

# 信赖域方法

- 信赖域方法
  - \* 信赖域方法概述
  - \* Algorithm 4.1 (Trust Region).
  - \* Theorem 4.1
    - 基于柯西点的算法
      - 柯西点
        - \* Algorithm 4.2 (Cauchy Point Calculation).
      - 为什么要改良柯西法
      - 狗腿法
        - \* Lemma 4.2
      - 二维子空间极小化
    - 全局收敛性
      - 柯西点收益
        - 引理4.3
        - 定理4.4
      - 平稳点的收敛性
        - 定理4.5
        - 定理4.6
    - 子问题的迭代解法
      - \* Alorithm 4.3 (Trust Reegion Subproblem)
        - 困难的情况
        - 定理4.1的证明
          - 引理4.7
        - 基于近似精确解的算法的收敛性
          - 定理4.8
    - 信赖域算法的局部收敛性
      - \* 定理4.9
    - 其他改进方法
      - Scaling
        - \* Algorithm 4.4 Generalized Cauchy Point Calculation
      - 其他形式的信赖域方法

	LSM	TR
Step.1	确定一个搜索方向 $p_k$	确定近似模型 $m_k(p)$ 和信赖域半径 $\Delta_k$

	LSM	TR
Step.2	确定合适的步长 $\alpha_k$	寻找 $m_k$ 在信赖域中的最小值 $p_k$ ，检查是否合适。若合适，则 $x_{k+1} = x_k + p_k$ ；若不合适，缩小 $\Delta_k$ ，再次寻找最小值 $p_k$

线性搜索法(LSM)和信赖域方法(TR)都是基于目标函数的二次模型迭代，不同的是：**LSM**用二次模型产生搜索方向，再沿着这个方向寻找一个合适的步长；而**TR**则是假设这个模型是原函数的近似模型，寻找这个模型的最小值，再检查这个最小值是否合适，如果不合适，它们就会缩小信赖域的范围，再寻找一个新的最小值，通常，信赖域范围改变时，最小值的方向也会改变。

信赖域的大小是**TR**很关键的参数。如果信赖域过小，则可能失去快速下降的机会；然而如果信赖域过大，则可能导致信赖域模型的最小值与原函数的最小值距离很远，这时我们又必须缩小信赖域重新寻找迭代点。在**实际应用中**，我们常常根据算法最近的表现效果去选择我们的信赖域大小。

本章，我们假设每次在 $x_k$ 处迭代用的模型 $m_k$ 都是二次函数。更进一步， $m_k$ 是 $f$ 在 $x_k$ 处的泰勒展开(4.1)

$$f(x_k + p) = f_k + g_k^T p + \frac{1}{2} p^T \nabla^2 f(x_k + tp) p$$

其中， $f_k = f(x_k)$ ,  $g_k = \nabla f(x_k)$ ,  $t \in (0, 1)$ . 利用 $B_k$ 逼近Hessian矩阵，我们得到(4.2)

$$m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p$$

其中， $B_k$ 是对称阵。模型的误差是 $O(\|p\|^2)$

当 $B_k = \nabla^2 f(x_k)$ ，模型的误差是 $O(\|p\|^3)$ 。此时，这种方法被称为信赖域牛顿方法。

为了一般性，大部分时候我们都只假设 $B_k$ 是对称且一致有界的。

在每一步迭代过程，我们会求解下述子问题(4.3):

$$\min_{p \in R^n} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p \quad \text{s.t. } \|p\| \leq \Delta_k$$

其中 $\Delta_k > 0$ 是信赖域半径，大多数情况我们使用Euclidean范数。因此**TR**要求我们求解一系列目标和约束都是二次函数的子问题。当 $B_k$ 正定且 $\|B_k^{-1} g_k\| \leq \Delta_k$ ，子问题(4.3)是很简单的，且 $p_k^B = -B_k^{-1} g_k$ ，此时我们称 $p_k^B$ 是完全步。其他时候，(4.3)往往不是显然的，但是它往往可以通过少量的计算进行求解。大部分时候我们只需要近似解就能获得很好的收敛效果。

