

信赖域算法

文再文

北京大学北京国际数学研究中心

教材《最优化:建模、算法与理论》配套电子教案

<http://bicmr.pku.edu.cn/~wenzw/optbook.html>

致谢:本教案由**邓展望**协助准备

- 1 信赖域算法框架
- 2 信赖域子问题
- 3 柯西点
- 4 收敛性
- 5 应用举例

信赖域算法框架

- 1 在当前迭代点 x^k 建立局部模型

$$d^k = \arg \min_d (g^k)^\top d + d^\top B d, \text{ s.t. } \|d\|_2 \leq \Delta_k$$

- 2 求出局部模型的最优解
- 3 更新模型信赖域的半径：
 - 模型足够好 \Rightarrow 增大半径
 - 模型比较差 \Rightarrow 缩小半径
 - 否则半径不变
- 4 对模型进行评价：
 - 好 \Rightarrow 子问题的解即下一个迭代点
 - 差 \Rightarrow 迭代点不改变

信赖域子问题

- 根据带拉格朗日余项的泰勒展开

$$f(x^k + d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla^2 f(x^k + td) d$$

其中 $t \in (0, 1)$ 为和 d 有关的正数。

- 和牛顿法相同，利用 $f(x)$ 的二阶近似来刻画 $f(x)$ 在点 x^k 处的性质：

$$m_k(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T B^k d$$

其中 B^k 是对称矩阵，并且是海瑟矩阵的近似矩阵。

- 由于泰勒展开的局部性，需对上述模型添加信赖域约束：

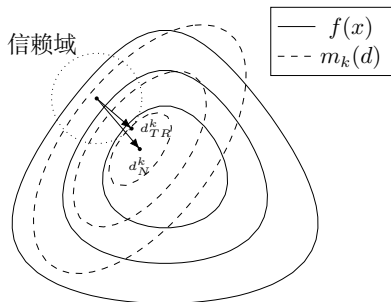
$$\Omega_k = \{x^k + d \mid \|d\| \leq \Delta_k\},$$

其中 $\Delta_k > 0$ 是信赖域半径。

信赖域子问题

- 因此信赖域算法每一步都需要求解如下子问题:

$$\min_{d \in \mathbb{R}^n} m_k(d), \quad \text{s.t.} \quad \|d\| \leq \Delta_k. \quad (1)$$



- 图中实线表示 $f(x)$ 的等高线，虚线表示 $m_k(d)$ 的等高线， d_N^k 表示解无约束问题得到的下降方向， d_{TR}^k 表示解信赖域子问题得到的下降方向。

模型近似程度好坏的的衡量

- 引入如下定义来衡量 $m_k(d)$ 近似程度的好坏：

$$\rho_k = \frac{f(x^k) - f(x^k + d^k)}{m_k(0) - m_k(d^k)} \quad (2)$$

其中 d^k 为解信赖域子问题得到的迭代方向.根据 ρ_k 的定义可知,其为函数值实际下降量与预估下降量(即二阶近似模型下降量)的比值.

- 如果 ρ_k 接近1,说明 $m_k(d)$ 来近似 $f(x)$ 是比较成功的,则应该扩大 Δ_k ;如果 ρ_k 非常小甚至为负,就说明我们过分地相信了二阶模型 $m_k(d)$,此时应该缩小 Δ_k .使用这个机制可以动态调节 Δ_k ,让二阶模型 $m_k(d)$ 的定义域处于一个合适的范围.

Algorithm 1 信赖域算法

- 1: 给定最大半径 Δ_{\max} , 初始半径 Δ_0 , 初始点 x^0 , $k \leftarrow 0$.
- 2: 给定参数 $0 \leq \eta < \bar{\rho}_1 < \bar{\rho}_2 < 1$, $\gamma_1 < 1 < \gamma_2$.
- 3: **while** 未达到收敛准则 **do**
- 4: 计算子问题(2)得到迭代方向 d^k .
- 5: 根据(3)式计算下降率 ρ_k .
- 6: 更新信赖域半径:

$$\Delta_{k+1} = \begin{cases} \gamma_1 \Delta_k, & \rho_k < \bar{\rho}_1, \\ \min\{\gamma_2 \Delta_k, \Delta_{\max}\}, & \rho_k > \bar{\rho}_2 \text{ 以及 } \|d^k\| = \Delta_k, \\ \Delta_k, & \text{其他.} \end{cases}$$

- 7: 更新自变量:

$$x^{k+1} = \begin{cases} x^k + d^k, & \rho_k > \eta, \\ x^k, & \text{其他.} \end{cases} \quad /* \text{ 只有下降比例足够大才更新} */$$

- 8: $k \leftarrow k + 1$.
- 9: **end while**

- 1 信赖域算法框架
- 2 信赖域子问题**
- 3 柯西点
- 4 收敛性
- 5 应用举例

信赖域子问题最优性条件

定理 (信赖域子问题最优性条件)

d^* 是信赖域子问题

$$\min \quad m(d) = f + g^T d + \frac{1}{2} d^T B d, \quad \text{s.t.} \quad \|d\| \leq \Delta \quad (3)$$

的全局极小解当且仅当 d^* 是可行的且存在 $\lambda \geq 0$ 使得

$$(B + \lambda I) d^* = -g, \quad (4a)$$

$$\lambda(\Delta - \|d^*\|) = 0, \quad (4b)$$

$$B + \lambda I \succeq 0. \quad (4c)$$

信赖域子问题最优性条件

必要性：问题(3)的拉格朗日函数为

$$L(d, \lambda) = f + g^T d + \frac{1}{2} d^T B d - \frac{\lambda}{2} (\Delta^2 - \|d\|^2),$$

其中乘子 $\lambda \geq 0$.

