谨以此文纪念冯康教授 (1920-1993)

信赖域方法的收敛性 [*1)]

袁亚湘

(中国科学院计算中心)

ON THE CONVERGENCE OF TRUST REGION ALGORITHMS

Yuan Ya-xiang

(*Computing Center, Academia Sinica*)

Abstract

Trust region algorithms for nonlinear optimization and their convergence properties are discussed. Convergence results and techniques for convergence analysis are studied. An A (δ, η) descent trial step is defined, and is used to obtain a unified proof for global convergence of trust region algorithms

1. 信赖域方法

信赖域方法是非线性优化的一类重要的数值计算方法. 它在近二十年来受到了非线性优化研究界非常的重视. 特别是最近几年，一直是非线性优化的研究热点. 目前，信赖域方法已经和传统的线收索方法并列为非线性规划的两类主要数值方法.

信赖域方法的研究起始于 Powell[32]. 但是，人们发现信赖域方法的基本技巧在一定意义下等价于十分著名的求解非线性最小二乘的 Levenberg-Marquadt 方法. 对于非线性最小二乘问题：

$$\min_{x\in\Re^n} \|F(x)\|_2^2, \tag{1.1}$$

其中 $F(x)=(f_1(x),\cdots,f_m(x))^T$ 和 $f_i(x)(i=1,\cdots,m)$ 是 $\Re^n$ 中的连续可微的函数，Gauss-Newton 方法是

$$x_{k+1}=x_k+d_k=x_k-A(x_k)^+F(x_k), \tag{1.2}$$

其中 $A(x)=\nabla F(x)^T$ 是 Jacobi 矩阵，A^+ 表示 A 的广义逆矩阵. 不难看出，Gauss-Newton 步 $d_k=-A(x_k)^+F(x_k)$ 是问题

$$\min_{d\in\Re^n} \|F(x_k)+A(x_k)d\|_2^2 \tag{1.3}$$

* 1993 年 10 月 12 日收到.

1) 由国家攀登项目“大规模科学与工程计算”资助.

的解. 问题 (1.3) 显然是原问题 (1.1) 在当前迭代点 x_k 的近似. 当 Jacobi 矩阵 $A(x_k)$ 几乎奇异时, Gauss-Newton 步 $d_k=-A(x_k)^+F(x_k)$ 很可能会非常长, 从而使 Gauss-Newton 方法出现数值困难. 为了克服这一困难, 由 Levenberg[21] 提出并由 Marquardt[23] 重新发现的 Levenberg-Marquardt 方法利用如下步:

$$d_k=-(A(x_k)^TA(x_k)+\lambda_kI)^{-1}A(x_k)^TF(x_k), \tag{1.4}$$

其中 $\lambda_k\geq 0$ 是一个参数, 它每次迭代修正 [25]. Levenberg-Marquardt 方法的思想就是通过引入参数 λ_k 来克服 $A(x_k)$ 几乎奇异所带来的困难. 利用恰当的参数 λ_k 可保证矩阵 $(A(x_k)^TA(x_k)+\lambda_kI)$ 可逆而且能避免出现过于大的 $||d_k||_2$. 容易看到, Levenberg-Marquadt 步 (1.4) 是问题

$$\min_{d\in\Re^n}||F(x_k)+A(x_k)d||_2^2+\lambda_k||d||_2^2 \tag{1.5}$$

的解. 问题 (1.5) 是 (1.3) 的修改. 所加的一项 $\lambda_k||d||_2^2$ 可看作是罚项, 从而可阻止 $||d_k||$ 过大.

定义

$$\Delta_k=||(A(x_k)^TA(x_k)+\lambda_kI)^{-1}A(x_k)^TF(x_k),||_2\ , \tag{1.6}$$

则不难证明 Levenberg-Marquardt 步 (1.4) 也是如下问题:

$$\min_{d\in\Re^n}||F(x_k)+A(x_k)d||_2^2, \tag{1.7}$$

$$\text{subject to}\quad ||d||_2\leq\Delta_k \tag{1.8}$$

的解. 约束条件 (1.8) 正是一信赖域约束. 故显然可见, 问题 (1.7)-(1.8) 是一信赖域子问题. 因此, 从这个意义下传统的 Levenberg-Marquardt 方法也是一特殊的信赖域方法. 也正由于此, 人们提及信赖域方法的历史总是追溯到 Levenberg-Marquardt 方法.