## 信赖域方法概述

信赖域方法在选择信赖域半径时有一个重要指标： $\rho_k$ ，它是用来刻画模型逼近效果和下降效果的，其定义如下：(4.4)

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$$

首先，根据定义分母恒非负，因此，当 $\rho_k < 0$ ，函数值增加，显然要拒绝当前迭代，缩小半径重新计算；当 $\rho_k$ 约等于1，说明模型的近似效果很好，信赖域半径保持；当 $\rho_k > 0$ 但是远小于1，则说明模型

的近似效果出了问题，根据模型性质，我们应适当缩小半径，重新进行计算。

可以看出这个指标有重要的实践意义，同时，在后续分析中，我们还可以看到它贡献了极高的理论价值。

下面是信赖域方法的一个框架

### Algorithm 4.1 (Trust Region).

```

Given  $\Delta_m > 0$ ,  $\Delta_0 \in (0, \Delta_m)$ ,  $\eta \in [0, 1/4)$ ;
for  $k = 0, 1, 2, \dots$ 
    Obtain  $p(k)$  by (approximately) solving (4.3);
    Evaluate  $\rho(k)$  from (4.4);
    if  $\rho(k) < 1/4$ 
         $\Delta(k+1) = 1/4 * \Delta(k)$ 
    else
        if  $\rho(k) > 3/4$  and  $\text{norm}(p) = \Delta(k)$ 
             $\Delta(k+1) = \min(2\Delta(k), \Delta_m)$ 
        else
             $\Delta(k+1) = \Delta(k)$ ;
    if  $\rho(k) > \eta$ 
         $x(k+1) = x(k) + p(k)$ 
    else
         $x(k+1) = x(k)$ 

```

值得注意的是，我们仅在模型拟合效果好且  $\|p_k\| = \Delta_k$  时才扩大信赖域。如果更新点在信赖域内，那么这个时候，我们称信赖域的半径在这个过程中是没有贡献的，因此我们保持其半径不变。

为了在实际应用中使用上述算法，我们主要要做的是求解信赖域子问题(4.3)。在分析这个子问题的时候，我们去掉下标，有(4.5):

$$\min_{p \in R^n} m(p) = f + g^T p + \frac{1}{2} p^T B p \quad \text{s.t. } \|p\| \leq \Delta$$

为了刻画(4.5)的解的性质，我们用下述定理说明，(4.5)的解  $p^*$  满足下述方程(4.6):

$$(B + \lambda I)p^* = -g$$

其中， $\lambda \geq 0$

### Theorem 4.1

$p^*$  是 (4.5) 的全局解当且仅当  $p^*$  是可行解，且存在  $\lambda \geq 0$  s.t.(4.8)

$$\begin{aligned} (B + \lambda I)p^* &= -g, \\ \lambda(\Delta - \|p^*\|) &= 0, \\ (B + \lambda I) &\text{ 半正定} \end{aligned}$$

我们将在4.3节证明该定理。

从结果上看，(4.8b)说明了信赖域子问题的最优解有如下两种形式：

1.  $\lambda = 0$ , 此时  $B$  半正定,  $Bp^* = -g$
2.  $\|p^*\| = \Delta$ , 此时  $\lambda p^* = -Bp^* - g = -\nabla m(p^*)$ , 即与模型负梯度方向共线

本章将从(4.3)的一个简单估计——柯西点法出发, 介绍两种常用的信赖域方法, 再对它们的性质进行分析。而在7.1会用共轭梯度法给出第三种方法。

## 基于柯西点的算法

### 柯西点

因为最速下降法有全局收敛性, 所以我们考虑在信赖域中引入类似的手段, 不过步长会被信赖域半径限制, 由此得到的方法, 即柯西点方法。

#### Algorithm 4.2 (Cauchy Point Calculation).

计算  $p_k^s$  s.t.(4.9)

$$p_k^s = \arg \min_{p \in R^n} f_k + g_k^T p \quad \text{s.t. } \|p\| \leq \Delta_k$$