- 由KKT条件， d^* 是可行解，且 $\nabla_d L(d^*, \lambda) = (B + \lambda I)d^* + g = 0$. 此外由互补条件 $\frac{\lambda}{2} (\Delta^2 - \|d^*\|^2) = 0$, 整理后就是(4a)式和(4b)式.
- 为了证明(4c)式，我们任取 d 满足 $\|d\| = \Delta$ ，根据最优性，有

$$m(d) \geq m(d^*) = m(d^*) + \frac{\lambda}{2} (\|d^*\|^2 - \|d\|^2).$$

利用(4a)式消去 g ，代入上式整理有 $(d - d^*)^T (B + \lambda I) (d - d^*) \geq 0$ ，由 d 的任意性可知 $B + \lambda I$ 半正定.

信赖域子问题最优性条件

再证明充分性. 定义辅助函数

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

由条件(4c)可知 $\hat{m}(d)$ 关于 d 是凸函数. 根据条件(4a), d^* 满足凸函数一阶最优性条件, 可推出 d^* 是 $\hat{m}(d)$ 的全局极小值点, 进而对任意可行解 d , 我们有

$$m(d) \geq m(d^*) + \frac{\lambda}{2} (\|d^*\|^2 - \|d\|^2).$$

由互补条件(4b)可知 $\lambda(\Delta^2 - \|d^*\|^2) = 0$, 代入上式消去 $\|d^*\|^2$ 得

$$m(d) \geq m(d^*) + \frac{\lambda}{2} (\Delta^2 - \|d\|^2) \geq m(d^*).$$

求解信赖域子问题迭代法

- 上述定理提供了一个寻找 d 的方法:
 - $\lambda = 0$, 测试是否 $B \succeq 0$, $d = -B^{-1}g$ 且 $\|d\| \leq \Delta$ 满足.
 - 选择充分大的 λ 使得 $B + \lambda I \succeq 0$ 且 $\|d(\lambda)\| = \Delta$, 其中

$$d(\lambda) = -(B + \lambda I)^{-1}g. \quad (5)$$

关键: 求解关于 λ 的方程 $\|d(\lambda)\| = \Delta$ 或者 $1/\|d(\lambda)\| = 1/\Delta$.

- 设 B 有特征值分解 $B = Q\Lambda Q^T$, 其中 $Q = [q_1, q_2, \dots, q_n]$ 是正交矩阵, $\Lambda = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 是对角矩阵, $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ 是 B 的特征值.

$$d(\lambda) = -Q(\Lambda + \lambda I)^{-1}Q^Tg = -\sum_{i=1}^n \frac{q_i^T g}{\lambda_i + \lambda} q_i. \quad (6)$$

这正是 $d(\lambda)$ 的正交分解, 由正交性可容易求出

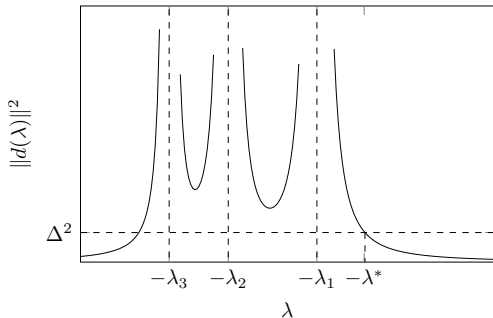
$$\|d(\lambda)\|^2 = \sum_{i=1}^n \frac{(q_i^T g)^2}{(\lambda_i + \lambda)^2}. \quad (7)$$

求解信赖域子问题迭代法

- 由(7)式, 当 $\lambda > -\lambda_1$ 且 $q_1^T g \neq 0$ 时, $\|d(\lambda)\|^2$ 是关于 λ 的严格减函数, 且有

$$\lim_{\lambda \rightarrow \infty} \|d(\lambda)\| = 0, \quad \lim_{\lambda \rightarrow -\lambda_1+} \|d(\lambda)\| = +\infty.$$

- 由连续函数介值定理, $\|d(\lambda)\| = \Delta$ 的解必存在且唯一.
- 所以寻找 λ^* 已经转化为一个一元方程求根问题, 可使用牛顿法求解.



求解信赖域子问题牛顿法

求解 $\phi(\lambda) = 1/\Delta - 1/\|d(\lambda)\| = 0$ 且 $B + \lambda I \succeq 0$.