Levenberg-Marquardt 方法每次迭代修改 λ_k, 从而隐含地起到调节 Δ_k, 限制步长 $||d_k||_2$ 的作用. 信赖域方法直接调节 Δ_k, 故能更好地控制 $||d_k||_2$ 的大小, 而且方法直观以及具有很强的逼近背景.

信赖域方法的构造和逼近是密切相关的. 给出一个当前迭代点 x_k, 构造一近似于原问题的逼近模型. 由于该模型主要是基于原问题在 x_k 的信息, 故有理由认为此模型仅在 x_k 附近可以很好的描述原问题. 所以人们仅在 x_k 附近的某一邻域内相信该模型. 信赖域方法的子问题都是在当前迭代点 x_k 附近的某一邻域内求逼近模型的最优点. 该邻域被称为信赖域, 它常是以 x_k 为中心的广义球. 信赖域的大小通过迭代逐步调节. 一般说来, 如果在当前迭代模型较好地逼近原问题, 则信赖域可扩大, 否则信赖域应缩小.

信赖域方法的关键组成部分是如何求得信赖域试探步以及怎样决定试探步是否可被接受. 试探步一般是一子问题的解, 所以如何求得信赖域试探步实质上归结于子问题的构造. 决定试探步是否可被接受通常是利用某一价值函数, 看试探步能否使价值函数

下降．对于无约束优化问题，价值函数显然就是目标函数．对于约束优化问题，价值函数常常是一罚函数．

和线搜索方法相比，信赖域方法还不够成熟．有效的信赖域方法实用软件，特别是对于求解约束优化问题的软件，依然十分缺少．所以，目前在实践应用中信赖域方法还没有线搜索方法那样广泛．但是，信赖域方法具有两个很好的性质．一是它有很好的可靠性和强适性 (robust)；另一是它有很强的收敛性．毫无疑问，信赖域方法将受到更大的重视，它的应用将更广．

下面介绍无约束优化和约束优化的信赖域方法以及它们的收敛性分析技巧．本文不介绍非光滑优化的信赖域方法，有关这方面的研究可见 [15], [42], [44] 和 [45] 等．

2. 无约束优化

无约束优化问题是在整个 $\Re^n$ 上求解函数的极值点，可写为

$$\min_{x\in\Re^n} f(x), \tag{2.1}$$

其中 $f(x)$ 是定义在 $\Re^n$ 上的函数．利用二次逼近，可构造信赖域子问题如下：

$$\min_{d\in\Re^n} g_k^T d + \frac{1}{2}d^T B_k d = \phi_k(d), \tag{2.2}$$

$$\text{subject to} \quad ||d||_2 \le \Delta_k, \tag{2.3}$$

其中 $g_k = \nabla f(x_k)$, B_k 是一 $n\times n$ 实对称阵， $\Delta_k > 0$ 是信赖域半径．

定理 2.1. 设 $g\in\Re^n$, $\Delta>0$ 和 $B\in\Re^{n\times n}$ 对称，则 $s\in\Re^n$ 是问题：

$$\min_{d\in\Re^n} g^T d + \frac{1}{2}d^T B d = \phi(d), \tag{2.4}$$

$$\text{subject to} \quad ||d||_2 \le \Delta \tag{2.5}$$

之解的充分必要条件是存在唯一的 $\lambda\ge 0$ 使得

$$(B+\lambda I)s = -g, \tag{2.6}$$

$$||s||_2 \le \Delta, \quad \lambda(\Delta - ||s||_2) = 0, \tag{2.7}$$

且 $B+\lambda I$ 是半正定矩阵．

关于此定理的证明可见 [26]. 基于这一定理，可将求解信赖域子问题化为计算满足 (2.6)-(2.7) 的 λ. λ 可利用 Newton 法计算，具体算法已由 [18] 和 [28] 给出．子问题的解具有很好的下降性质：