计算  $\tau_k$  s.t.(4.10)

$$\tau_k = \arg \min_{\tau \geq 0} m_k(\tau p_k^s) \quad \text{s.t. } \|\tau p_k^s\| \leq \Delta_k$$

令  $p_k^C = \tau_k p_k^s$

实际上, 根据线性模型的性质

$$p_k^s = -\frac{\Delta_k}{\|g_k\|} g_k$$

将上述结果代入(4.10), 柯西点可表示为

$$p_k^C = -\tau_k \frac{\Delta_k}{\|g_k\|} g_k$$

其中

$$\tau_k = \begin{cases} 1 & \text{if } g_k^T B_k g_k \leq 0 \\ \min(\|g_k\|^3 / (\Delta_k g_k^T B_k g_k), 1) & \text{otherwise} \end{cases}$$

柯西法操作代价很小, 而且在判断子问题的有效性时有着十分重要的作用: 如果子问题解对于近似模型的优化效果比柯西法好正常数倍, 那么这种求解子问题的方法是全局收敛的。

### 为什么要改良柯西法

柯西法虽然能够实现全局收敛，但是我们注意到，柯西法的求解方向其实就是最速下降法的方向，而且它的步长还不一定能取到当前最优步长，因此，仅用柯西法进行计算，效果会很差。

本节我们只关注如何去求解子问题，所以我们不使用下标，而且本节我们记(4.5)的解为 $p^*(\Delta)$ ，用此强调 $\Delta$ 在子问题里的重要性。

## 狗腿法

当 $B$ 正定的时候，我们引入狗腿法。

而当 $B$ 正定时， $m$ 的无约束最优极小子是 $p^B = -B^{-1}g$ 。当这个点是(4.5)的可行解时，显然 $p^*(\Delta) = p^B$

然而当 $\Delta$ 很小的时候，我们可以舍弃(4.5)中的二次项，于是有 $p^*(\Delta) = -\Delta \frac{g}{\|g\|}$ ，而这种近似的最优解我们记为(4.15)

$$p^U = -\frac{g^T g}{g^T B g} g$$

而狗腿法便是用两条线段构造了一个寻找近似解的方法 $\tilde{p}(\tau)$ (4.16)

$$\tilde{p}(\tau) = \begin{cases} \tau p^U & 0 \leq \tau \leq 1 \\ p^U + (\tau - 1)(p^B - p^U) & 1 \leq \tau \leq 2 \end{cases}$$

而 $p^*(\Delta) = \min_{\tau} \tilde{p}(\tau)$ ,  $\forall \tilde{p}(\tau) \in \Delta$ ，而这个问题根据下述引理其实是容易求解的

### Lemma 4.2

假设 $B$ 正定，我们有

1.  $\|\tilde{p}(\tau)\|$ 随 $\tau$ 单增
2.  $m(\tilde{p}(\tau))$ 随 $\tau$ 单减

于是当 $\|p^B\| \leq \Delta$ 时， $p^*(\Delta) = p^B$ ，当 $\|p^B\| \geq \Delta$ 时， $\|p\| = \Delta$ 有唯一解，我们只需要解一个一次方程或二次方程：

$$\|p^U + (\tau - 1)(p^B - p^U)\|^2 = \Delta^2$$

虽然我们可以用修正的Hessian矩阵去处理非正定的情况，但是这样的处理会失去直观性，因为 $p^B$ 此时已经不是原子问题的无约束最优解了。所以这样处理没有意义。

## 二维子空间极小化

当 $B$ 正定时，狗腿法可以稍加修改，让 $p$ 在 $p^U$ 和 $p^B$ 张成的子空间中搜索，于是子问题(4.5)变成(4.17)

$$\min_p m(p) = f + g^T p + \frac{1}{2} p^T B p \quad s.t. \|p\| \leq \Delta, p \in \text{span}[g, B^{-1}g]$$

首先，显然柯西点是这个方法的可行解，因此这个方法具有全局收敛性；显然，狗腿法的解也是这个问题的可行解；同时，这个子问题其实是容易求解的(经过一些代数运算，我们可以把它转化为四次方程求根的问题)。