- $\phi'(\lambda) = \|d(\lambda)\|^{-3} d(\lambda)^\top d'(\lambda)$. 对(4a) 关于 λ 求导数:

$$d(\lambda) + (B + \lambda I)d'(\lambda) = 0 \implies d'(\lambda) = -(B + \lambda I)^{-1}d(\lambda).$$

- 假设 $B + \lambda I = R^\top R$. 因此

$$\begin{aligned} d(\lambda)^\top d'(\lambda) &= -d(\lambda)^\top (B + \lambda I)^{-1}d(\lambda) = -d(\lambda)^\top R^{-1}R^{-\top}d(\lambda) \\ &= -\|R^{-\top}d(\lambda)\|_2^2. \end{aligned}$$

- 牛顿法:

$$\begin{aligned} \lambda^{(l+1)} &= \lambda^{(l)} - \frac{\phi(\lambda^{(l)})}{\phi'(\lambda^{(l)})} \\ &= \lambda^{(l)} + \frac{\|d(\lambda)\|_2^2}{\|R^{-\top}d(\lambda)\|_2^2} \frac{\|d(\lambda)\|_2 - \Delta}{\Delta} \end{aligned}$$

困难情形

- 上面的分析中假定了 $q_1^T g \neq 0$ ，在实际中这个条件未必满足。当 $q_1^T g = 0$ 时，(7) 式将没有和 λ_1 相关的项。此时未必存在 $\lambda^* > -\lambda_1$ 使得 $\|d(\lambda^*)\| = \Delta$ 成立。记

$$M = \lim_{\lambda \rightarrow -\lambda_1^+} \|d(\lambda)\|$$

- 当 $M \geq \Delta$ 时，仍然可以根据介值定理得出 $\lambda^* (> -\lambda_1)$ 的存在性；
- 当 $M < \Delta$ 时，无法利用前面的分析求出 λ^* 和 d^* ，此时信赖域子问题变得比较复杂。

实际上， $q_1^T g = 0$ 且 $M < \Delta$ 的情形被称为“困难情形(hard case)”。此情形发生时，区间 $(-\lambda_1, +\infty)$ 中的点无法使得(4b) 成立，而定理?? 的结果说明 $\lambda^* \in [-\lambda_1, +\infty)$ ，因此必有 $\lambda^* = -\lambda_1$ 。

困难情形

- 为了求出 d^* ，可以利用（奇异）线性方程组(4a)解的结构，其通解可以写为

$$d(\alpha) = - \sum_{i=2}^n \frac{q_i^T g}{\lambda_i - \lambda_1} q_i + \alpha q_1, \quad \alpha \in \mathbb{R}.$$

- 由正交性，

$$\|d(\alpha)\|^2 = \alpha^2 + \sum_{i=2}^n \frac{(q_i^T g)^2}{(\lambda_i - \lambda_1)^2}.$$

注意在困难情形中有 $M = \sqrt{\sum_{i=2}^n \frac{(q_i^T g)^2}{(\lambda_i - \lambda_1)^2}} < \Delta$ ，因此必存在 α^* 使得 $\|d(\alpha^*)\| = \Delta$ 。这就求出了 d^* 的表达式。

截断共轭梯度法介绍

下面再介绍一种信赖域子问题的求解方法.

- 既然信赖域子问题的解不易求出, 则求出其近似解, Steihaug 在1983年对共轭梯度法进行了改造, 使其成为能求解子问题的算法. 此算法能够应用在大规模问题中, 是一种非常有效的信赖域子问题的求解方法.
- 由于子问题和一般的二次极小化问题相差一个约束, 如果先不考虑其中的约束 $\|d\| \leq \Delta$ 而直接使用共轭梯度法求解, 在迭代过程中找到合适的迭代点作为信赖域子问题的近似解, 检测到负曲率或者到达信赖域边界 $\|d\| = \Delta$ 即终止. 这就是截断共轭梯度法(见书P260)的基本思想.

共轭梯度法介绍

- 标准共轭梯度法求解二次极小化问题

$$\min_s q(s) \stackrel{\text{def}}{=} g^T s + \frac{1}{2} s^T B s,$$

- 给定初值 $s^0 = 0, r^0 = g, p^0 = -g$, 迭代过程为

$$\begin{aligned}\alpha_{k+1} &= \frac{\|r^k\|^2}{(p^k)^T B p^k}, \\ s^{k+1} &= s^k + \alpha_k p^k, \\ r^{k+1} &= r^k + \alpha_k B p^k, \\ \beta_k &= \frac{\|r^{k+1}\|^2}{\|r^k\|^2}, \\ p^{k+1} &= -r^{k+1} + \beta_k p^k,\end{aligned}$$

其中迭代序列 $\{s^k\}$ 最终的输出即为二次极小化问题的解，算法的终止准则是判断 $\|r^k\|$ 是否足够小。

截断共轭梯度法

- 截断共轭梯度法则是标准共轭梯度法增加了两条终止准则，并对最后一步的迭代点 s^k 进行修正来得到信赖域子问题的解。
- 矩阵 B 不一定正定，在迭代过程中可能会产生如下三种情况：
 - ① $(p^k)^T B p^k \leq 0$ ，即 B 不是正定矩阵。遇到这种情况立即终止算法。但根据这个条件也找到了一个负曲率方向，此时只需要沿着这个方向走到信赖域边界即可。
 - ② $(p^k)^T B p^k > 0$ 但 $\|s^{k+1}\| \geq \Delta$ ，这表示若继续进行共轭梯度法迭代，则点 s^{k+1} 将处于信赖域之外或边界上，此时必须马上停止迭代，并在 s^k 和 s^{k+1} 之间找一个近似解。
 - ③ $(p^k)^T B p^k > 0$ 且 $\|r^{k+1}\|$ 充分小，这表示若共轭梯度法成功收敛到信赖域内。子问题(1)和不带约束的二次极小化问题是等价的。
- 从上述终止条件来看截断共轭梯度法仅仅产生了共轭梯度法的部分迭代点，这也是该方法名字的由来。

