运筹学习题课

侯馨翔

2021 年 12 月 14 日

1 拟牛顿法收敛性

1.1 拟牛顿法的全局收敛性

BFGS格式：

$$B_{k+1} = B_k + \frac{y_k y_k^T}{s_k^T y_k} - \frac{B_k s_k (B_k s_k)^T}{(s_k)^T B_k s_k} \tag{1.1}$$

定义 1.1 (Wolfe 准则)**.** 设d_k是点x_k处的下降方向，若

$$f(x_k + \alpha d_k) \le f(x_k) + c_1 \alpha \nabla f(x_k)^T d_k \tag{1.2}$$

$$\nabla f(x_k + \alpha d_k)^T d_k \ge c_2 \nabla f(x_k)^T d_k \tag{1.3}$$

则称步长α满足$Wolfe$准则，其中c_1， $c_2 \in (0,1)$为给定的常数且$c_1 < c_2$.

定理 1.1 (Zoutendijk)**.** 考虑一般的迭代格式

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.4}$$

其中d_k是搜索方向，α_k是步长，且在迭代过程中$Wolfe$准则满足.假设目标函数f下有界、连续可微且梯度$L-Lip$连续，即

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n,$$

那么

$$\sum_{k=1}^{\infty} \cos^2\theta_k \|\nabla f(x_k)\|^2 < +\infty, \tag{1.5}$$

其中$\cos\theta_k$为负梯度$-\nabla f(x_k)$和下降方向d_k夹角的余弦，即

$$\cos\theta_k = \frac{-\nabla f(x_k)^T d_k}{\|\nabla f(x_k)\| \|d_k\|}$$

不等式 *(1.5)*也被称为*Zoutendijk*条件

证明. 由条件1.3,

$$(\nabla f(x_{k+1}) - \nabla f(x_k))^T d_k \geq (c_2 - 1)\nabla f(x_k)^T d_k.$$

由柯西不等式和梯度$L - Lip.$连续性质，

$$(\nabla f(x_{k+1}) - \nabla f(x_k))^T d_k \leq \|\nabla f(x_{k+1}) - \nabla f(x_k)\| \|d_k\| \leq \alpha_k L \|d_k\|^2.$$

结合上述两市可得

$$\alpha_k \geq \frac{c_2 - 1}{L} \frac{\nabla f(x_k)^T d_k}{\|d_k\|^2}.$$

注意到$\nabla f(x_k)^T d_k < 0$，将上式代入条件1.2，则

$$f(x_{k+1}) \leq f(x_k) + c_1 \frac{c_2 - 1}{L} \frac{(\nabla f(x_k)^T d_k)^2}{\|d_k\|^2}.$$

根据θ_k的定义，此不等式可等价表述为

$$f(x_{k+1}) \leq f(x_k) + c_1 \frac{c_2 - 1}{L} \cos^2\theta_k \|\nabla f(x_k)\|^2.$$

再关于k求和，我们有

$$f(x_{k+1}) \leq f(x_0) - c_1 \frac{1 - c_2}{L} \sum_{j=0}^{k} \cos^2\theta_j \|\nabla f(x_j)\|^2.$$

又因为函数f是下有界的，且由$0 < c_1 < c_2 < 1$可知$c_1(1 - c_2) > 0$，因此当$k \to \infty$时，

$$\sum_{j=1}^{\infty} \cos^2\theta_j \|\nabla f(x_j)\|^2 < +\infty.$$

□

定理 1.2. 设$x, y, u, v \in \mathbb{R}^n$，则

$$\det(I_{n\times n} + xy^T + uv^T) = (1 + y^T x)(1 + v^T u) - (x^T v)(y^T u). \tag{1.6}$$

证明. 利用降阶公式可得：

$$\begin{aligned}\det(I_{n\times n} + xy^T + uv^T) = \quad & \det\left(I_{n\times n} + \begin{pmatrix} x & u \end{pmatrix}\begin{pmatrix} y^T \\ v^T \end{pmatrix}\right) \\ = \quad & \det\left(I_{2\times 2} + \begin{pmatrix} x^T \\ u^T \end{pmatrix}\begin{pmatrix} y & v \end{pmatrix}\right)\end{aligned}$$

□

定理 1.3 (BFGS全局收敛性)**.** 假设初始矩阵B^0是对称正定矩阵，目标函数$f(x)$是二阶连续可微函数，且下水平集

$$\mathcal{L} = \{x \in \mathbb{R}^n | f(x) \le f(x_0)\} \tag{1.7}$$

是凸的，并且存在整数m以及M使得对于任意的$z \in \mathbb{R}^n$以及任意的$x \in \mathcal{L}$有

$$m\|z\|^2 \le z^T \nabla^2 f(x) z \le M\|z\|^2 \tag{1.8}$$

则采用*BFGS*格式并结合*Wolfe*线搜索的拟牛顿算法全局收敛到$f(x)$的极小值点x^*.

证明. 为了方便，定义

$$m_k = \frac{y_k^T s_k}{s_k^T s_k}, \quad M_k = \frac{y_k^T y_k}{y_k^T s_k}.$$

由1.8以及$y_k = \int_0^1 \nabla^2 f(x_k + t(x_{k+1} - x_k)) s_k dt$可得

$$m_k \ge m, \quad M_k \le M. \tag{1.9}$$

根据1.1，可得

$$\mathrm{Tr}(B_{k+1}) = \mathrm{Tr}(B_k) - \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \frac{\|y_k\|^2}{y_k^T s_k} \tag{1.10}$$

又由1.2，可得

$$\det B_{k+1} = \det B_k \frac{y_k^T s_k}{s_k^T B_k s_k} \tag{1.11}$$

定义

$$\cos\theta_k = \frac{s_k^T B_k s_k}{\|s_k\|\|B_k s_k\|}, \quad q_k = \frac{s_k^T B_k s_k}{s_k^T s_k},$$

那么将1.10整理可得

$$\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} = \frac{\|B_k s_k\|^2 \|s_k\|^2}{(s_k^T B_k s_k)^2} \frac{s_k^T B_k s_k}{\|s_k\|^2} = \frac{q_k}{\cos^2\theta_k} \tag{1.12}$$

同样由1.11可得

$$\det B_{k+1} = \det B_k \frac{y_k^T s_k}{s_k^T s_k} \frac{s_k^T s_k}{s_k^T B_k s_k} = \det B_k \frac{m_k}{q_k}$$

引进矩阵辅助函数

$$\psi(B) = \mathrm{Tr}(B) - \ln\det B,$$

则我们有

$$\begin{aligned}
&\psi(B_{k+1}) \\
&= \mathrm{Tr}(B_k) + M_k - \frac{q_k}{\cos^2\theta_k} - \ln\det B_k - \ln m_k + \ln q_k \\
&= \psi(B_k) + (M_k - \ln m_k - 1) + (1 - \frac{q_k}{\cos^2\theta_k} + \ln\frac{q_k}{\cos^2\theta_k}) + \ln\cos^2\theta_k.
\end{aligned} \tag{1.13}$$

根据1.13，以及$\psi(B) > 0, 1 - t + \ln t \le 0, \forall t$可以得到

$$0 < \psi(B_{k+1}) \le \psi(B_0) + (k+1)c + \sum_{j=0}^{k} \ln\cos^2\theta_j. \tag{1.14}$$

其中，$c = M - \ln m - 1$。不妨假设$c > 0$。注意到$s_k = -\alpha_k (B_k)^{-1}\nabla f(x_k)$是搜索方向，那么$\cos\theta_k$即是搜索方向与梯度方向夹角余弦。根据定理1.1可知$\|\nabla f(x_k)\|$大于某个非零常数仅当$\cos\theta_k \to 0$。因此，为了证明$\|\nabla f(x_k)\| \to 0$，我们仅需证明$\cos\theta_k \to 0$不成立。