定理 2.2. 设 s 是问题 (2.4)-(2.5) 的解，则必有

$$\phi(0)-\phi(s) \ge \frac{1}{2}||g||_2 \min\left[\Delta, \frac{||g||_2}{1+||B||_2}\right], \tag{2.8}$$

$$-g^T s \ge \frac{1}{2}||g||_2 \min\left[\Delta, \frac{||g||_2}{1+||B||_2}\right]. \tag{2.9}$$

不等式 (2.8) 的证明最早由 Powell[32] 给出. 关于 (2.9) 的证明可见 [29]. 事实上，人们发现信赖域方法的收敛性主要是基于试探步满足不等式 (2.8), 所以，大多数信赖域方法并不精确求解子问题 (2.2)–(2.3), 而是计算一满足：

$$\phi_k(0)-\phi_k(s_k)\geq\delta||g_k||_2\min\left[\Delta_k,\frac{||g_k||_2}{1+||B_k||_2}\right] \tag{2.10}$$

的试探步 s_k, 其中 δ 是一正常数.

定义 2.1. 设 $\delta>0,\eta\geq 0$ 是两常数，如果 $s_k\in\Re^n$ 满足 (2.10) 和不等式

$$||s_k||_2\leq(1+\eta)\Delta_k, \tag{2.11}$$

则称 s_k 是子问题 (2.2)–(2.3) 的 (δ,η) 下降试探步.

显然，子问题 (2.2)–(2.3) 的精确解是 $(0.5,0)$ 下降试探步. 事实上，对任何子空间 $S_k\subset\Re^n$, 只要 $g_k\in S_k$, 则子问题 (2.2)–(2.3) 在 S_k 上的解也是 $(0.5,0)$ 下降试探步. 为了子问题容易求解且使算法具有较好的收敛性， S_k 可为 $\mathrm{Span}(g_k,B_k^{-1}g)$. 在 Dennis 和 Mei[8], Powell[31], Shultz, Schnabel 和 Byrd[35] 以及 Thomas[39] 等的方法都是在子空间上非精确求解 (2.2)–(2.3). 子问题 (2.2)–(2.3) 也可通过预条件共轭梯度法非精确求解，例如 [38].

对任一试探步 s_k, 我们称模型函数的下降量

$$Pred_k=\phi_k(0)-\phi_k(s_k) \tag{2.12}$$

为预估下降量，称目标函数的下降量

$$Ared_k=f(x_k)-Axrf(x_k+s_k) \tag{2.13}$$

为真实下降量. 由 (2.10) 可知, 如果 $||g_k||_2\neq 0$ 而且 s_k 是一 (δ,η) 下降试探步, 则 $Pred_k>0$. 于是可定义比值：

$$r_k=\frac{Ared_k}{Pred_k}. \tag{2.14}$$

这一比值在决定是否接受试探步以及如何调节信赖域半径方面起关键作用.

下面给出的是一个一般性的无约束优化信赖域方法：

算法 2.2.

步 1. 给出 $x_1\in\Re^n$, $\Delta_1>0$, $\varepsilon\geq 0$,

$$0<c_3<c_4<1<c_1,\ 0\leq c_0\leq c_2<1,\ c_2>0,\ k:=1.$$

步 2. 如果 $||g_k||_2\leq\varepsilon$ 则停；求 (2.2)–(2.3) 的 (δ,η) 下降试探步 s_k.

步 3. 计算 r_k;

$$x_{k+1}=\begin{cases}x_k, & \text{如果 } r_k\leq c_0,\\ x_k+s_k, & \text{否则}.\end{cases} \tag{2.15}$$

选取 Δ_{k+1} 满足

$$\Delta_{k+1} \in \begin{cases} [c_3\|s_k\|_2,\ c_4\Delta_k], & \text{如果 } r_k < c_2, \\ [\Delta_k,\ c_1\Delta_k], & \text{否则}. \end{cases} \tag{2.16}$$

步 4. 修正 B_{k+1}; $k := k+1$; 转步 2.