截断共轭梯度法

$$\min_s \quad q(s) \stackrel{\text{def}}{=} g^T s + \frac{1}{2} s^T B s, \quad \text{s.t. } \|s\| \leq \Delta.$$

Algorithm 2 截断共轭梯度法1 (Steihaug-CG)

- 1: 给定精度 $\varepsilon > 0$, 初始化 $s^0 = 0, r^0 = g, p^0 = -g, k \leftarrow 0$.
- 2: **if** $\|p^0\| \leq \varepsilon$ **then**
- 3: 算法停止, 输出 $s = 0$.
- 4: **end if**
- 5: **LOOP**
- 6: **if** $(p^k)^T B p^k \leq 0$ **then**
- 7: 计算 $\tau > 0$ 使得 $\|s^k + \tau p^k\| = \Delta$.
- 8: 算法停止, 输出 $s = s^k + \tau p^k$.
- 9: **end if**
- 10: 计算 $\alpha_k = \frac{\|r^k\|^2}{(p^k)^T B p^k}$, 更新 $s^{k+1} = s^k + \alpha_k p^k$.
- 11: 接下页

截断共轭梯度法

Algorithm 3 截断共轭梯度法2 (Steihaug-CG)

- 1: **if** $\|s^{k+1}\| \geq \Delta$ **then**
 - 2: 计算 $\tau > 0$ 使得 $\|s^k + \tau p^k\| = \Delta$.
 - 3: 算法停止, 输出 $s = s^k + \tau p^k$.
 - 4: **end if**
 - 5: 计算 $r^{k+1} = r^k + \alpha_k B p^k$.
 - 6: **if** $\|r^{k+1}\| < \varepsilon \|r^0\|$ **then**
 - 7: 算法停止, 输出 $s = s^{k+1}$.
 - 8: **end if**
 - 9: 计算 $\beta_k = \frac{\|r^{k+1}\|^2}{\|r^k\|^2}$, 更新 $p^{k+1} = -r^{k+1} + \beta_k p^k$.
 - 10: $k \leftarrow k + 1$.
 - 11: **ENDLOOP**
-

截断共轭梯度法

截断共轭梯度法的迭代序列 $\{s^k\}$ 有非常好的性质，实际上我们可以证明如下定理：

定理

设 $q(s)$ 是任意外迭代步信赖域子问题的目标函数，令 $\{s^j\}$ 是由截断共轭梯度算法产生的迭代序列，则在算法终止前 $q(s^j)$ 是严格单调递减的，即

$$q(s^{j+1}) < q(s^j). \quad (8)$$

并且 $\|s^j\|$ 是严格单调递增的，即

$$0 = \|s^0\| < \|s^1\| < \dots < \|s^{j+1}\| < \dots \leq \Delta. \quad (9)$$

截断共轭梯度法

Proof.

设迭代在第 t 步终止. 根据算法2, 在终止前, 若 $(p^j)^T B p^j > 0, j < t$ 一直成立. 此时算法即共轭梯度法, 容易证明(8)式和(9)式.

又 $q(s)$ 在点 s^j 处的梯度为 r^j , 由共轭梯度法性质 $(r^j)^T p^i = 0, i < j$, 所以 $(r^j)^T p^j = (r^j)^T (-r^j + \beta_{j-1} p^{j-1}) = -\|r^j\|^2 < 0$, 即 p^j 是下降方向.

而 α_j 的选取为精确步长, 因此有 $q(s^{j+1}) < q(s^j)$. 此外由 s^j 的定义, $s^j = \sum_{i=0}^{j-1} \alpha_i p^i, \alpha_i > 0$. 再根据共轭梯度法的性质:

$$(p^j)^T s^j = \sum_{i=0}^{j-1} \alpha_i (p^j)^T p^i = \sum_{i=0}^{j-1} \alpha_i \frac{\|r^j\|^2}{\|r^i\|^2} \|p^i\|^2 > 0.$$

结合以上表达式可得

$$\|s^{j+1}\|^2 = \|s^j + \alpha_j p^j\|^2 = \|s^j\|^2 + 2\alpha_j (p^j)^T s^j + \alpha_j^2 \|p^j\|^2 > \|s^j\|^2. \quad \square$$

截断共轭梯度法

实际上，我们还可进一步说明截断共轭梯度算法的输出 s 满足如下关系：

$$q(s) \leq q(s^t), \quad \|s^t\| \leq \|s\|,$$

其中 t 为算法终止时的迭代数。这只需要分别讨论三种终止条件即可。

- 1 若 $(p^t)^T B p^t \leq 0$ ，则 p^t 是负曲率方向，沿着负曲率方向显然有 $q(s) \leq q(s^t)$ 。注意到此时 $\|s\| = \Delta$ ，因此有 $\|s^t\| \leq \|s\| = \Delta$ 。
- 2 若 $(p^t)^T B p^t > 0$ 但 $\|s^{t+1}\| \geq \Delta$ ，根据最速下降法的性质， $q(s^t + \alpha p^t)$ 关于 $\alpha \in (0, \alpha_t]$ 单调下降，根据 τ 的取法显然有 $q(s) \leq q(s^t)$ 。此时依然有 $\|s\| = \Delta$ ，因此 $\|s^t\| \leq \|s\| = \Delta$ 仍成立。
- 3 若 $(p^t)^T B p^t > 0$ 且 $\|r^{t+1}\| \leq \varepsilon \|r^0\|$ ，此时算法就是共轭梯度法，结论自然成立。