反证法：假设$\cos\theta_k \to 0$，那么存在$k_1 > 0$，对于任意的$j > k_1$，

$$\ln\cos^2\theta_j < -2c.$$

结合1.14，当$k > k_1$时，

$$\begin{aligned}
0 < \psi(B_{k+1}) &\le \psi(B_0) + (k+1)c + \sum_{j=1}^{k_1} \ln\cos^2\theta_j + \sum_{j=k_1}^{k} (-2c) \\
&= \psi(B_0) + \sum_{j=0}^{k_1} \ln\cos^2\theta_j + 2ck_1 + C - ck.
\end{aligned} \tag{1.15}$$

上式右边对于充分大的k是负的，而不等式左边是0，矛盾。因此存在一个子列$\{j_k\}_{k=1,2,...}$使得$\cos\theta_{jk} \geq \delta > 0$。根据定理1.1，可以得到$\liminf_{k\to\infty} \|\nabla f(x_k)\| \to 0$。又因为问题对$x \in \mathcal{L}$是强凸的，所以有$x_k \to x^*$ $\square$

1.2 拟牛顿法超线性收敛速度

本节适用于一般非线性目标函数（不仅仅是凸函数）。首先我们需要一些额外的假设。

定理 1.4. *设$f : \mathbb{R}^n \to \mathbb{R}$是二阶连续可微。考虑迭代步$x_{k+1} = x_k + \alpha_k d_k$，其中$d_k$是下降方向，当$c_1 \leq 1/2$时$\alpha_k$满足Wolfe条件。若$\{x_k\}$收敛到点$x^*$使得$\nabla f(x^*) = 0$且$\nabla^2 f(x^*)$时正定的，并且满足*

$$\liminf_{k\to\infty} \frac{\|\nabla f(x_k) + \nabla^2 f(x_k) d_k\|}{\|d_k\|} = 0 \tag{1.16}$$

则

(1)对于$k > k_0$的步长取$\alpha_k = 1$

(2)若$\alpha_k = 1 \quad \forall k > k_0$，$\{x_k\}$超线性收敛至$x^*$.

若d_k是拟牛顿搜索方向，那么1.16等价于

$$\liminf_{k\to\infty} \frac{\|(B_k - \nabla^2 f(x^*)) d_k\|}{\|d_k\|} = 0. \tag{1.17}$$

首先，我们记

$$\sum_{k=1}^{\infty} \|x_k - x^*\| < \infty \tag{1.18}$$

假设 1.1. *海瑟矩阵G在x^*上Lipschitz连续，也就是说对于在x^*附近的点x，均满足*

$$\|G(x) - G(x^*)\| \leq L\|x - x^*\|,$$

其中L为一个正常数。

首先，我们有

$$\tilde{s}_k = G_*^{1/2} s_k, \quad \tilde{y}_k = G_*^{-1/2} y_k, \quad \tilde{B}_k = G_*^{-1/2} B_k G_*^{-1/2},$$

其中$G_* = G(x^*)$。我们定义

$$\cos\tilde{\theta}_k = \frac{\tilde{s}_k^T \tilde{B}_k \tilde{s}_k}{\|\tilde{s}_k\|\|\tilde{B}_k \tilde{s}_k\|}, \quad \tilde{q}_k = \frac{\tilde{s}_k^T \tilde{B}_k \tilde{s}_k}{\|\tilde{s}_k\|^2},$$

令

$$\tilde{M}_k = \frac{\|\tilde{y}_k\|^2}{\tilde{y}_k^T \tilde{s}_k}, \quad \tilde{m}_k = \frac{\tilde{y}_k^T \tilde{s}_k}{\tilde{s}_k^T \tilde{s}_k}$$

由BFGS的迭代公式可以得到

$$\tilde{B}_{k+1} = \tilde{B}_k - \frac{\tilde{B}_k \tilde{s}_k \tilde{s}_k^T \tilde{B}_k}{\tilde{s}_k^T \tilde{B}_k \tilde{s}_k} + \frac{\tilde{y}_k \tilde{y}_k^T}{\tilde{y}_k^T \tilde{s}_k}$$

由于这个表达式与BFGS的更新公式由相同的形式，那么由1.13同样地可得

$$\psi(\tilde{B}_{k+1}) = \psi(\tilde{B}_k) + (\tilde{M}_k - \ln\tilde{m}_k - 1) + [1 - \frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k} + \ln\frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k}] + \ln\cos^2\tilde{\theta}_k. \tag{1.19}$$

而由$y_k = G_k s_k$可知

$$y_k - G_* s_k = (G_k - G_*) s_k$$

而且

$$\tilde{y}_k - \tilde{s}_k = G_*^{-1/2}(G_k - G_*)G_*^{-1/2}\tilde{s}_k.$$

由假设1.1可知

$$\|\tilde{y}_k - \tilde{s}_k\| \le \|G_*^{-1/2}\|^2 \|\tilde{s}_k\| \|G_k - G_*\| \le \|G_*^{-1/2}\|^2 \|\tilde{s}_k\| L\epsilon_k,$$

其中$\epsilon_k = \max\{\|x^{k+1} - x^*\|, \|x_k - x^*\|\}$。所以我们可以得到

$$\frac{\|\tilde{y}_k - \tilde{s}_k\|}{\|\tilde{s}_k\|} \le \bar{c}\epsilon_k, \tag{1.20}$$

其中$\bar{c}$为正常数。

定理 1.5. *设f为二阶连续可微并且在假设1.1下使用BFGS算法迭代收敛到最优点x^*.假设1.18成立，那么x_k以超线性收敛速度收敛至x^*.*

证明. 由1.20我们有

$$\|\tilde{y}_k\| - \|\tilde{s}_k\| \le \bar{c}\epsilon_k \|\tilde{s}_k\|, \quad \|\tilde{s}_k\| - \|\tilde{y}_k\| \le \bar{c}\epsilon_k \|\tilde{s}_k\|,$$

因此

$$(1 - \bar{c}\epsilon_k)\|\tilde{s}_k\| \le \|\tilde{y}_k\| \le (1 + \bar{c}\epsilon_k)\|\tilde{s}_k\|. \tag{1.21}$$

由1.20和1.21，我们有

$$(1-\bar{c}\epsilon_k)^2\|\tilde{s}_k\|^2 - 2\tilde{y}_k^T\tilde{s}_k + \|\tilde{s}_k\|^2 \le \|\tilde{y}_k\|^2 - 2\tilde{y}_k^T\tilde{s}_k + \|\tilde{s}_k\|^2 \le \bar{c}^2\epsilon_k^2\|\tilde{s}_k\|^2,$$

因此

$$2\tilde{y}_k^T\tilde{s}_k \ge (1-2\bar{c}\epsilon_k+\bar{c}^2\epsilon_k+1-\bar{c}^2\epsilon_k^2)\|\tilde{s}_k\|^2 = 2(1-\bar{c}\epsilon_k)\|\tilde{s}_k\|^2.$$

由$\tilde{m}_k$的定义可以得到

$$\tilde{m}_k = \frac{\tilde{y}_k^T\tilde{s}_k}{\|\tilde{s}_k\|^2} \ge 1-\bar{c}\epsilon_k \tag{1.22}$$

那么由1.21和1.22可以得到

$$\tilde{M}_k = \frac{\|\tilde{y}_k\|^2}{\tilde{y}_k^T\tilde{s}_k} \le \frac{1+\bar{c}\epsilon_k}{1-\bar{c}\epsilon_k} \tag{1.23}$$

由于$x_k \to x^*$，我们有$\epsilon_k \to 0$，由1.23存在一个正常数$c > \bar{c}$，当k足够大时

$$\tilde{M}_k \le 1 + \frac{2\bar{c}}{1-\bar{c}\epsilon_k}\epsilon_k \le 1 + c\epsilon_k \tag{1.24}$$

我们利用函数$h(t) = 1 - t + \ln t$的非正性，我们有

$$\frac{-x}{1-x} - \ln(1-x) = h(\frac{1}{1-x}) \le 0.$$

对充分大的k我们可以假设$\bar{c}\epsilon_k < \frac{1}{2}$，因此

$$\ln(1-\bar{c}\epsilon_k) \ge \frac{-\bar{c}\epsilon_k}{1-\bar{c}\epsilon_k} \ge -2\bar{c}\epsilon_k.$$