在算法中， $c_i\ (i=0,\cdots,4)$ 是由用户选择的常数. 常用的数值是 $c_0 = 0, c_1 = 2, c_2 = c_3 = 0.25, c_4 = 0.5$. 其它选取可参阅 [15], [17], [26] 和 [33] 等. 常数 c_0 一般是零 (如 [13], [32]) 或是很小的正常数 (如 [10], [37]). 取 $c_0 = 0$ 的优点是任何使目标函数下降的试探步都会被接受，从而保证了算法不会放弃任一已计算过函数值的 "好点". 这一点当目标函数的值很难求得时尤为重要. 不足之处是当 $c_0 = 0$ 时收敛性结果较弱，即仅能证明：

$$\liminf_{k\to\infty} \|g_k\|_2 = 0. \tag{2.17}$$

而在 $c_0 > 0$ 的假定下，可证明：

$$\lim_{k\to\infty} \|g_k\|_2 = 0. \tag{2.18}$$

当然，对于任意给定的 $\varepsilon > 0$, 弱收敛性结果 (2.17) 也能使算法经过有限次迭代后满足终止性条件 $\|g_k\|_2 \le \varepsilon$.

下面给出的是算法 2.2 的全局收敛性结果：

定理 2.3. 设函数 $f(x)$ 在 $\Re^n$ 上连续可微，如果 $(1+\eta)c_4 < 1$ 以及

$$\sum_{k=1}^{\infty} \frac{1}{M_k} = \infty\ , \tag{2.19}$$

其中

$$M_k = \max_{1\le i\le k} \|B_i\|_2 + 1\ , \tag{2.20}$$

则对任何给定的 $\varepsilon > 0$ 算法 2.2 必定经过有限次迭代后终止或者产生点列 x_k 使得 $f(x_k)$ 趋于 $-\infty$.

证明. 定义集合

$$S = \{k | r_k < c_2\}. \tag{2.21}$$

根据算法可知，

$$\Delta_{k+1} \le c_4\Delta_k, \quad \forall k \in S, \tag{2.22}$$

以及

$$f(x_k) - f(x_{k+1}) \ \ge\ c_2 Pred_k \tag{2.23}$$

$$\ge\ c_2\delta\|g_k\|_2 \min\left[\Delta_k, \frac{\|g_k\|_2}{1+\|B_k\|_2}\right], \quad \forall k \notin S. \tag{2.24}$$

如果定理非真，则对一切正整数 k 都有

$$\|g_k\|_k \geq \varepsilon \tag{2.25}$$

而且 $f(x_k)$ 下方有界. 于是

$$\infty > \sum_{k\notin S}[f(x_k)-f(x_{k+1})]/c_2\delta\varepsilon \geq \sum_{k\notin S}\min\left[\Delta_k, \frac{\varepsilon}{1+\|B_k\|_2}\right]. \tag{2.26}$$

由于

$$Pred_k - Ared_k = o(\|s_k\|) + O(\|s_k\|^2\|B_k\|) \tag{2.27}$$

以及集合 S 的定义可证

$$\frac{1}{1+\|B_k\|_2} = O(\|s_k\|_2) = O(\Delta_k), \quad k\in S. \tag{2.28}$$

从 M_k 的单调性和 Δ_k 所满足的 (2.16) 可得

$$\frac{1}{M_k} = O(\Delta_k), \quad \forall k. \tag{2.29}$$

利用 (2.26) 即知

$$\sum_{k\notin S}\frac{1}{M_k} < \infty. \tag{2.30}$$

根据 Powell[33] 的一个著名的引理，由 M_k 的单调性，(2.29), (2.30) 和 Δ_k 所满足的 (2.16) 可推出

$$\sum_{k=1}^{\infty}\frac{1}{M_k} < \infty. \tag{2.31}$$

上式与定理的假定 (2.19) 相矛盾，所以定理为真.

定理 2.4. 设函数 $f(x)$ 在 $\Re^n$ 上连续可微，如果 $c_0>0$, $\|B_k\|_2\leq M$ 对一切 k 成立，则对于 $\epsilon=0$ 算法 2.2 必定经过有限次迭代后终止或者产生点列 x_k 使得

$$\lim_{k\to\infty} f(x_k) = -\infty, \tag{2.32}$$

$$\lim_{k\to\infty}\|g_k\|_2 = 0 \tag{2.33}$$

中必有一个成立.