提纲

- 1 信赖域算法框架
- 2 信赖域子问题
- 3 柯西点**
- 4 收敛性
- 5 应用举例

柯西点定义

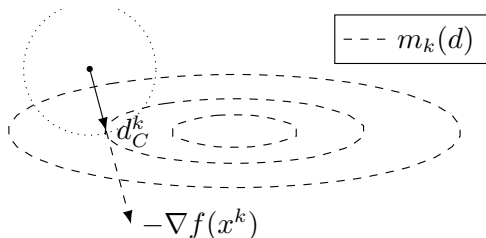
定义 (柯西点)

设 $m_k(d)$ 是 $f(x)$ 在点 $x = x^k$ 处的二阶近似， τ_k 为如下优化问题的解：

$$\begin{aligned} \min \quad & m_k(-\tau \nabla f(x^k)), \\ \text{s.t.} \quad & \|\tau \nabla f(x^k)\| \leq \Delta_k, \tau \geq 0. \end{aligned}$$

则称 $x_C^k \stackrel{\text{def}}{=} x^k + d_C^k$ 为柯西点，其中 $d_C^k = -\tau_k \nabla f(x^k)$ 。

柯西点实际上是在约束下对 $m_k(d)$ 进行了一次精确线搜索的梯度法



柯西点性质

给定 $m_k(d)$ ，柯西点可以显式计算出来。为了方便我们用 g^k 表示 $\nabla f(x^k)$ ，根据 τ_k 的定义，容易计算出其表达式为

$$\tau_k = \begin{cases} \frac{\Delta_k}{\|g^k\|}, & (g^k)^\top B^k g^k \leq 0, \\ \min \left\{ \frac{\|g^k\|^2}{(g^k)^\top B^k g^k}, \frac{\Delta_k}{\|g^k\|} \right\}, & \text{其他.} \end{cases}$$

引理 (柯西点的下降量)

设 d_C^k 为求解柯西点产生的下降方向，则

$$m_k(0) - m_k(d_C^k) \geq c_1 \|g^k\| \min \left\{ \Delta_k, \frac{\|g^k\|}{\|B^k\|_2} \right\}. \quad (10)$$

其中 $c_1 = \frac{1}{2}$

柯西点性质

分三种情况证明该结论.(方便起见证明中忽略上标 k)

- 首先考虑 $g^T B g \leq 0$ 时的情况

$$\begin{aligned}m(d_C) - m(0) &= m(-\Delta g / \|g\|) - f \\&= -\frac{\Delta}{\|g\|} \|g\|^2 + \frac{1}{2} \frac{\Delta^2}{\|g\|^2} g^T B g \\&\leq -\Delta \|g\| \\&\leq -\|g\| \min\left(\Delta, \frac{\|g\|}{\|B\|}\right)\end{aligned}$$

即(10)式成立.

柯西点性质

- 然后考虑 $g^T B g > 0$, $\frac{\|g\|^3}{\Delta g^T B g} \leq 1$ 时的情况

$$\begin{aligned} m(d_C) - m(0) &= -\frac{\|g\|^4}{g^T B g} + \frac{1}{2} g^T B g \frac{\|g\|^4}{(g^T B g)^2} \\ &= -\frac{1}{2} \frac{\|g\|^4}{g^T B g} \\ &\leq -\frac{1}{2} \frac{\|g\|^4}{\|B\| \|g\|^2} \\ &= -\frac{1}{2} \frac{\|g\|^2}{\|B\|} \\ &\leq -\frac{1}{2} \|g\| \min\left(\Delta, \frac{\|g\|}{\|B\|}\right) \end{aligned}$$

即(10)式成立.

柯西点性质

- 最后考虑 $\frac{\|g\|^3}{\Delta g^T B g} \geq 1$ 时的情况

$$\begin{aligned}m(d_C) - m(0) &= -\frac{\Delta}{\|g\|} \|g\|^2 + \frac{1}{2} \frac{\Delta^2}{\|g\|^2} g^T B g \\ &\leq -\Delta \|g\| + \frac{1}{2} \frac{\Delta^2}{\|g\|^2} \frac{\|g\|^3}{\Delta} \\ &= -\frac{1}{2} \Delta \|g\| \\ &\leq -\frac{1}{2} \|g\| \min \left(\Delta, \frac{\|g\|}{\|B\|} \right)\end{aligned}$$

即(10)式成立.

- 1 信赖域算法框架
- 2 信赖域子问题
- 3 柯西点
- 4 收敛性
- 5 应用举例

全局收敛性

回顾信赖域算法，我们引入了一个参数 η 来确定是否应该更新迭代点。这分为两种情况：当 $\eta = 0$ 时，只要原目标函数有下降量就接受信赖域迭代步的更新；当 $\eta > 0$ 时，只有当改善量 ρ_k 达到一定程度时再进行更新。在这两种情况下得到的收敛性结果是不同的，我们分别介绍这两种结果。在 $\eta = 0$ 的条件下有如下收敛性定理：

定理 (全局收敛性1)