结合1.22可以得到

$$\ln\tilde{m}_k \ge \ln(1-\bar{c}\epsilon_k) \ge -2\bar{c}\epsilon_k > -2c\epsilon_k \tag{1.25}$$

我们由1.19,1.24,1.25可得

$$0 < \psi(\tilde{B}_{k+1}) \le \psi(\tilde{B}_k) + 3c\epsilon_k + \ln\cos^2\tilde{\theta}_k + [1 - \frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k} + \ln\frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k}]. \tag{1.26}$$

由1.18可得

$$\sum_{j=0}^{\infty}\left(\ln\frac{1}{\cos^2\tilde{\theta}_j} - \left[1 - \frac{\tilde{q}_j}{\cos^2\tilde{\theta}_j} + \ln\frac{\tilde{q}_j}{\cos^2\tilde{\theta}_j}\right]\right) \le \psi(\tilde{B}_0) + 3c\sum_{j=0}^{\infty}\epsilon_j < +\infty.$$

由于$\ln\left(1/\cos^2\tilde{\theta}_j\right) \ge 0\forall j$，我们可以得到

$$\lim_{j\to\infty}\ln\frac{1}{\cos^2\tilde{\theta}_j} = 0, \quad \lim_{j\to\infty}\left(1 - \frac{\tilde{q}_j}{\cos^2\tilde{\theta}_j} + \ln\frac{\tilde{q}_j}{\cos^2\tilde{\theta}_j}\right) = 0,$$

这表明了

$$\lim_{j\to\infty}\cos\tilde{\theta}_j = 1,\quad \lim_{j\to\infty}\tilde{q}_j = 1 \tag{1.27}$$

由1.12我们有

$$\begin{aligned}\frac{\|G_*^{-1/2}(B_k-G_*)s_k\|^2}{\|G_*^{1/2}s_k\|^2} &= \frac{\|(\tilde{B}_k-I)\tilde{s}_k\|^2}{\|\tilde{s}_k\|^2}\\ &= \frac{\|\tilde{B}_k\tilde{s}_k\|^2-2\tilde{s}_k^T\tilde{B}_k\tilde{s}_k+\tilde{s}_k^T\tilde{s}_k}{\tilde{s}_k^T\tilde{s}_k}\\ &= \frac{\tilde{q}_k^2}{\cos\tilde{\theta}_k^2}-2\tilde{q}_k+1.\end{aligned}$$

由1.27右边收敛至0，我们可以得到

$$\lim_{k\to\infty}\frac{\|(B_k-G_*)s_k\|}{\|s_k\|}=0.$$

收敛。1.17和1.4表示当$\alpha_k=1$将会满足在最优解附近Wolfe条件，因此收敛速率为超线性的。 □

2 共轭梯度法

2.1 FR共轭梯度法

采用精确线性搜索的再开始共轭梯度法

算法：

给出初始点x_0，容限$\epsilon > 0$.

Step 1: 令$k=0$，计算$g_0=g(x_0)$.

Step 2: 若$\|g_0\|\le\epsilon$，停止迭代，输出$x^*=x_0$；否则，令$d_0=-g_0$.

Step 3: 一维搜索求α_k，使得

$$f(x_k+\alpha_k d_k)=\min_{\alpha}\{f(x_k+\alpha d_k)|\alpha\ge 0\}$$

Step 4: 令$x_{k+1} = x_k + \alpha_k d_k, k := k+1$.

Step 5: 计算$g_k = g(x_k)$.

Step 6: 如果$\|g_k\| \le \epsilon$，停止迭代；否则转Step 7.

Step 7: 若$k = n$，令$x_0 := x_k$，并转Step 1；否则转Step 8.

Step 8: 计算

$$\beta = g_k^T g_k / g_{k-1}^T g_{k-1}, \quad d_k = -g_k + \beta d_{k-1}.$$

Step 9: 如果$d_k^T g_k > 0$，令$x_0 := x_k$，并转Step 1；否则转Step 3.

2.2 FR共轭梯度法收敛性

定理 2.1 (FR共轭梯度法全局收敛性定理). 假定在有界水平集$\mathcal{L} = \{x \in R^n | f(x) \le f(x_0)\}$上$f : R^n \to R$连续可微，那么采用

$$\beta = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \tag{2.1}$$

和精确线性搜索的FR共轭梯度法产生的序列$\{x_k\}$至少有一个聚点是驻点，即

(1)当$\{x_k\}$是有穷点列时，其最后一个点x^*是f的驻点.

(2)当$\{x_k\}$是无穷点列时，它必有极限点，且其任一极限点是f的驻点.

证明. (1)当$\{x_k\}$是有穷点列时，由算法的终止性条件可知，其最后一点x^*满足$\nabla f(x^*)$，故x^*是f的驻点.

(2)当$\{x_k\}$是无穷点列时，则有$\nabla f(x_k) \ne 0$，由于$d_k = -g_k + \beta_{k-1} d_{k-1}$，故

$$g_k^T d_k = -\|g_k\|^2 + \beta_{k-1} g_k^T d_{k-1} = -\|g_k\|^2 < 0, \tag{2.2}$$

从而d_k是下降方向，$\{f(x_k)\}$是单调下降序列，$\{x_k\} \subset \mathcal{L}$，所以$\{x_k\}$是有界点列，必有极限点.

设x^*是$\{x_k\}$的极限点，则存在子列$\{x_k\}_{K_1}$收敛到x^*，这里K_1是子序列的指标集.由于$\{x_k\}_{K_1} \subset \{x_k\}$，故$\{f(x_k)\}_{K_1} \subset \{f(x_k)\}$，从而由$f$的连续性可知，对于$k \in K_1$，有

$$f(x^*) = f\left(\lim_{k\to\infty} x_k\right) = \lim_{k\to\infty} f(x_k) = f^* \tag{2.3}$$

类似地，$\{x_{k+1}\}$也是有界点列，故存在子列$\{x_{k+1}\}_{K_2}$收敛到x^*，这里K_2是$\{x_{k+1}\}$的子序列的指标集，并且

$$f(\bar{x}^*) = f\left(\lim_{k\to\infty} x_{k+1}\right) = \lim_{k\to\infty} f(x_{k+1}) = f^* \tag{2.4}$$

于是

$$f(\bar{x}^*) = f(x^*) = f^*. \tag{2.5}$$

现在用反证法证明$\nabla f(x^*) = 0$，假定$\nabla f(x^*) \neq 0$，则对于充分小的α，有

$$f(x^* + \alpha d^*) < f(x^*),$$

由于

$$f(x_{k+1}) = f(x_k + \alpha_k d_k) \leq f(x_k + \alpha d_k), \quad \forall \alpha > 0,$$

故对于$k \in K_2$，令$k \to \infty$，可得

$$f(\bar{x}^*) \leq f(x^* + \alpha d^*) < f(x^*), \tag{2.6}$$

这与2.5矛盾.因此，$\nabla f(x^*) = 0$，即x^*是f的驻点. □

2.3 FR共轭梯度法收敛速率

假设 2.1. *(1).* $f \in C^3(R^n, R)$，即$f : R^n \to R$三次连续可微。

*(2)*设存在常数$m > 0$和$M > 0$，使得

$$m\|y\|^2 \leq y^T \nabla^2 f(x) y \leq M\|y\|^2, \quad \forall y \in R^n, x \in L,$$

其中$L = \{x \in R^n | f(x) \leq f(x_0)\}$是有界水平集

定理 2.2. 假定假设*2.1*满足，那么，每r步再开始的FR共轭梯度法产生的迭代点列$\{x_k\}$$n$步二阶收敛，即存在常数$c>0$，使得

$$\lim_{k\to\infty}\sup\frac{\|x_{kr+n}-x^*\|}{\|x_{kr}-x^*\|^2}\leq c<\infty. \tag{2.7}$$

这个定理的证明分两步完成，第一步证明2.3，第二步证明2.4

定理 2.3. 设定理*2.2*的条件满足，并设在每一个点x_{kr}定义二次函数$\hat{f}_{kr}$为

$$\hat{f}_{kr}(x)=f(x_{kr})+\nabla f(x_{kr})^T(x-x_{kr})+\frac{1}{2}(x-x_{kr})^T\nabla^2 f(x_{kr})(x-x_{kr}). \tag{2.8}$$

设$\psi_{\hat{f}_{kr}}$表示应用到$\hat{f}_{kr}$上的共轭梯度法，并且令$d^0_{kr}=d_{kr}=-\nabla f(x_{kr})$.