证明. 假定算法不能有限终止且 (2.33) 不成立，则存在正常数 $\bar{\varepsilon}>0$ 以及无穷多个 k 使得 $\|g_k\|_2\geq\bar{\varepsilon}$. 定义集合 $T=\{k|x_{k+1}\neq x_k, \|g_k\|_2\geq\bar{\varepsilon}/2\}$, 由于 $c_0>0$, 存在正常数 $\bar{\delta}$, 使得

$$\sum_{k\in T}[f(x_k)-f(x_{k+1})] \geq \sum_{k\in T}\bar{\delta}\min[\Delta_k, 1]. \tag{2.34}$$

如果定理非真，则 (2.32) 也不成立，故而

$$\sum_{k\in T}\Delta_k < \infty. \tag{2.35}$$

由上式即知

$$\sum_{k=1}^{\infty} \Delta_k < \infty \tag{2.36}$$

和

$$\lim ||g_k||_2 \geq \bar{\varepsilon}. \tag{2.37}$$

从 (2.36) 和 (2.37) 可证

$$\lim_{k\to\infty} r_k = 1. \tag{2.38}$$

上面极限表明，当 k 充分大时，有 $\Delta_{k+1} \geq \Delta_k$. 这与 (2.36) 矛盾，由此说明定理成立.

定理 2.5. 假定算法 2.2 产生的点列 x_k 收敛于 x^*, $\nabla^2 f(x)$ 在 x^* 附近连续而且 $\nabla^2 f(x^*)$ 正定，假设 s_k 不仅是一 (δ, η) 下降步而且满足

$$\phi_k(\alpha s_k) \geq \phi_k(s_k), \quad \forall \alpha \in [0, 1], \tag{2.39}$$

$$s_k = -B_k^{-1} g_k, \text{ 若 } ||s_k||_2 < \Delta_k, \tag{2.40}$$

如果还有

$$\lim_{k\to\infty} ||g(x_k + s_k) - g_k - B_k s_k||_2 / ||s_k||_2 = 0, \tag{2.41}$$

则 x_k 超线性收敛于 x^*.

证明. 从 (2.41) 得到

$$s_k^T (g(x_k + s_k) - g_k - B_k s_k) = o(||s_k||_2^2), \tag{2.42}$$

故知

$$s_k^T B_k s_k - s_k^T \nabla^2 f(x^*) s_k = o(||s_k||_2^2), \tag{2.43}$$

$$Ared_k - Pred_k = o(||s_k||_2^2). \tag{2.44}$$

由于 s_k 满足 (2.39),

$$Pred_k \geq \frac{1}{2} s_k^T B_k s_k \geq \frac{1}{2} s_k^T \nabla^2 f(x^*) s_k + o(||s_k||_2^2), \tag{2.45}$$

所以 $r_k \to 1$. 因此对所有充分大的 k 都有 $||s_k||_2 < \Delta_k$, 故有 $B_k s_k - g_k = 0$. 将其代入 (2.41) 即得

$$\lim ||g(x_{k+1})||_2 / ||s_k||_2 = 0. \tag{2.46}$$

从上式可证

$$\lim ||x_{k+1} - x^*||_2 / ||x_k - x^*||_2 = 0. \tag{2.47}$$

故知定理为真.

关于如何修正 B_k 可保证 (2.41), 可见 [32] 和 [20].

如果 $B_k = \nabla^2 f(x_k)$, 则定理 2.5 的结论可加强为二阶收敛. 这方面的工作可见 [17], [35], [36] 以及 [38].

3. 约束优化

对于约束优化问题:

$$\min_{x\in\Re^n} f(x), \tag{3.1}$$

$$\begin{aligned} \text{s.t.}\quad c_i(x) &= 0, \qquad i=1,2,\cdots,m_e; &(3.2)\\ c_i(x) &\geq 0, \qquad i=m_e+1,\cdots,m, &(3.3)\end{aligned}$$

信赖域方法的构造主要取决于试探步 s_k 的选取和如何判别 s_k 是否可被接受. 与无约束优化一样, s_k 的选取归结为子问题的构造. 约束优化的信赖域子问题有好几种. 第一种具有如下形式:

$$\min_{d\in\Re^n} g_k^T d+\frac{1}{2}d^T B_k d=\phi_k(d), \tag{3.4}$$

$$\begin{aligned} \text{s.t.}\quad \theta c_i(x_k)+d^T\nabla c_i(x_k) &= 0, \qquad i=1,2,\cdots,m_e; &(3.5)\\ \theta c_i(x_k)+d^T\nabla c_i(x_k) &\geq 0, \qquad i=m_e+1,\cdots,m; &(3.6)\\ ||d||_2 &\leq \Delta_k, &(3.7)\end{aligned}$$