设近似海瑟矩阵 B^k 有界，即 $\|B^k\|_2 \leq \beta, \forall k$ ， $f(x)$ 在下水平集 $\mathcal{L} = \{x \mid f(x) \leq f(x^0)\}$ 上有下界，且 $\nabla f(x)$ 在 \mathcal{L} 的一个开邻域 $S(R_0)$ 内利普希茨连续。若 d^k 为信赖域子问题的近似解且满足(10)式，信赖域算法选取参数 $\eta = 0$ ，则

$$\liminf_{k \rightarrow \infty} \|\nabla f(x^k)\| = 0,$$

即 x^k 的聚点中包含稳定点。

全局收敛性1证明

首先由 ρ_k 的定义可以得到：

$$\begin{aligned} |\rho_k - 1| &= \left| \frac{(f(x^k) - f(x^k + d^k)) - (m_k(0) - m_k(d^k))}{m_k(0) - m_k(d^k)} \right| \\ &= \left| \frac{m_k(d^k) - f(x^k + d^k)}{m_k(0) - m_k(d^k)} \right| \end{aligned} \quad (11)$$

再由泰勒展式可得： $(g(x^k) = \nabla f(x^k))$

$$f(x^k + d^k) = f(x^k) + g(x^k)^T d^k + \int_0^1 [g(x^k + td^k) - g(x^k)]^T d^k dt$$

其中 $t \in (0, 1)$,再由 m 的定义可得(接下页)

全局收敛性1证明

$$\begin{aligned} |m_k(d^k) - f(x^k + d^k)| &= \left| \frac{1}{2} d^{kT} B^k d^k - \int_0^1 [g(x^k + td^k) - g(x^k)]^T d^k dt \right| \\ &\leq (\beta/2) \|d^k\|^2 + \beta_1 \|d^k\|^2, \end{aligned} \quad (12)$$

其中 β_1 为Lipschitz常数,并假设 $\|d^k\| \leq R_0$ 以保证 x^k 和 $x^k + td^k$ 均位于 $S(R_0)$ 中.再用反证法,反设结论不成立,则存在 ε 和指标 K 有:

$$\|g^k\| \geq \varepsilon, \forall k \geq K$$

从(10)式可知

$$m_k(0) - m_k(d^k) \geq c_1 \|g^k\| \min \left\{ \Delta_k, \frac{\|g^k\|}{\|B^k\|_2} \right\} \geq c_1 \varepsilon \min \left\{ \Delta_k, \frac{\varepsilon}{\beta} \right\} \quad (13)$$

全局收敛性1证明

由(12)和(13)可知

$$|\rho_k - 1| \leq \frac{\Delta_k^2 (\beta/2 + \beta_1)}{c_1 \epsilon \min(\Delta_k, \epsilon/\beta)}. \quad (14)$$

定义

$$\bar{\Delta} = \min \left((1 - 2\bar{\rho}_1) \frac{c_1 \epsilon}{(\beta/2 + \beta_1)}, R_0 \right).$$

对于所有满足 $\Delta_k \leq \bar{\Delta}$ 的充分大 k ，下面给出 $|\rho_k - 1|$ 的上界。注意到 $c_1 \leq 1$ ， $\min(\Delta_k, \epsilon/\beta) = \Delta_k$ 。由此可得：

$$|\rho_k - 1| \leq \frac{\Delta_k^2 (\beta/2 + \beta_1)}{c_1 \epsilon \Delta_k} \leq \frac{\bar{\Delta} (\beta/2 + \beta_1)}{c_1 \epsilon} \leq 1 - 2\bar{\rho}_1.$$

因此 $\rho_k > \bar{\rho}_1$ 。再由算法1可知 $\Delta_k \leq \bar{\Delta}$ 时 $\Delta_{k+1} \geq \Delta_k$ 成立。

全局收敛性1证明

因此如果 $k-1$ 失败且 $\Delta_k \leq \bar{\Delta}$, 则一定 $\Delta_{k-1} \geq \bar{\Delta}$. 所以我们有

$$\Delta_k \geq \min(\Delta_K, \gamma_1 \bar{\Delta}), \quad \forall k \geq K. \quad (15)$$

假设有含有无穷项的子列 \mathcal{K} 指标集, 使得 $\rho_k \geq \bar{\rho}_1, k \in \mathcal{K}$. 则对于 $k \in \mathcal{K}$, 并且 $k \geq K$, 由(13)可知

$$f(x^k) - f(x^{k+1}) \geq \bar{\rho}_1 [m_k(0) - m_k(d^k)] \geq \bar{\rho}_1 c_1 \varepsilon \min(\Delta_k, \varepsilon/\beta). \quad (16)$$

再由于 f 的下水平集有界, 由该不等式可以得到

$$\lim_{k \in \mathcal{K}, k \rightarrow \infty} \Delta_k = 0,$$

与(15)矛盾, 由此可知 \mathcal{K} 不存在, 所以 k 足够大时, Δ_k 将会在每次迭代中缩小 γ_1 倍, 所以 $\lim_{k \rightarrow \infty} \Delta_k = 0$, 与(15)矛盾, 因此原假设不成立.

全局收敛性

定理(全局收敛性1)表明若无条件接受信赖域子问题的更新, 则信赖域算法仅仅有子序列的收敛性, 迭代点序列本身不一定收敛. 根据下面的定理则说明选取 $\eta > 0$ 可以改善收敛性结果.

定理 (全局收敛性2)

在定理(全局收敛性1)的条件下, 若信赖域算法选取参数 $\eta > 0$, 且信赖域子问题近似解 d^k 满足(10)式, 则

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

和牛顿类算法不同, 信赖域算法具有全局收敛性, 因此它对迭代初值选取的要求比较弱. 而牛顿法的收敛性极大地依赖初值的选取.