又设

$$x^0_k=x_{kr},\quad x^1_{kr}=\psi_{\hat{f}_{kr}}\left(x^0_{kr}\right),\cdots,x^n_{kr}=\psi_{\hat{f}_{kr}}\left(x^{n-1}_{kr}\right)$$

那么，如果对于 $i=0,1,\cdots,j(k)-1$,

$$\left\|\alpha_{kr+i}d_{kr+i}-\alpha^i_{kr}d^i_{kr}\right\|=O\left(\|x_{kr}-x^*\|^2\right) \tag{2.9}$$

成立, 其中 $j(k)$ 是小于等于 n 的整数, 满足

$$x^{j(k)}_{kr}=\bar{x}\left(\hat{f}_{kr}\right),\quad x^{j(k)-1}_{kr}\neq\bar{x}\left(\hat{f}_{kr}\right),$$

$\bar{x}\left(\hat{f}_{kr}\right)$ 表示函数 $\hat{f}_{kr}$ 的极小点, 则再开始FR 共轭梯度法产生的点列满足*2.7*.

证明. 由于$\hat{f}_{kr}$是二次正定函数，共轭梯度法至多n步达到其极小点. 由算法可知,

$$f\left(x_{k+1}\right)\leq f\left(x_k\right),\quad\forall k \tag{2.10}$$

故

$$f\left(x_{k+1}\right)-f\left(x^*\right)\leq f\left(x_k\right)-f\left(x^*\right) \tag{2.11}$$

又由Taylor 定理,

$$\begin{aligned}f\left(x_k\right)-f\left(x^*\right)&=\nabla f\left(x^*\right)^T\left(x_k-x^*\right)+\left(x_k-x^*\right)^T\nabla^2 f(\eta)\left(x_k-x^*\right)\\&=(x_k-x^*)^T\nabla^2 f(\eta)(x_k-x^*)\end{aligned} \tag{2.12}$$

其中 $\eta = x_k + \alpha\left(x^* - x_k\right), \alpha \in (0,1)$.再利用2.1，则有

$$m\left\|x_k - x^*\right\|^2 \leq f\left(x_k\right) - f\left(x^*\right) \leq M\left\|x_k - x^*\right\|^2. \tag{2.13}$$

综合2.11和2.13, 得

$$\begin{aligned} m\left\|x_{k+1} - x^*\right\|^2 &\leq \left|f\left(x_{k+1}\right) - f\left(x^*\right)\right| \\ &\leq \left|f\left(x_k\right) - f\left(x^*\right)\right| \leq M\left\|x_k - x^*\right\|^2, \end{aligned}$$

即有

$$\left\|x_{k+1} - x^*\right\| \leq \sqrt{M/m}\left\|x_k - x^*\right\|. \tag{2.14}$$

因此,

$$\begin{aligned} \left\|x_{kr+n} - x^*\right\| &\leq (M/m)^{(n-j(k))/2}\left\|x_{kr+j(k)} - x^*\right\| \\ &\leq (M/m)^{n/2}\left\|x_{kr+j(k)} - x^*\right\| \end{aligned} \tag{2.15}$$

这表明，为了证明结论2.7, 只要证明

$$\left\|x_{kr+j(k)} - x^*\right\| = O\left(\left\|x_{kr} - x^*\right\|^2\right) \tag{2.16}$$

由于

$$\left\|x_{kr+j(k)} - x^*\right\| \leq \left\|x_{kr+j(k)} - \bar{x}\left(\hat{f}_{kr}\right)\right\| + \left\|\bar{x}\left(\hat{f}_{kr}\right) - x^*\right\|, \tag{2.17}$$

其中$\bar{x}\left(\hat{f}_{kr}\right)$是$\hat{f}_{kr}$的极小点,可以表示为

$$\bar{x}\left(\hat{f}_{kr}\right) = x_{kr} - \left[\nabla^2 f\left(x_{kr}\right)\right]^{-1}\nabla f\left(x_{kr}\right),$$

但这恰恰是在x_{kr}点应用到f上的牛顿法，这样

$$\left\|\bar{x}\left(\hat{f}_{kr}\right) - x^*\right\| = \left\|\hat{\phi}\left(x_{kr}\right) - x^*\right\|.$$

由于牛顿法二阶收敛, 便有

$$\left\|\hat{\phi}\left(x_{kr}\right) - x^*\right\| = O\left(\left\|x_{kr} - x^*\right\|^2\right). \tag{2.18}$$

因此, 从2.18,2.17和2.18可知,我们需要证明

$$\begin{aligned} \left\|x_{kr+j(k)} - \bar{x}\left(\hat{f}_{kr}\right)\right\| &= \left\|x_{kr+j(k)} - x_{kr}^{j(k)}\right\| \\ &= O\left(\left\|x_{kr} - x^*\right\|^2\right). \end{aligned} \tag{2.19}$$

令$x_{kr}=x_{kr}^0$,这样,

$$
\begin{aligned}
\left\|x_{kr+j(k)}-x_{kr}^{j(k)}\right\| &= \left\|\sum_{i=0}^{j(k)-1}\left[\left(x_{kr+i+1}-x_{kr+i}\right)-\left(x_{kr}^{i+1}-x_{kr}^{i}\right)\right]\right\| \\
&= \left\|\sum_{i=0}^{j(k)-1}\left[\alpha_{kr+i}d_{kr+i}-\alpha_{kr}^{i}d_{kr}^{i}\right]\right\| \\
&\leq \sum_{i=0}^{j(k)-1}\left\|\alpha_{kr+i}d_{kr+i}-\alpha_{kr}^{i}d_{kr}^{i}\right\|.
\end{aligned} \tag{2.20}
$$

从2.17、2.18、2.19和2.20可知,2.9意味着2.7成立. □

第二部分要证明2.9对于共轭梯度法成立.为此, 我们先给出几个引理.

引理 2.1.

$$\alpha_k=\frac{\|g_k\|^2}{d_k^T\hat{B}_kd_k} \tag{2.21}$$

其中

$$\widehat{B}_k=\int_0^1\nabla^2 f\left(x_k+\xi\alpha_kd_k\right)d\xi. \tag{2.22}$$

证明. $g_{k+1}=g_k+\alpha_k\hat{B}_kd_k$,由于$g_{k+1}^Td_k=0$,故有$g_k^Td_k+\alpha_kd_k^T\hat{B}_kd_k=0$,从而

$$\alpha_k=\frac{-g_k^Td_k}{d_k^T\hat{B}_kd_k}=\frac{\|g_k\|^2}{d_k^T\hat{B}_kd_k}.$$

□

引理 2.2.

$$\|g_{k+1}\|\leq(1+M/m)\|d_k\|. \tag{2.23}$$

证明. 由于$d_{k+1}=-g_{k+1}+\beta_kd_k$和$g_{k+1}^Td_k=0$,故

$$\|d_{k+1}\|^2=\|g_{k+1}\|^2+\beta_k^2\|d_k\|^2. \tag{2.24}$$

这表明

$$\|g_{k+1}\|\leq\|d_{k+1}\| \tag{2.25}$$

利用2.1、2.21和2.25,

$$\alpha_k = \frac{\|g_k\|^2}{d_k^T \hat{B}_k d_k} \leq \frac{\|g_k\|^2}{m\|d_k\|^2} \leq \frac{1}{m}. \tag{2.26}$$

于是可得

$$\begin{aligned} \|g_{k+1}\| &\leq \|g_k\| + |\alpha_k| \left\|\hat{B}_k d_k\right\| \\ &\leq \|d_k\| + \frac{M}{m}\|d_k\| \\ &= \left(1 + \frac{M}{m}\right)\|d_k\|. \end{aligned}$$