这个方法也可以经过一些调整去处理 $B$ 非正定的情况，而且这种做法是直观的，实用的且有理论保障的。这里我们只介绍一些核心思想：

当 $B$ 有负特征值，我们将(4.17)中的子空间修正为

$$\text{span}[g, (B + \alpha I)^{-1}g], \quad \text{for some } \alpha \in (-\lambda_1, -2\lambda_2]$$

其中 $\lambda_1$ 是 $B$ 的最小负特征值。

当 $(B + \alpha I)^{-1}g \leq \Delta$ , (4.19)

$$p = -(B + \alpha I)^{-1}g + v$$

其中 $v$ 满足 $v^T(B + \alpha I)^{-1}g \leq 0$ ，它保证了 $\|p\| \geq \|(B + \alpha I)^{-1}g\|$

当 $B$ 有零特征值但是没有负特征值时，我们取 $p = p^C$

大部分运算代价产生在 $B$ 和 $B + \alpha I$ 的矩阵分解上，相对来说 $\alpha$ 的估计和(4.17)的计算是比较简单的。

## 全局收敛性

### 柯西点收益

为了分析全局收敛性，我们首先估计柯西点的收益，然后再利用这个估计去证明算法4.1产生的 $\{g_k\}$ 在0处有一个聚点，而且当且仅当 $\eta > 0$ 时收敛。

我们第一个核心结果是证明：狗腿法和二维子空间法还有7.2里的方法关于(4.3)的估计满足如下不等式(4.20):

$$m_k(0) - m_k(p_k) \geq c_1 \|g_k\| \min \left( \Delta_k, \frac{\|g_k\|}{\|B_k\|} \right)$$

其中 $c_1 \in (0, 1]$ ，这个估计在后续分析中很有用。注意到，如果RHS中的最小值是 $\Delta_k$ 时，这个估计和Wolfe条件的第一个约束异曲同工：都是希望收益应该是梯度和步长倍数。

首先我们将证明柯西点是满足(4.20)的，其中 $c_1 = \frac{1}{2}$ 。

### 引理4.3

柯西点满足(4.20)，且 $c_1 = \frac{1}{2}$ ，即(4.21)

$$m_k(0) - m_k(p_k^C) \geq \frac{1}{2} \|g_k\| \min \left( \Delta_k, \frac{\|g_k\|}{\|B_k\|} \right)$$

### 定理4.4

$\|p_k\| \leq \Delta_k$  且  $m_k(0) - m_k(p_k) \geq c_2(m_k(0) - m_k(p_k^C))$ , 则  $p_k$  满足(4.20), 且  $c_1 = c_2/2$ . 特别的, 当  $p_k$  是(4.3)的精确解, 它满足(4.20), 且  $c_1 = \frac{1}{2}$ .

## 平稳点的收敛性

信赖域方法的全局收敛性取决于  $\eta$  是否等于0

1. 当  $\eta = 0$ , 我们只能证明  $\{g_k\}$  有下极限0
2. 当  $\eta > 0$ , 即我们要求每次迭代  $f$  必须要下降, 那么我们可以得到  $\{g_k\} \rightarrow 0$

证明过程中, 我们做如下假设:

1.  $\|B_k\|$  一致有界
2.  $f$  在水平集(4.24)  $S \stackrel{\text{def}}{=} \{x | f(x) \leq f(x_0)\}$  上下有界

为了叙述方便, 我们定义如下开邻域:

$$S(R_0) \stackrel{\text{def}}{=} \{x | \|x - y\| < R_0 \text{ for some } y \in S\}$$

另外, 为了得到更广泛的结果, 我们将允许(4.3)的近似解超出信赖域范围(4.25)

$$\|p_k\| \leq \gamma \Delta_k, \quad \text{for some constant } \gamma \geq 1$$

首先说明  $\eta = 0$  的情况

### 定理4.5

取  $\eta = 0$ , 假设  $\|B_k\| \leq \beta$ ,  $f$  在水平集  $S$  上下有界且在领域  $S(R_0)$  上 Lipschitz 连续可微, 且(4.3)的估计解  $p_k$  满足(4.20)和(4.25), 则(4.26)成立

