}_k^N - \boldsymbol{s}_k^c), & \|\boldsymbol{s}_k^c\| < \Delta_k \text{且} \|\boldsymbol{s}_k^N\| > \Delta_k \\ \boldsymbol{x}_k - G_k^{-1}\boldsymbol{g}_k, & \|\boldsymbol{s}_k^c\| < \Delta_k \text{且} \ \|\boldsymbol{s}_k^N\| \leq \Delta_k \end{cases} \tag{4.5.15}
$$

上述折线方法满足:

1) 从点$\boldsymbol{x}_k$到Cauchy点C.P.，到牛顿点$\boldsymbol{x}_{k+1}^N$的距离单调增加；
2) 从Cauchy点C.P.到牛顿点$\boldsymbol{x}_{k+1}^N$，模型函数值单调减少。

双折线法

在Powell单折线方法的基础上，Dennis和Mei(1979)提出了双折线法，其具体做法是把Cauchy点和牛顿方向上的$\hat{N}$点连接起来，并将这条连线与信赖域边界的交点取为$\boldsymbol{x}_{k+1}$. 双折线迭代格式为：

$$\boldsymbol{x}_{k+1} = \begin{cases} \boldsymbol{x}_k - \frac{\Delta_k}{\|\boldsymbol{g}_k\|_2}\boldsymbol{g}_k, & \|\boldsymbol{s}_k^c\| \geq \Delta_k \\ \boldsymbol{x}_k + \boldsymbol{s}_k^c + \lambda(\boldsymbol{s}_k^{\hat{N}} - \boldsymbol{s}_k^c), & \|\boldsymbol{s}_k^c\| < \Delta_k \text{且} \|\boldsymbol{s}_k^{\hat{N}}\| > \Delta_k \\ \boldsymbol{x}_k - \eta G_k^{-1}\boldsymbol{g}_k, & \|\boldsymbol{s}_k^c\| < \Delta_k \text{且} \|\boldsymbol{s}_k^{\hat{N}}\| \leq \Delta_k \end{cases} \tag{4.5.16}$$

其中，

$$\boldsymbol{s}_k^{\hat{N}} = \eta \boldsymbol{s}_k^N, \eta \in (\gamma, 1), \gamma = \frac{\|\boldsymbol{g}_k\|_2^4}{(\boldsymbol{g}_k^T G_k \boldsymbol{g}_k)(\boldsymbol{g}_k^T G_k^{-1} \boldsymbol{g}_k)}. \tag{4.5.17}$$

当$\eta = 1$, (4.5.16)就是Powell单折线法(4.5.15)。一般取$\eta = 0.8\gamma + 0.2$.

双折线法

上述双折线法也满足：

1)从点$\boldsymbol{x}_k$到Cauchy点C.P.，到$\boldsymbol{x}_{k+1}^{\hat{N}}$的距离单调增加；

2)从Cauchy点C.P.到点$\boldsymbol{x}_{k+1}^{\hat{N}}$,模型函数值单调减少。

在产生C.P.点和$\hat{N}$点后，所求的新点$\boldsymbol{x}_{k+1}(\lambda)$由(4.5.16)产生，解方程$\|s_k^c+\lambda(\eta \boldsymbol{s}_k^N-\boldsymbol{s}_k^c)\|_2=\Delta_k$.得到$\lambda$。

若$\boldsymbol{x}_{k+1}(\lambda)$满足下降性要求:

$$f(\boldsymbol{x}_{k+1}(\lambda)) \leq f(\boldsymbol{x}_k)+\rho \boldsymbol{g}_k^T(\boldsymbol{x}_{k+1}(\lambda)-\boldsymbol{x}_k), \rho \in (0, \frac{1}{2}), \tag{4.5.18}$$

则接受$\boldsymbol{x}_{k+1}(\lambda)$为新点$\boldsymbol{x}_{k+1}$，并根据信赖域**算法**4.5.1步5校正信赖域半径。

双折线法

若$\boldsymbol{x}_{k+1}(\lambda)$不满足(4.5.18), 则令$\boldsymbol{x}_{k+1} := \boldsymbol{x}_k$, 并缩小信赖域半径。

在折线法中，可以考虑将牛顿点改为拟牛顿点或改进的牛顿点，从而可以得到折线法的若干其他变形.

折线法算例

例4.5.1 设$f(\boldsymbol{x}) = x_1^4 + x_1^2 + x_2^2$，当前点$\boldsymbol{x}_k = (1,1)^T, \Delta_k = \frac{1}{2}$，试用双折线法求$\boldsymbol{x}_{k+1}$。

解： 由于$\boldsymbol{x}_k = (1,1)^T$，由计算可得

$$\boldsymbol{g}_k = \begin{pmatrix} 6, & 2 \end{pmatrix}^T, \qquad G_k = \begin{bmatrix} 14 & 0 \\ 0 & 2 \end{bmatrix},$$

$$\boldsymbol{s}_k^N = -G_k^{-1}\boldsymbol{g}_k = (-\frac{3}{7}, -1)^T.$$

$$\boldsymbol{s}_k^c = -\frac{\|\boldsymbol{g}_k\|_2^2}{\boldsymbol{g}_k^T G_k \boldsymbol{g}_k}\boldsymbol{g}_k = -\left(\frac{40}{512}\right)\begin{pmatrix} 6 \\ 2 \end{pmatrix} \cong \begin{pmatrix} -0.469 \\ -0.156 \end{pmatrix}.$$

折线法算例

由于$\|\boldsymbol{s}_k^c\| \cong 0.496 < \Delta_k$, 计算到$\hat{N}$的步长$\boldsymbol{s}_k^{\hat{N}}$,有

$$\gamma = \frac{\|\boldsymbol{g}_k\|_2^4}{(\boldsymbol{g}_k^T G_k \boldsymbol{g}_k)(\boldsymbol{g}_k^T G_k^{-1} \boldsymbol{g}_k)} = \frac{(40)^2}{(512)(\frac{32}{7})} \cong 0.684,$$

$$\eta = 0.8\gamma + 0.2 \cong 0.747,$$

$$\boldsymbol{s}_k^{\hat{N}} = \eta \boldsymbol{s}_k^N \cong \begin{pmatrix} -0.320 \\ -0.747 \end{pmatrix}.$$

由于$\|\boldsymbol{s}_k^{\hat{N}}\|_2 \cong 0.813 > \Delta_k$, 故取双折线步长为:

$$\boldsymbol{s}_k = \boldsymbol{s}_k^c + \lambda(\boldsymbol{s}_k^{\hat{N}} - \boldsymbol{s}_k^c), \lambda \in (0, 1),$$

使得$\|\boldsymbol{s}_k\| = \Delta_k$.

折线法算例

解二次方程

$$\|\boldsymbol{s}_k^c + \lambda(\boldsymbol{s}_k^{\hat{N}} - \boldsymbol{s}_k^c)\|_2^2 = \Delta_k^2$$

得

$$\lambda \cong 0.867.$$

因此

$$\boldsymbol{s}_k = \boldsymbol{s}_k^c + 0.867(\boldsymbol{s}_k^{\hat{N}} - \boldsymbol{s}_k^c) \cong \begin{pmatrix} -0.340 \\ -0.669 \end{pmatrix},$$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{s}_k \cong \begin{pmatrix} -0.660 \\ -0.331 \end{pmatrix}.$$

□