其中 $\theta\in(0,1]$. 为了使 θ 尽可能靠近 1, 可在 (3.4) 的目标函数中加上一罚项 $\sigma(\theta-1)^2$. 利用 (3.4)-(3.7) 的方法, 有 Byrd, Schnabel 和 Shultz[4] 以及 Vardi[41] 等. 子问题 (3.4)-(3.6) 是二次规划问题, 可由已有方法求解.

另一类子问题是

$$\min_{d\in\Re^n} g_k^T d+\frac{1}{2}d^T B_k d=\phi_k(d), \tag{3.8}$$

$$\text{s.t.}\ ||(c_k+A_k^T d)^-||_2\leq\xi_k, \tag{3.9}$$

$$||d||_2\leq\Delta_k, \tag{3.10}$$

其中 $c_k=c(x_k)=(c_1(x),\cdots,c_m(x))^T$, $A_k=A(x_k)=\nabla c(x_k)^T$, $\xi_k\geq 0$ 是参数, 上标 "$-$" 表示

$$v_i^-=v_i,\quad i=1,\cdots,m_e \tag{3.11}$$

$$v_i^-=\min[0,v_i],\quad i=m_e+1,\cdots,m. \tag{3.12}$$

利用 (3.8)-(3.10) 的方法, 有 Celis, Denis 和 Tapia[5] 以及 Powell 和 Yuan[34]. Yuan[46] 讨论了子问题 (3.8)-(3.10) 的性质, 还给出了求解该子问题的一个对偶算法 [47]. 这个子问题还可化为一个单参数估计问题, 见 [50].

子问题还可利用罚函数导出, 基于非光滑精确罚函数的子问题是

$$\min g_k^T d+\frac{1}{2}d^T B_k d+\sigma_k||(c_k+A_k^T d)^-||=\Phi_k(d), \tag{3.13}$$

$$\text{s.t.}\quad ||d||\leq\Delta_k. \tag{3.14}$$

Fletcher[14] 取 $\|\cdot\| = \|\cdot\|_1$, Yuan[48] 考虑了 $\|\cdot\| = \|\cdot\|_\infty$ 而且给出了如何逐步修改 σ_k 的技巧.

计算试探步也可利用约束零空间以及积极集合等技巧，如 [1], [2], [27] 和 [30] 等.

判别 s_k 是否被接受将依赖于某一价值函数 $P_k(x)$, 它常是一精确罚函数.

约束优化信赖域方法一般需要构造一个近似价值函数 $\bar{\phi}_k(d)$, 使其在 $\|d\|$ 很小时满足

$$\bar{\phi}_k(d) - \bar{\phi}_k(0) = P_k(x_k + d) - P_k(x_k) + o(\|d\|), \tag{3.15}$$

而且有

$$pred_k = \bar{\phi}_k(0) - \bar{\phi}_k(s_k) \geq \bar{\delta}\varepsilon_k \min[\Delta_k, \epsilon_k/\|B_k\|], \tag{3.16}$$

其中 $\bar{\delta}$ 是正常数， ε_k 是 KT 条件的违反度：

$$\varepsilon_k = \|c_k^-\| + \|g_k - A_k\lambda_k\|, \tag{3.17}$$

λ_k 满足

$$(\lambda_k)_i \geq 0, \quad i > m_e, \tag{3.18}$$

是在当前迭代点 x_k 对 Langrange 乘子的估计.