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0$$

思路:

基于  $f$  下有界和信赖域方法的单减性质构造矛盾

$\Delta_k$  足够小时,  $\rho_k$  比较大, 所以信赖域大小不会变化, 因此我们能对这样的  $\Delta_k$  的上界做一个估计  $\bar{\Delta}$ ; 当信赖域仅会在大于它的时候缩小, 因此

$$\Delta_k \geq \min\{\bar{\Delta}/4, \Delta_K\}, \forall k \geq K$$

结合(4.20)考虑  $\rho_k$  比较大的子列, 得到  $f_k$  的估计, 与其下有界形成矛盾。

### 定理4.6

取  $\eta \in (0, \frac{1}{4})$ , 假设  $\|B_k\| \leq \beta$ ,  $f$  在水平集  $S$  上下有界且在领域  $S(R_0)$  上 Lipschitz 连续可微, 且(4.3)的估计解  $p_k$  满足(4.20)和(4.25), 则有(4.33)

$$\lim_{k \rightarrow \infty} \|g_k\| = 0$$

思路:

我们先证明, 当  $g_k \neq 0$  时, 信赖域方法总能让  $x$  离开  $x_k$  的小邻域  
后续证明则是基于  $\{f_k\}$  的单调下降性质构造矛盾

## 子问题的迭代解法

本节我们将应用子问题的特征(4.6), 结合牛顿法去求解满足(4.5)的  $\lambda$ , 同时还会去证明定理4.1。

4.1中的方法其实并没有很好地求解(4.5), 虽然它们用了一些Hessian矩阵的信息, 也有收敛性, 计算代价也合理。

当问题的规模不是很大的时候, 我们应该是可以通过进一步分析模型得到子问题更好的估计。本节我们将通过对  $B$  做一些(通常3次)矩阵分解得到一个好的估计, 并跟前面只用一次分解的狗腿法和二维子空间法作对比。

这个做法基于定理4.1的结果, 引入了一个精彩的一维牛顿法。本质上, 它是想找到(4.5)的解所满足的(4.6)中的  $\lambda$ 。

定理4.1说明(4.7)的解有如下两种情况

1.  $\lambda = 0$  (4.8a) (4.8c) 且  $\|p\| \leq \Delta$
2.  $\lambda$  足够大使得  $B + \lambda I$  正定且(4.37)  $\|p(\lambda)\| = \Delta$ , 其中  $p(\lambda) = -(B + \lambda I)^{-1}g$

注意到(4.37)其实是一个关于  $\lambda$  的一维求根问题。

为了证明  $\lambda$  的存在性, 我们对  $B$  做特征值分解, 并基于特征值分解去研究  $\|p(\lambda)\|$  的性质。由于  $B$  正定, 所以我们有  $B = Q\Lambda Q^T$ , 其中  $Q$  为正交阵,  $\Lambda$  为对角阵, 且

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

其中  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  是  $B$  的特征值。显然  $B + \lambda I = Q(\Lambda + \lambda I)Q^T$ , 对于  $\lambda \neq \lambda_j$ , 我们有(4.38)

$$p(\lambda) = -Q(\Lambda + \lambda I)^{-1}Q^Tg = -\sum_{j=1}^n \frac{q_j^T g}{\lambda_j + \lambda} q_j$$

其中  $q_j$  是  $Q$  的第  $j$  列, 因此, 由  $q_1, q_2, \dots, q_n$  的正交性, 我们有(4.39)

$$\|p(\lambda)\|^2 = \sum_{j=1}^n \frac{(q_j^T g)^2}{(\lambda_j + \lambda)^2}$$

于是, 当  $\lambda > -\lambda_1$  时, 我们有  $\lambda_j + \lambda > 0$ , 因此当  $\lambda \in (-\lambda_1, \infty)$ ,  $\|p(\lambda)\|$  是  $\lambda$  的连续函数。实际上, 我们有(4.40)



$$\lim_{\lambda \rightarrow \infty} \|p(\lambda)\| = 0$$

同时(4.41)

$$\lim_{\lambda \rightarrow -\lambda_1} \|p(\lambda)\| = \infty$$

特别的, 当 $q_1^T g \neq 0$ 时, 方程 $\|p(\lambda)\| = \Delta$ 在 $(-\lambda_1, \infty)$ 内存在唯一解 $\lambda^*$