全局收敛性2证明

Proof.

取 $g^m \neq 0$ ($g^m = \nabla f(x^m)$), 否则, $g^m \equiv 0$, 矛盾, 设 β_1 为 Lipschitz 常数, 则有

$$\|g(x) - g^m\| \leq \beta_1 \|x - x^m\|.$$

对任意 $x \in S(R_0)$ 均成立, 再定义 ϵ 和 R 如下:

$$\epsilon = \frac{1}{2} \|g^m\|, \quad R = \min \left(\frac{\epsilon}{\beta_1}, R_0 \right).$$

并且注意到 $\mathcal{B}(x^m, R) = \{x \mid \|x - x^m\| \leq R\}$ 包含在 $S(R_0)$ 中, 故 Lipschitz 连续性成立, 则有

$$x \in \mathcal{B}(x^m, R) \Rightarrow \|g(x)\| \geq \|g^m\| - \|g(x) - g^m\| \geq \frac{1}{2} \|g^m\| = \epsilon$$

全局收敛性2证明

Proof.

若 $\{x^k\}_{k \geq m}$ 均在 $\mathcal{B}(x^m, R)$ 中, 则有 $\|g^k\| \geq \epsilon > 0$ 对所有 $k \geq m$ 成立, 则由全局收敛性1的证明可知此结论不成立, 所以设 $l \geq m$, 且 x^{l+1} 是第一个离开 $\mathcal{B}(x^m, R)$ 的迭代点. 则对于 $k = m, m+1, \dots, l$ 由(13)可得:

$$\begin{aligned} f(x^m) - f(x^{l+1}) &= \sum_{k=m}^l f(x^k) - f(x^{k+1}) \\ &\geq \sum_{k=m, x^k \neq x^{k+1}}^l \eta [m_k(0) - m_k(d^k)] \\ &\geq \sum_{k=m, x^k \neq x^{k+1}}^l \eta c_1 \min\left(\Delta_k, \frac{\epsilon}{\beta}\right) \end{aligned} \quad (17)$$

全局收敛性2证明

Proof.

所以若 $\Delta \leq \frac{\epsilon}{\beta}$ 恒成立, 则有

$$f(x^m) - f(x^{l+1}) \geq \sum_{k=m, x^k \neq x^{k+1}}^l \eta c_1 \epsilon \Delta_k \geq \eta c_1 \epsilon R = \eta c_1 \epsilon \min\left(\frac{\epsilon}{\beta_1}, R_0\right) \quad (18)$$

否则, 可得 $\Delta_k > \epsilon/\beta$ 对某些 $k = m, m+1, \dots, l$ 成立, 则有

$$f(x^m) - f(x^{l+1}) \geq \eta c_1 \epsilon \frac{\epsilon}{\beta} \quad (19)$$

由于 $\{f(x^k)\}_{k=0}^{\infty}$ 单调递减并且有下界, 则有

$$f(x^k) \downarrow f^*$$

全局收敛性2证明

Proof.

因此,利用(18)与(19)式, 则有

$$\begin{aligned} f(x^m) - f^* &\geq f(x^m) - f(x^{l+1}) \\ &\geq \eta c_1 \varepsilon \min\left(\frac{\varepsilon}{\beta}, \frac{\varepsilon}{\beta_1}, R_0\right) \\ &= \frac{1}{2} \eta c_1 \|g^m\| \min\left(\frac{\|g^m\|}{2\beta}, \frac{\|g^m\|}{2\beta_1}, R_0\right) > 0 \end{aligned} \quad (20)$$

再由于 $f(x^k) \downarrow f^*$, 必有 $g^m \rightarrow 0$, 得证. □

局部收敛性

- 在构造信赖域子问题时利用了 $f(x)$ 的二阶信息，它在最优点附近应该具有牛顿法的性质。特别地，当近似矩阵 B^k 取为海瑟矩阵 $\nabla^2 f(x^k)$ 时，根据信赖域子问题的更新方式，二次模型 $m_k(d)$ 将会越来越逼近原函数 $f(x)$ 。

定理

设 $f(x)$ 在最优点 $x = x^*$ 的一个邻域内二阶连续可微，且 $\nabla f(x)$ 利普希茨连续，在最优点 x^* 处二阶充分条件成立，即 $\nabla^2 f(x) \succ 0$ 。若迭代点列 $\{x^k\}$ 收敛到 x^* ，并且在迭代中选取 B^k 为海瑟矩阵 $\nabla^2 f(x^k)$ ，且对充分大的 k ，任意满足(10)式的信赖域子问题算法产生的迭代方向 d^k 均满足

$$\|d^k - d_N^k\| = o(\|d_N^k\|), \quad (21)$$

其中 d_N^k 为第 k 步迭代的牛顿方向且满足假设 $\|d_N^k\| \leq \frac{\Delta_k}{2}$ 。则当 k 足够大时，信赖域约束 Δ_k 将未被激活。

局部收敛性证明

Proof.