□

引理 2.3. 对于 *FR* 公式,

$$\beta_k = 1 + \frac{(g_{k+1} + g_k)^T \hat{B}_k d_k}{d_k^T \hat{B}_k d_k}, \tag{2.27}$$

$$|\beta_k| \leq \left(1 + \frac{M}{m}\right)^2. \tag{2.28}$$

证明. 对于 FR 公式,

$$\begin{aligned} \beta_k &= \frac{\|g_{k+1}\|^2}{\|g_k\|^2} = \frac{g_{k+1}^T (g_{k+1} - g_k) + g_{k+1}^T g_k}{\|g_k\|^2} \\ &= \frac{g_{k+1}^T \hat{B}_k d_k}{d_k^T \hat{B}_k d_k} + \frac{g_{k+1}^T g_k}{\|g_k\|^2} \end{aligned} \tag{2.29}$$

由于

$$\begin{aligned} \frac{g_{k+1}^T g_k}{\|g_k\|^2} &= \frac{\left(g_k + \alpha_k \hat{B}_k d_k\right)^T g_k}{\|g_k\|^2} \\ &= 1 + \alpha_k \frac{g_k^T \hat{B}_k d_k}{\|g_k\|^2} \\ &= 1 + \frac{g_k^T \hat{B}_k d_k}{d_k^T \hat{B}_k d_k}, \end{aligned} \tag{2.30}$$

于是由2.29和2.30得

$$\beta_k = 1 + \frac{(g_{k+1} + g_k)^T \hat{B}_k d_k}{d_k^T \hat{B}_k d_k}$$

上式即为2.27。利用2.27, 2.25, 2.23有

$$\begin{aligned}|\beta_k| &\leq 1+\left|\frac{\left(g_{k+1}+g_k\right)^T \hat{B}_k d_k}{d_k^T \hat{B}_k d_k}\right| \\ &\leq 1+\frac{\left(\|g_{k+1}\|+\|g_k\|\right) M \|d_k\|}{m \|d_k\|^2} \\ &\leq 1+\frac{M}{m}\left(\frac{\|g_{k+1}\|}{\|d_k\|}+\frac{\|g_k\|}{\|d_k\|}\right) \\ &\leq 1+\frac{M}{m}\left(1+\frac{M}{m}+1\right) \\ &\leq \left(1+\frac{M}{m}\right)^2 .\end{aligned}$$

□

引理 2.4.

$$\|g_k\| \leq M\|x_k - x^*\| \tag{2.31}$$

$$\|d_{k+1}\| \leq (1+M/m)^2\|d_k\| \tag{2.32}$$

证明. $g_k = \nabla f(x^*) + \int_0^1 \nabla^2 f(x^* + \xi(x_k - x^*))d\xi(x_k - x^*)$,即可得2.31.
又

$$d_{k+1} = -g_{k+1} + \beta_k d_k,$$

故

$$\|d_{k+1}\| \leq \|g_{k+1}\| + |\beta_k| \, \|d_k\| , \tag{2.33}$$

利用2.23, 2.28可得

$$\|d_{k+1}\| \leq (1+M/m)^2 \, \|d_k\| .$$

□

引理 2.5.

$$\left\|\nabla^2 f\left(x_{k+i}\right)-\nabla^2 f\left(x_k\right)\right\| = O\left(\|d_k\|\right). \tag{2.34}$$

$$\left\|\hat{B}_{k+i}-\nabla^2 f\left(x_k\right)\right\| = O\left(\|d_k\|\right) \tag{2.35}$$

证明.

$$\left\|\nabla^2 f\left(x_{k+i}\right)-\nabla^2 f\left(x_k\right)\right\| \leq \sum_{l=0}^{i-1}\left\|\nabla^2 f\left(x_{k+l+1}\right)-\nabla^2 f\left(x_{k+l}\right)\right\|$$

由于

$$
\begin{aligned}
& \left\|\nabla^2 f\left(x_{k+l+1}\right)-\nabla^2 f\left(x_{k+l}\right)\right\| \\
= & \left\|\nabla^2 f\left(x_{k+l}+\alpha_{k+l} d_{k+l}\right)-\nabla^2 f\left(x_{k+l}\right)\right\| \\
= & \left\|\int_0^1 \nabla^3 f\left(x_{k+l}+\xi \alpha_{k+l} d_{k+l}\right) \alpha_{k+l} d_{k+l} d \xi\right\| \\
\leq & \sup_{\xi \in(0,1)}\left\|\nabla^3 f\left(x_{k+l}+\xi \alpha_{k+l} d_{k+l}\right)\right\|\left|\alpha_{k+l}\right|\left\|d_{k+l}\right\| .
\end{aligned} \tag{2.36}
$$

由于$x_k \rightarrow x^*, \nabla f\left(x^*\right)=0, f \in C^3$, 故存在$K_1$和$\bar{M}$, 使得对于 $k \geq K_1,\left\|\nabla^3 f\left(x_k\right)\right\| \leq \bar{M}$. 又利用2.26和2.32,2.36可以写成

$$
\left\|\nabla^2 f\left(x_{k+l+1}\right)-\nabla^2 f\left(x_{k+l}\right)\right\| \leq \frac{\bar{M}}{m}\left\|d_{k+l}\right\|=O\left(\left\|d_k\right\|\right) . \tag{2.37}
$$

因此,

$$
\left\|\nabla^2 f\left(x_{k+i}\right)-\nabla^2 f\left(x_k\right)\right\|=O\left(\left\|d_k\right\|\right) .
$$

此即2.34. 下面证明2.35.

$$
\|\hat{B}_{k+i}-\nabla^2 f\left(x_k\right)\| \leq\|\hat{B}_{k+i}-\nabla^2 f\left(x_{k+i}\right)\|+\left\|\nabla^2 f\left(x_{k+i}\right)-\nabla^2 f\left(x_k\right)\right\| \tag{2.38}
$$

由于

$$
\begin{aligned}
& \left\|\hat{B}_{k+i}-\nabla^2 f\left(x_{k+i}\right)\right\| \\
= & \left\|\int_0^1 \nabla^2 f\left(x_{k+i}+\xi \alpha_{k+i} d_{k+i}\right) d \xi-\nabla^2 f\left(x_{k+i}\right)\right\| \\
\leq & \int_0^1\left\|\nabla^2 f(x_{k+i}+\xi \alpha_{k+i} d_{k+i})-\nabla^2 f(x_{k+i})\right\| d \xi,
\end{aligned} \tag{2.39}
$$

$$
\begin{aligned}
& \|\nabla^2 f\left(x_{k+i}+\xi \alpha_{k+i} d_{k+i}\right)-\nabla^2 f\left(x_{k+i}\right)\| \\
= & \left\|\int_0^1 \nabla^3 f\left(x_{k+i}+\eta \xi \alpha_{k+i} d_{k+i}\right) \xi \alpha_{k+i} d_{k+i} d \eta\right\| \\
\leq & \sup_{\eta \in(0,1)}\left\|\nabla^3 f\left(x_{k+i}+\eta \alpha_{k+i} d_{k+i}\right)\right\|\left|\alpha_{k+i}\right|\left\|d_{k+i}\right\| \\
\leq & \frac{M}{m}\left\|d_{k+i}\right\|,
\end{aligned} \tag{2.40}
$$

因此,利用2.32,

$$
\left\|\hat{B}_{k+i}-\nabla^2 f\left(x_k\right)\right\| \leq \frac{\bar{M}}{m}\left\|d_{k+i}\right\|=O\left(\left\|d_k\right\|\right) . \tag{2.41}
$$

于是, 由2.38,2.34和2.41得2.35. □

引理 2.6. 对于FR公式, 有

$$\begin{aligned}\left\|d_{k+i+1}-d_k^{i+1}\right\| =&O_1\left(\left\|d_{k+i}-d_k^i\right\|\right)+O_2\left(\left\|g_{k+i+1}-g_k^{i+1}\right\|\right)\\&+O_3\left(\left\|g_{k+i}-g_k^i\right\|\right)+O_4\left(\left\|d_k\right\|^2\right).\end{aligned}\tag{2.42}$$