下面我们开始分析, 当 $q_1^T g \neq 0$ 时, 如何找到 $\lambda^*$

首先, 如果 $B$ 正定, 且 $\|B^{-1}g\| \leq \Delta$ , 则 $\lambda = 0$ 满足(4.8), 故 $\lambda^* = 0$

否则, 我们将用牛顿求根公式去寻找 $\lambda > -\lambda_1$  s.t. (4.42)

$$\phi_1(\lambda) = \|p(\lambda)\| - \Delta = 0$$

然而这个方法并不是很好。当 $\lambda$ 接近 $-\lambda_1$ 时, 我们可以用有理函数对 $\phi_1$ 做一个估计, 即

$$\phi_1(\lambda) \approx \frac{C_1}{\lambda + \lambda_1} + C_2$$

其中 $C_1 > 0$ ,  $C_2$ 均为常数。可以看到这个估计是非线性的, 因此原函数也是, 从而, 牛顿法将会变得不是很可靠。然而如果我们对(4.42)做一些改变, 我们可以改善这一结果。下面定义

$$\phi_2(\lambda) = \frac{1}{\Delta} - \frac{1}{\|p(\lambda)\|}$$

可以看到, 当 $\lambda$ 接近 $-\lambda_1$ 时, 我们有

$$\phi_2(\lambda) \approx \frac{1}{\Delta} - \frac{\lambda + \lambda_1}{C_3}$$

其中 $C_3 > 0$ 。即 $\phi_2$ 在根附近是接近线性的, 于是只要保证 $\lambda > -\lambda_1$ , 牛顿求根公式将有很好的效果。其具体形式如下:

$$\lambda^{(l+1)} = \lambda^{(l)} - \frac{\phi_2(\lambda^{(l)})}{\phi_2'(\lambda^{(l)})}$$

经过一些基本操作, 这个迭代公式具体实现方式如下:

#### Algorithm 4.3 (Trust Region Subproblem)

```

Given lambda(0), Delta>0;
for l = 0,1,2,...
    Factor B+lambda(l)I = R'R;
    Solve R'Rp(l) = -g, R'q(l)=p(l);
    Set (4.44)
        lambda(l+1) = lambda(l)+(norm(p(l))/norm(q(l)))^2*((norm(p(l))-Delta)/Delta));
end

```

对于这个算法，实际使用的时候一定要再添加一些保护措施；比如 $\lambda^{(l)} < -\lambda_1$ 时，分解不存在。这个算法的改良版本大部分情况下是可以收敛到(4.37)的解的。

这个算法每次迭代主要代价在于矩阵分解。实际应用时，我们不总要求算法迭代到收敛解，而是满足于两次或三次迭代后得到的近似解。

## 困难的情况

前述分析我们要求 $q_1^T g \neq 0$ ，其实即使很多负特征值是重特征值即 $0 > \lambda_1 = \lambda_2 = \dots$ 时，这个方法也是有效的，只需 $Q_1^T g \neq 0$ ，其中 $Q_1$ 是 $\lambda_1$ 的特征向量张成的子空间。

然而，当这个条件不满足时，情况就会变得复杂，因为此时(4.41)对于 $\lambda_j = \lambda_1$ 不成立，方程(4.37)可能在 $(-\lambda_1, \infty)$ 没有解，但是定理4.1告诉我们， $\lambda$ 应该取在区间 $[-\lambda_1, \infty)$ 内，因此 $\lambda = -\lambda_1$ ，为了找到 $p$ ，光去掉(4.38)中 $\lambda_j = \lambda_1$ 的项是不够的，我们注意到由于 $(B - \lambda_1 I)$ 是奇异阵，我们能找到 $\|z\| = 1$  s.t.  $(B - \lambda_1 I)z = 0$ 。其实， $z$ 是 $B$ 关于 $\lambda_1$ 的特征向量，由 $Q$ 的正交性，我们有 $q_j^T z = 0, \lambda_j \neq \lambda_1$ ，于是我们设(4.45)

$$p = \sum_{j: \lambda_j \neq \lambda_1} \frac{q_j^T g}{\lambda_j + \lambda} q_j + \tau z$$