我们证明在 k 充分大时 $\|d_N^k\| \leq \frac{1}{2}\Delta_k$ 与 $\|d^k\| \leq \Delta_k$. 则结论得证.
设 k 充分大时有 $o(\|d_N^k\|) < \|d_N^k\|$, 当 $\|d_N^k\| \geq \frac{1}{2}\Delta_k$ 时, 则有 $\|d^k\| \geq 2\|d_N^k\|$, 但若 $\|d_N^k\| > \frac{1}{2}\Delta_k$, 我们有 $\|d^k\| \geq \Delta_k \geq 2\|d_N^k\|$, 两种情况均有: $\|d^k\| \leq 2\|d_N^k\| \leq 2\|\nabla^2 f(x^k)^{-1}\| \|g^k\|$ 再由(10)可知:

$$\begin{aligned} m_k(0) - m_k(d^k) &\geq c_1 \|g^k\| \min\left(\Delta_k, \frac{\|g^k\|}{\|\nabla^2 f(x^k)\|}\right) \\ &\geq c_1 \frac{\|d^k\|}{2\|\nabla^2 f(x^k)^{-1}\|} \min\left(\|d^k\|, \frac{\|d^k\|}{2\|\nabla^2 f(x^k)\| \|\nabla^2 f(x^k)^{-1}\|}\right) \\ &\geq c_1 \frac{\|d^k\|^2}{4\|\nabla^2 f(x^k)^{-1}\|^2 \|\nabla^2 f(x^k)\|} \end{aligned}$$

局部收敛性证明

Proof.

再由于 $x^k \rightarrow x^*$, 并利用 $\nabla^2 f(x)$ 的连续性与 $\nabla^2 f(x^*)$ 的正定性知当 k 充分大时

$$\frac{c_1}{4 \left\| \nabla^2 f(x^k)^{-1} \right\|^2 \left\| \nabla^2 f(x^k) \right\|} \geq \frac{c_1}{8 \left\| \nabla^2 f(x^*)^{-1} \right\|^2 \left\| \nabla^2 f(x^*) \right\|} \stackrel{\text{def}}{=} c_3$$

因此可得当 k 足够大时有：

$$m_k(0) - m_k(d^k) \geq c_3 \|d^k\|^2$$

再利用 $\nabla^2 f(x)$ 的 Lipschitz 连续性并利用泰勒展式可得

$$\left| (f(x^k) - f(x^k + d^k)) - (m_k(0) - m_k(d^k)) \right|$$

局部收敛性证明

Proof.

$$\begin{aligned} &= \left| \frac{1}{2} d^k T \nabla^2 f(x^k) d^T - \frac{1}{2} \int_0^1 d^{kT} \nabla^2 f(x^k + td^k) d^k dt \right| \\ &\leq \frac{L}{4} \|d^k\|^3 \end{aligned}$$

其中 L 为Lipschitz常数, 因此由 ρ_k 的定义可得 k 足够大时:

$$|\rho_k - 1| \leq \frac{\|d^k\|^3 (L/4)}{c_3 \|d^k\|^2} = \frac{L}{4c_3} \|d^k\| \leq \frac{L}{4c_3} \Delta_k$$

又由于信赖域半径只可能在 $\rho_k < \frac{1}{4}$ 时减小, 所以由上式可得序列 Δ_k 不收敛到0, 又由于 $x^k \rightarrow x^*$, 所以我们有 $\|d_N^k\| \rightarrow 0$, 因此 $\|d^k\| \rightarrow 0$, 所以信赖域半径约束在 k 足够大时未激活, 并且 $\|d_N^k\| \leq \frac{1}{2} \Delta_k$ 最终一直成立

局部收敛性

推论 (信赖域算法的局部收敛速度)

在定理5的条件下, 信赖域算法产生的迭代序列 $\{x^k\}$ 具有Q-超线性收敛速度.

Proof.

根据牛顿法收敛性定理, 对牛顿方向,

$$\|x^k + d_N^k - x^*\| = \mathcal{O}(\|x^k - x^*\|^2),$$

因此得到估计 $\|d_N^k\| = \mathcal{O}(\|x^k - x^*\|)$. 又根据牛顿法的局部收敛性定理,

$$\|x^k + d^k - x^*\| \leq \|x^k + d_N^k - x^*\| + \|d^k - d_N^k\| = o(\|x^k - x^*\|).$$

这说明信赖域算法是Q-超线性收敛的. □

收敛性

- 容易看出，若在迭代后期 $d^k = d_N^k$ 能得到满足，则信赖域算法是Q-二次收敛的。
- 很多算法都会有这样的性质，例如前面提到的截断共轭梯度法和dogleg方法。因此在实际应用中，截断共轭梯度法是最常用的信赖域子问题的求解方法，使用此方法能够同时兼顾全局收敛性和局部Q-二次收敛性。

- 1 信赖域算法框架
- 2 信赖域子问题
- 3 柯西点
- 4 收敛性
- 5 应用举例

- 考虑逻辑回归问题

$$\min_x \frac{1}{m} \sum_{i=1}^m \ln(1 + \exp(-b_i a_i^T x)) + \lambda \|x\|_2^2, \quad (22)$$

这里选取 $\lambda = \frac{1}{100m}$.

- 同样地，我们选取LIBSVM上的数据集，调用信赖域算法求解代入数据集后的问题(22)，其迭代收敛过程见图1.
- 其中使用截断共轭梯度法来求解信赖域子问题，精度设置同牛顿法一致.

应用举例

从图中可以看到，在精确解附近梯度范数具有Q-超线性收敛性质。由于这个问题是强凸的，所以选取一个较大的初始信赖域半径 (\sqrt{n})。在数据集a9a和ijcnn1的求解中，信赖域子问题的求解没有因为超出信赖域边界而停机，因此牛顿法的数值表现一致。