引理 2.7.

$$\begin{aligned}\left\|g_{k+i+1}-g_k^{i+1}\right\| \leq& \left\|g_{k+i}-g_k^i\right\|+O\left(\left\|d_k\right\|^2\right)\\&+M\|\alpha_{k+i}d_{k+i}-\alpha_k^i d_k^i\|.\end{aligned}\tag{2.43}$$

证明.

$$\begin{aligned}\|g_{k+i+1}-g_k^{i+1}\| &= \|g_{k+i}+\alpha_{k+i}\hat{B}_{k+i}d_{k+i}-g_k^i-\alpha_k^i\nabla^2 f\left(x_k\right)d_k^i\|\\
&\leq \left\|g_{k+i}-g_k^i\right\|+\left\|\left(\nabla^2 f\left(x_k\right)-\hat{B}_{k+i}\right)\alpha_k^i d_k^i\right\|\\
&\quad+\left\|\hat{B}_{k+i}\left(\alpha_k^i d_k^i-\alpha_{k+i}d_{k+i}\right)\right\|\\
&\leq \left\|g_{k+i}-g_k^i\right\|+\frac{1}{m}\left\|\nabla^2 f\left(x_k\right)-\hat{B}_{k+i}\right\|\left\|d_k^i\right\|\\
&\quad+M\left\|\alpha_k^i d_k^i-\alpha_{k+i}d_{k+i}\right\|\\
&\leq \left\|g_{k+i}-g_k^i\right\|+O\left(\left\|d_k\right\|^2\right)+M\left\|\alpha_k^i d_k^i-\alpha_{k+i}d_{k+i}\right\|.\end{aligned}$$

□

引理 2.8.

$$\begin{aligned}\left\|\alpha_{k+i}d_{k+i}-\alpha_k^i d_k^i\right\| =&O_1\left(\left\|g_{k+i}-g_k^i\right\|\right)+O_2\left(\left\|d_{k+i}-d_k^i\right\|\right)\\&+O_3\left(\left\|d_k\right\|^2\right).\end{aligned}\tag{2.44}$$

证明. 由2.21可知

$$
\begin{aligned}
&\left\|\alpha_{k+i}d_{k+i}-\alpha_k^i d_k^i\right\| = \left\|\frac{\|g_{k+i}\|^2 d_{k+i}}{d_{k+i}^T\hat{B}_{k+i}d_{k+i}} - \frac{\|g_k^i\|^2 d_k^i}{{d_k^i}^T\nabla^2 f(x_k) d_k^i}\right\| \\
&= \frac{1}{c_{k+i}}\left\|\|g_{k+i}\|^2\left({d_k^i}^T\nabla^2 f(x_k) d_k^i\right)d_{k+i} - \left\|g_k^i\right\|^2\left(d_{k+i}^T\hat{B}_{k+i}d_{k+i}\right)d_k^i\right\| \\
&\le \frac{1}{c_{k+i}}\{\left\|\left(g_{k+i}^T\left(g_{k+i}-g_k^i\right)\right)\left(d_k^i\nabla^2 f(x_k) d_k^i\right)d_{k+i}\right\| \\
&+\left\|\left(g_{k+i}^T g_k^i\right)\left(\left(d_k^i-d_{k+i}\right)^T\nabla^2 f(x_k) d_k^i\right)d_{k+i}\right\| \\
&+\left\|\left(\left(g_{k+i}-g_k^i\right)^T g_k^i\right)\left(d_{k+i}^T\nabla^2 f(x_k) d_k^i\right)d_{k+i}\right\| \\
&+\left\|\|g_k^i\|\|^2\left(d_{k+i}^T\nabla^2 f(x_k)\left(d_k^i-d_{k+i}\right)\right)d_{k+i}\right\| \\
&+\left\|\|g_k^i\|^2\left(d_{k+i}^T\nabla^2 f(x_k) d_{k+i}\right)\left(d_{k+i}-d_k^i\right)\right\| \\
&\left\|\|g_k^i\|^2\left(d_{k+i}^T\left(\nabla^2 f(x_k)-\hat{B}_{k+i}\right)d_{k+i}\right)d_k^i\right\|\}.
\end{aligned}
\tag{2.45}
$$

注意到

$$
\frac{1}{c_{k+i}} \le \frac{1}{m^2\|d_{k+i}\|^2\|d_k^i\|^2}
$$

利用2.25,2.23,2.35, 便得所需要的结果2.44. □

定理 2.4. 设*2.2*的条件满足, 那么,

$$
d_k \to 0, \tag{2.46}
$$

$$
\left\|\alpha_{kr+i}d_{kr+i}-\alpha_{kr}^i d_{kr}^i\right\| = O\left(\|d_{kr}\|^2\right), i=0,1,\cdots,j(k)-1, \tag{2.47}
$$

$$
\left\|\alpha_{kr+i}d_{kr+i}-\alpha_{kr}^i d_{kr}^i\right\| = O\left(\|x_{kr}-x^*\|^2\right), i=0,1,\cdots,j(k)-1, \tag{2.48}
$$

其中, 对于所有k和$r \ge n, d_{kr}=-g_{kr}$.

证明. 由2.32,$\|d_{k+1}\| \le (1+M/m)^2\|d_k\|$, 故

$$
\|d_{kr+i}\| \le \left[(1+M/m)^2\right]^i\|d_{kr}\|, \quad i=0,1,\cdots,j(k)-1.
$$

由于算法收敛, 2.31意味着$g_k \to 0$,因此,$d_{kr} \to 0$,由2.32,有$d_k \to 0$,2.46得到. 为了证明2.47, 我们用归纳法证明: 对于 $i=0,1,\cdots$, $j(k/r)-1$, 以及 k 是 r 的倍数, 有

$$
\left\|g_{k+i}-g_k^i\right\| = O\left(\|d_k\|^2\right), \tag{2.49}
$$

$$\left\|d_{k+i}-d_k^i\right\|=O\left(\|d_k\|^2\right), \tag{2.50}$$

$$\left\|\alpha_{k+i}d_{k+i}-\alpha_k^i d_k^i\right\|=O\left(\|d_k\|^2\right) \tag{2.51}$$

首先, 对于$i=0$, 由于$g_k=g_k^0, d_k=d_k^0$和2.44, 可知2.49-2.51成立. 今假定2.49-2.51对于i成立, 我们证明它们对于$i+1$也成立.由2.43和归纳法假设,

$$\begin{aligned}\left\|g_{k+i+1}-g_k^{i+1}\right\| &\leq \left\|g_{k+i}-g_k^i\right\|+O\left(\|d_k\|^2\right)\\ &+M\left\|\alpha_{k+i}d_{k+i}-\alpha_k^i d_k^i\right\|\\ &=O\left(\|d_k\|^2\right).\end{aligned} \tag{2.52}$$

由2.42,对于FR公式,

$$\begin{aligned}\left\|d_{k+i+1}-d_k^{i+1}\right\| =&O_1\left(\left\|d_{k+i}-d_k^i\right\|\right)+O_2\left(\left\|g_{k+i+1}-g_k^{i+1}\right\|\right)\\ &+O_3\left(\left\|g_{k+i}-g_k^i\right\|\right)+O_4\left(\|d_k\|^2\right).\end{aligned}$$

这样, 由归纳法假设, 我们有

$$\left\|d_{k+i+1}-d_k^{i+1}\right\|=O\left(\|d_k\|^2\right). \tag{2.53}$$

最后,利用2.44,2.52和归纳法假设,有

$$\begin{aligned}&\left\|\alpha_{k+i+1}d_{k+i+1}-\alpha_k^{i+1}d_k^{i+1}\right\|\\ &=O_1\left(\left\|g_{k+i+1}-g_k^{i+1}\right\|\right)+O_2\left(\left\|d_{k+i+1}-d_k^{i+1}\right\|\right)+O_3\left(\|d_k\|^2\right)\\ &=O\left(\|d_k\|^2\right).\end{aligned} \tag{2.54}$$