对于一些常量 $\tau$ ，我们有

$$\|p\|^2 = \sum_{j: \lambda_j \neq \lambda_1} \frac{(q_j^T g)^2}{(\lambda_j + \lambda)^2} + \tau^2$$

于是，只要选取合适的 $\tau$ ，我们总能保证 $\|p\| = \Delta$ ，而且此时取 $\lambda = -\lambda_1$ ，(4.8)成立

## 定理4.1的证明

下面我们证明定理4.1，证明主要基于引理4.7。引理4.7求解二次优化的极小点，当Hessian矩阵半正定时尤为有趣。

### 引理4.7

(4.46)

$$m(p) = g^T p + \frac{1}{2} p^T B p$$

其中 $B$ 为对称矩阵，我们有下述结论：

1. 当且仅当 $B$ 半正定且 $g \in R(B)$ 时， $m$ 有最小值。进一步的，当 $B$ 半正定时，任何 $p$  s.t  $Bp = -g$ 都是 $m$ 的极小子
2. 当且仅当 $B$ 正定时， $m$ 有唯一极小子

核心操作：

$$m(p + w) \geq m(p)$$

下面我们证明定理4.1，即在引理4.7中考虑信赖域约束 $\|p\| \leq \Delta$ 的作用

充分性

我们可以根据引理引入以 $p^*$ 为极小子的函数

$$\hat{m}(p) = g^T p + \frac{1}{2} p^T (B + \lambda I) p = m(p) + \frac{\lambda}{2} p^T p$$

然后根据其极值和(4.8)的特征式得到结果

必要性

分 $\|p^*\| < \Delta$ (无约束)和 $\|p^*\| = \Delta$ 的情况讨论，前者结论显然，而后者则需引入一个拉格朗日函数

$$\mathcal{L}(p, \lambda) = m(p) + \frac{\lambda}{2} (p^T p - \Delta^2)$$

得到(4.8a)，再根据一些代数运算得到(4.8c)，最后通过反证法说明存在 $\lambda > 0$

## 基于近似精确解的算法的收敛性

如同我们之前在分析算法4.3时提到的一样，我们在求解子问题(4.5)时，并不需要太多的精度。同时我们会通过增加一些保护措施使得牛顿求根过程的结果能够满足定理4.5 4.6的要求，特别的，我们会要求(4.52)

$$\begin{aligned} m(0) - m(p) &\geq c_1(m(0) - m(p^*)) \\ \|p\| &\leq \gamma \Delta \end{aligned}$$

其中 $p^*$ 是(4.3)的精确解， $c_1 \in (0, 1], \gamma > 0$

(4.52)与之前的标准(4.20)主要的不同是在于(4.52)对二次信息的利用率更高，典型的，对于鞍点 $x_k$ ，(4.20)的右端是0，即我们会在鞍点停止，而(4.52)右端仍是正的，即能让算法离开鞍点。

近似精确解对于二次项的关注仅当该项反映了函数的真实表现时有效——实际上，也仅有信赖域牛顿方法满足该要求。

在分析全局收敛性质的时候，精确的Hessian矩阵信息能够保证极限不仅仅是稳定点。下述结论说明极限点还满足二阶必要条件。

### 定理4.8

假设定理4.6的条件满足且 $f$ 在水平集上二阶连续可微。假设 $B_k = \nabla^2 f(x_k)$ ，则(4.3)的近似解 $p_k$ 满足(4.52)。且 $\lim_{k \rightarrow \infty} \|g_k\| = 0$ 。

如果(4.24)的水平集是紧集，则算法将收敛至满足二阶必要条件的 $x_k$ 处，或 $\{x_k\}$ 有极限点 $x^*$ 满足二阶必要条件。

## 信赖域算法的局部收敛性

使用Hessian矩阵的信赖域方法的全局收敛性已经被证明，下面我们分析它的局部收敛性。其核心在于去证明信赖域边界最终不影响秋季，即我们希望在接近真解时，子问题(近似)解将存在于信赖域内部并且近似于真实牛顿法。我们把满足后续说法的迭代步称为渐进相似于牛顿步。

首先我们证明对于一般的信赖域算法4.1，如果在牛顿步在信赖域内时它的迭代步渐进相似于牛顿步，那么其信赖域约束最终将不影响算法，并且该算法能达到超线性收敛。

这个结果假设当 $x_k$ 靠近满足二阶充分条件的真解时，将在子问题中取 $B_k = \nabla^2 f(x_k)$ 。同时，它假设算法使用的近似解 $p_k$ 与 $p_k^C$ 对于 $m_k$ 的下降效果相似。

## 定理4.9

假设 $f$ 在满足二阶充分条件的 $x^*$ 附近二次Lipschitz连续可微。假设 $\{x_k\}$ 收敛到 $x^*$ ，且当 $k$ 充分大时，子问题(4.3)设 $B_k = \nabla^2 f(x_k)$ 选择 $p_k$ ，且 $p_k$ 满足(4.20)且在 $\|p_k^N\| \leq \frac{1}{2}\Delta_k$ 时，渐进相似与牛顿步 $p_k^N$ ，即(4.53)

$$\|p_k - p_k^N\| = o(\|p_k^N\|)$$

则信赖域界 $\Delta_k$ 会对于大的 $k$ 失去作用，且 $\{x_k\}$ 超线性收敛至 $x^*$ 。

证明思路：

再次利用 $\|\rho_k - 1\|$ 的估计和全局收敛的结果，得到 $\Delta_k$ 大于某个正数

## 其他改进方法

### Scaling

在 $f$ 失衡的时候，用圆形区域显然不是一个很好的选择，同时，信赖域方法在 $f$ 变化缓慢的方向近似效果可能更好，于是，椭圆形信赖域是一种值得考虑的方法，即(4.56)

$$\|Dp\| \leq \Delta$$

其中， $D$ 是对角元素为正数的对角阵，它导出下述问题：

$$\min_{p \in R^n} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p \quad \text{s.t. } \|Dp\| \leq \Delta_k$$

当 $f(x)$ 对于分量 $x_i$ 敏感时，我们令 $d_{ii}$ 取较大的数，反之取较小的数。

我们可以从 $f$ 的二阶导数信息中得到构造 $D$ 的一些信息， $D$ 也可以在迭代过程中变化，大部分情况下 $d_{ii}$ 的取值会有一个范围限制，当然我们也没必要 $D$ 很精确地反映出尺度信息

### Algorithm 4.4 Generalized Cauchy Point Calculation

略

有一种简单的实现操作是对原子问题(4.57)进行仿射变换：

$$\tilde{p} \stackrel{\text{def}}{=} Dp$$

于是，得到

$$\min_{\tilde{p} \in R^n} \tilde{m}_k(\tilde{p}) = f_k + g_k^T D^{-1} \tilde{p} + \frac{1}{2} \tilde{p}^T D^{-1} B_k D^{-1} \tilde{p} \quad \text{s.t. } \|\tilde{p}\| \leq \Delta_k$$

## 其他形式的信赖域方法

信赖域方法也可以用其他范数进行定义，如

$$\|p_k\|_1 \leq \Delta \quad \|p_k\|_\infty \leq \Delta$$

或仿射版本

$$\|Dp_k\|_1 \leq \Delta \quad \|Dp_k\|_\infty \leq \Delta$$

实际上，对于大规模问题，**Hessian**矩阵不好求，此时使用无穷范数会改善问题的性质，它的求解可能会比原子问题更便捷。然而这方面的研究还比较少。