这样2.49-2.51证得,从而,2.47得到.由2.47,并注意到 $d_{kr}=-g_{kr}$和$\|g_{kr}\|\leq M\|x_{kr}-x^*\|$, 即可得到2.48.而这等价于2.9. 综合定理 4.3.8 和定理 4.3.17, 我们便得到再开始共轭梯度法的 n 步二阶收敛定理 4.3.7. □

3 信赖域

3.1 信赖域算法

信赖域子问题

$$\min_{d\in\mathbb{R}^n} m_k(d)=f(x_k)+\nabla f(x_k)^T d+\frac{1}{2}d^T B_k d, \quad \text{s.t.} \quad \|d\|\leq\Delta_k. \tag{3.1}$$

Algorithm 1 信赖域算法

给定最大半径Δ_{max}，初始半径Δ_0，初始点$x_0, k \leftarrow 0.$

给定参数$0 \le \eta < \bar{\rho}_1 < \bar{\rho}_2 < 1, \gamma_1 < 1 < \gamma_2.$

while 未达到收敛准则 **do**

  计算信赖域子问题3.1得到迭代方向$d_k.$

  根据$\rho_k = \frac{f(x_k)-f(x_k+d_k)}{m_k(0)-m_k(d_k)}$计算下降率$\rho_k$

  if $\rho_k < \bar{\rho}_1$ **then** 缩小信赖域半径：$\Delta_{k+1} = \gamma_1 \Delta_k.$

  else

    if $\rho_k > \bar{\rho}_2$以及$\|d_k\| = \Delta_k$ **then**

      扩大信赖域半径：$\Delta_{k+1} = \min\{\gamma_2 \Delta_k, \Delta_{max}\}.$

    else

      信赖域半径不变：$\Delta_{k+1} = \Delta_k.$

  if $\rho_k > \eta$ **then**

    更新$x_{k+1} = x_k + d_k.$

  else

    $x_{k+1} = x_k.$

  $k \rightarrow k+1.$

在此章节，我们取$\bar{\rho_1} = \frac{1}{4}$，$\bar{\rho_2} = \frac{3}{4}$，$\gamma_1 = \frac{1}{4}$，$\gamma_2 = 2$

3.2 信赖域算法收敛性

现在介绍信赖域算法的全局收敛性．回顾信赖域算法6.8，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．

3.2.1 信赖域算法全局收敛性

在$\eta = 0$的条件下有如下收敛性定理：

定理 3.1 (信赖域算法全局收敛性1). 设近似海瑟矩阵B_k有界，即$\|B_k\|_2 \le M, \forall k$，$f(x)$在下水平集$\mathcal{L} = \{x|f(x) \le f(x_0)\}$上有下界，且$\nabla f(x)$在$\mathcal{L}$的一个开邻域内$Lip$连续.若$d_k$为信赖域子问题的近似解且满足

$$m_k(0) - m_k(d_k) \ge c_1 \|\nabla f(x_k)\| \min\{\Delta_k, \frac{\|\nabla f(x_k)\|}{\|B_k\|_2}\}, \tag{3.2}$$

其中$c_1 \in (0, 1]$. 算法*3.1*选取参数$\eta = 0$，则

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0, \tag{3.3}$$

即x_k的聚点中包含稳定点.

证明. 通过对比ρ_k公式可以得到

$$\begin{aligned} |\rho_k - 1| &= \left| \frac{(f(x_k) - f(x_k + d_k)) - (m_k(0) - m_k(d_k))}{m_k(0) - m_k(d_k)} \right| \\ &= \left| \frac{m_k(d_k) - f(x_k + d_k)}{m_k(0) - m_k(d_k)} \right| \end{aligned}$$

由Taylor公式可以得到

$$f(x_k + d_k) = f(x_k) + \nabla f(x_k)^T d_k + \int_0^1 [\nabla f(x_k + td_k) - \nabla f(x_k)]^T d_k dt,$$

存在某个$t \in (0, 1)$，使得

$$\begin{aligned} |m_k(d_k) - f(x_k + d_k)| &= \left| \frac{1}{2} d_k^T B_k d_k - \int_0^1 [\nabla f(x_k + td_k) - \nabla f(x_k)]^T d_k dt \right| \\ &\le (M/2)\|d_k\|^2 + L\|d_k\|^2, \end{aligned} \tag{3.4}$$

其中，L是∇f在集合$S(R_0) = \{x | \|x - y\| < R_0, y \in S\}$的Lipschitz常数，假设$\|d_k\| \le R_0$.

反证法：假设存在$\epsilon > 0, K \in \mathbb{N}^*$使得

$$\|\nabla f(x_k)\| \ge \epsilon, \quad \forall k \ge K. \tag{3.5}$$

因此，对$k \ge K$，有

$$m_k(0) - m_k(d_k) \ge c_1 \|\nabla f(x_k)\| \min\left(\Delta_k, \frac{\|\nabla f(x_k)\|}{\|B_k\|}\right) \ge c_1 \epsilon \min\left(\Delta_k, \frac{\epsilon}{M}\right) \tag{3.6}$$

对于$\gamma \geq 1$，我们有

$$\|d_k\| \leq \gamma \Delta_k, \tag{3.7}$$

由3.7,3.4,3.6，我们有

$$|\rho_k - 1| \leq \frac{\gamma^2 \Delta_k^2 (M/2 + L)}{c_1 \epsilon \min(\Delta_k, \epsilon/M)}. \tag{3.8}$$

为了给出当k足够大时Δ的一个上界$\bar{\Delta}$，令

$$\bar{\Delta} = \min\left(\frac{1}{2}\frac{c_1 \epsilon}{\gamma^2 (M/2 + L)}, \frac{R_0}{\gamma}\right) \tag{3.9}$$

由于$c_1 \leq 1, \gamma \geq 1$，我们有$\bar{\Delta} \leq \epsilon/M$，并且$\Delta_k \in [0, \bar{\Delta}], \forall k$，因此$\min(\Delta_k, \epsilon/M) = \Delta_k$。

$$|\rho_k - 1| \leq \frac{\gamma^2 \Delta_k^2 (M/2 + L)}{c_1 \epsilon \Delta_k} = \frac{\gamma^2 \Delta_k (M/2 + L)}{c_1 \epsilon} \leq \frac{\gamma^2 \bar{\Delta} (M/2 + L)}{c_1 \epsilon} \leq \frac{1}{2}. \tag{3.10}$$

因此，$\rho_k > \frac{1}{4}$。由3.1可知，$\Delta_{k+1} \leq \Delta_k$仅当$\Delta_k \leq \bar{\Delta}$成立。因此可以得到

$$\Delta_k \geq \min(\Delta_K, \bar{\Delta}/4) \quad \forall k \geq K. \tag{3.11}$$

假设存在一个无限子列$\mathcal{K}$使得$\rho_k \geq \frac{1}{4}, \forall k \in \mathcal{K}$。对$k \in \mathcal{K}, k \geq K$我们由3.6有

$$\begin{aligned} f(x_k) - f(x_{k+1}) &= f(x_k) - f(x_k + d_k) \\ &\geq \frac{1}{4}[m_k(0) - m_k(d_k)] \\ &\geq \frac{1}{4} c_1 \epsilon \min(\Delta_k, \epsilon/\beta). \end{aligned}$$

由于f是下有界的，则会有

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

这与3.11矛盾因此不存在这样的子列，所以必有3.3成立。 □

定理 3.2 (信赖域算法全局收敛性2). *在3.1的条件下，若3.1选取参数$\eta > 0$，且信赖域子问题近似解d_k满足3.2，则*

$$\lim_{k \to \infty} \|\nabla f(x_k)\| = 0. \tag{3.12}$$

证明. 取正整数m使得$\nabla f(x_k) \neq 0$。令β_1为∇f在集合$S(R_0)$的Lipschitz常数，则我们有

$$\|\nabla f(x) - \nabla f(x_m)\| \leq \beta_1 \|x - x_m\|, \quad \forall x \in S(R_0)$$

定义
$$\epsilon = \frac{1}{2}\|\nabla f(x_m)\|, \quad R = \min\left(\frac{\epsilon}{\beta_1}, R_0\right).$$
令球
$$\mathcal{B}(x_m, R) = \{x | \|x - x_m\| \leq R\}$$
包含在$S(R_0)$，因此我们有
$$x \in \mathcal{B}(x_m, R) \Rightarrow \|\nabla f(x)\| \geq \|\nabla f(x_m)\| - \|\nabla f(x) - \nabla f(x_m)\| \geq \frac{1}{2}\|\nabla f(x_m)\| = \epsilon.$$
若$\{x_k\}_{k\geq m}$均在球$\mathcal{B}(x_m.R)$内，我们有$\|\nabla f(x_k)\| \geq \epsilon > 0, \forall k \geq m$.而这种情况不可能发生，因此序列$\{x_k\}_{k\geq m}$会离开$\mathcal{B}(x_m, R)$.
取$l \geq m$，使得x_{l+1}是第一个迭代到球$\mathcal{B}(x_m, R)$外的.由于$\|\nabla f(x_k)\| \geq \epsilon, k = m, m+1, \ldots, l$,根据3.6可知
$$\begin{aligned} f(x_m) - f(x_{l+1}) &= \sum_{k=m}^{l} f(x_k) - f(x_{k+1}) \\ &\geq \sum_{k=m, x_k \neq x_{k+1}}^{l} \eta[m_k(0) - m_k(d_k)] \\ &\geq \sum_{k=m, x_k \neq x_{k+1}}^{l} \eta c_1 \epsilon \min\left(\Delta_k, \frac{\epsilon}{M}\right) \end{aligned}$$
若$\Delta_k \leq \epsilon/M, k = m, m+1, \ldots, l$，则
$$f(x_m) - f(x_{l+1}) \geq \eta c_1 \epsilon \sum_{k=m, x_k \neq x_{k+1}}^{l} \Delta_k \geq \eta c_1 \epsilon R = \eta c_1 \epsilon \min\left(\frac{\epsilon}{\beta_1}, R_0\right). \tag{3.13}$$
否则，我们有$\Delta_k > \epsilon/\beta, k = m, m+1, \ldots, l$，且
$$f(x_m) - f(x_{l+1}) \geq \eta c_1 \epsilon \frac{\epsilon}{M}. \tag{3.14}$$
由于序列$\{f(x_k)\}_{k=0}^{\infty}$是单调递减且有下界，我们有
$$f(x_k) \downarrow f^* \tag{3.15}$$
因此由3.13和3.14，我们可以得到
$$\begin{aligned} f(x_m) - f^* &\geq f(x_m) - f(x_{l+1}) \\ &\geq \eta c_1 \epsilon \min\left(\frac{\epsilon}{M}, \frac{\epsilon}{\beta_1}, R_0\right) \\ &= \frac{1}{2}\eta c_1 \|\nabla f(x_m)\| \min\left(\frac{\|\nabla f(x_m)\|}{2M}, \frac{\|\nabla f(x_m)\|}{2\beta_1}, R_0\right) > 0. \end{aligned}$$
由于$f(x_m) - f^* \downarrow 0$，我们由$\nabla f(x_m) \to 0$. □

3.3 信赖域算法的局部收敛性与超线性收敛速度

定理 3.3. 设f时二阶可微，海瑟矩阵$\nabla^2 f(x)$在最优解x^*附近$Lipschitz$连续，对于迭代$x_{k+1} = x_k + d_k$，其中$d_k^N = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$，那么

*(1)*若起始点x_0充分靠近x^*，则迭代序列收敛至x^*

(2)$\{x_k\}$具有二阶收敛速度

(3)$\{\|\nabla f(x_k)\|\}$二阶收敛至*0*

定理 3.4. 设$f(x)$在最优点$x = x^*$的一个邻域内二阶连续可微，且$\nabla f(x) Lip$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^*，且在迭代中选取B_k为海瑟矩阵$\nabla^2 f(x_k)$，则对充分大的k，任意满足*3.2*的信赖域子问题算法产生的迭代方向d_k均满足

$$\|d_k - d_k^N\| = o(\|d_k^N\|), \tag{3.16}$$

其中d_k^N为第k步迭代的牛顿方向且满足假设$\|d_k^N\| \le \frac{\Delta_k}{2}$ 进一步地，信赖域算法产生的迭代序列$\{x_k\}$具有超线性收敛速度。

证明. 若对足够大的k有$\|d_k^N\| \le \frac{1}{2}\Delta_k, \|d_k\| \le \Delta_k$，那么在最优点附近的迭代步肯定满足3.16。

对足够大的k找到$m_k(0) - m_k(d_k)$的一个下界。假设k足够大是得$\|d_k^N\| \le o(\|d_k^N\|)$。当$\|d_k\| \le \|d_k^N\| + o(\|d_k^N\|) \le 2\|d_k^N\|$，而若$\|d_k^N\| > \frac{1}{2}\Delta_k$，那么我们有$\|d_k\| \le \Delta_k < 2\|d_k^N\|$。在两种情况下，我们都有

$$\|d_k\| \le 2\|d_k^N\| \le 2 \left\|\nabla^2 f(x_k)^{-1}\right\| \|\nabla f(x_k)\|$$

因此$\|\nabla f(x_k)\| \ge \frac{1}{2}\|d_k\| / \|\nabla^2 f(x_k)^{-1}\|$。

由3.2可知

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

由于$x_k \to x^*$，利用$\nabla^2 f(x)$的连续性与$\nabla^2 f(x^*)$的正定性可以得到当k足够大时，

$$\frac{c_1}{4\|\nabla^2 f(x_k)^{-1}\|^2\|\nabla^2 f(x_k)\|}\geq\frac{c_1}{8\|\nabla^2 f(x_k)^{-1}\|^2\|\nabla^2 f(x^*)\|}=c_3$$

其中$c_3>0$，因此对足够大的k我们有

$$m_k(0)-m_k(d_k)\geq c_3\|d_k\|^2. \tag{3.17}$$

由$\nabla^2 f(x)$在x^*附近的Lipschitz连续性可得

$$\begin{aligned}
&|(f(x_k)-f(x_k+d_k))-(m_k(0)-m_k(d_k))|\\
&=|\frac{1}{2}{d_k}^T\nabla^2 f(x_k)d_k-\frac{1}{2}\int_0^1 {d_k}^T\nabla^2 f(x_k+td_k)d_k dt\\
&\leq\frac{L}{4}\|d_k\|^3.
\end{aligned}$$

其中$L>0$是$\nabla^2 f(\cdot)$Lip常数。因此由ρ_k的定义我们可以得到对足够大的k，有

$$|\rho_k-1|\leq\frac{\|d_k\|^3(L/4)}{c_3\|d_k\|^2}=\frac{L}{4c_3}\|d_k\|\leq\frac{L}{4c_3}\Delta_k. \tag{3.18}$$

若信赖域半径下降则仅当$\rho_k<\frac{1}{4}$，那么由3.18可知$\{\Delta_k\}$是远离0的。由于$x_k\to x^*$，我们有$\|d_k^N\|\to 0$，所以由3.16可知$\|d_k\|\to 0$.因此当k足够大时信赖域半径的约束终将失效，且算法产生的迭代方向将会越来越接近牛顿方向，那么$\|d_k^N\|\leq\Delta_k$总能被满足。

为了证明超线性收敛性，我们利用3.3，那么我们有

$$\|x_k+d_k^N-x^*\|=o(\|x_k-x^*\|^2),$$

因此$\|d_k^N\|=\mathcal{O}(\|x_k-x^*\|)$.

由3.16可知

$$
\begin{aligned}
&\|x_k + d_k - x^*\| \\
&\le \|x_k + d_k^N - x^*\| + \|d_k^N - d_k\| = o(\|x_k - x^*\|^2) + o(\|d_k^N\|) = o(\|x_k - x^*\|),
\end{aligned}
$$

这说明信赖域算法是超线性收敛的。 □