证明信赖域方法收敛的基本方法是先证明 $\bar{\phi}_k(d)$ 满足 (3.15)–(3.16), 然后证明价值函数对充分大的 k 保持不变，即 $P_k(x) = P(x)$. 如果 ε_k 的任何子列均不趋于零，则在 $\|B_k\|$ 一致有界的假设下可推出

$$pred_k \geq \hat{\delta}\Delta_k, \tag{3.19}$$

其中 $\hat{\delta}$ 是正常数. 进而利用 $P(x)$ 下方有界可以证明

$$\sum_{k=1}^{\infty} \Delta_k < \infty. \tag{3.20}$$

于是 $\Delta_k \to 0$, 这一极限和关系式 (3.15) 可导出

$$r_k = \frac{P(x_k) - P(x_k + s_k)}{pred_k} \to 1. \tag{3.21}$$

从而 $\Delta_{k+1} \geq \Delta_k$, 这与 (3.20) 矛盾. 所以 $\{\varepsilon_k\}$ 必有子列收敛于零. 故知下面所给出的模型方法在上述假定成立下是收敛的：

算法 3.3. 步 1. 给出 $x_1 \in \Re^n$, $\Delta_1 > 0$, $\epsilon \geq 0$,

$$0 < c_3 < c_4 < 1 < c_1,\ 0 \leq c_0 \leq c_2 < 1,\ c_2 > 0,\ k := 1.$$

步 2. 如果 $\varepsilon_k \leq \varepsilon$ 则停；求试探步 s_k 使的 (3.16) 成立.

步 3. 计算 r_k; 根据 (2.15) 选取 x_{k+1}; 选取 Δ_{k+1} 满足 (2.16).

步 4. 构造 $\bar{\phi}_{k+1}(d)$; $k := k + 1$; 转步 2.

我们上面给出的是证明约束优化信赖域方法总体收敛的基本技巧，对于具体算法的收敛性证明的详细过程，可见 [3], [11], [30] 和 [34] 等.

关于信赖域方法的局部收敛性分析，一般都是基于 SQP 方法的超线性收敛性. 记 d_k^* 是在第 k 次迭代的 SQP 步，在一些假定下可证

$$\lim_{k\to\infty} ||x_k + d_k^* - x^*||/||x_k - x^*|| = 0. \tag{3.22}$$

所以要证明算法 (3.3) 产生的点列 Q - 超线性收敛

$$\lim_{k\to\infty} ||x_{k+1} - x^*||/||x_k - x^*|| = 0, \tag{3.23}$$

就等价于证明

$$||s_k - d_k^*|| = o(||d_k^*||), \tag{3.24}$$

$$x_{k+1} = x_k + s_k \tag{3.25}$$

对所有充分大的 k 都成立. 为使算法具有 (3.24), 产生试探步 s_k 的子问题必须较好地近似 SQP 子问题. (3.25) 则要求价值函数 $F_k(x)$ 在 $x_k + s_k$ 处的值比在 x_k 处小.

很多约束优化信赖域方法在 k 充分大时只要 $||s_k|| < \Delta_k$ 就有

$$s_k = d_k^*. \tag{3.26}$$

因而要证明超线性收敛性只需证对所有充分大的 k 试探步 s_k 是在信赖域的内部而且是可被接受的. 不幸的是，这并不是对所有方法都成立. 如果价值函数非光滑，则超线性收敛步 s_k 有可能不被接受 (见 [22] 和 [43]).

为了使超线性收敛步都能被接受的做法主要有两种：其一是引入二阶校正步 $\hat{s}_k$ 使得 $||\hat{s}_k|| = O(||s_k||^2)$ 且 $s_k + \hat{s}_k$ 是可接受的；其二是利用光滑罚函数作为价值函数就可保证在解的附近所有超线性收敛步都能被接受.

二阶校正步的技巧最早由 Fletcher[16] 提出并用于非光滑优化，它的超线性收敛性由 Yuan[45] 给出. 对于约束优化的线搜索方法，二阶校正步由 Coleman 和 Conn[6] 以及 Mayne 和 Polak[24] 分别讨论. Byrd, Schnabel 和 Shultz[4] 给出的约束优化信赖域方法利用了二阶校正步从而保证了算法的超线性收敛性. 二阶校正步一般可通过求解一原子问题的修正问题来得到，例如假设 s_k 是由 (3.13)–(3.14) 得到的，则计算二阶校正步 $\hat{s}_k$ 可求如下问题：

$$\min\ g_k^T(s_k + d) + \frac{1}{2}(s_k + d)^T B_k (s_k + d) + \sigma_k ||(c(x_k + s_k) + A_k^T(d))^-||_\infty, \tag{3.27}$$

$$\text{s.t.} \quad ||s_k + d|| \le \Delta_k. \tag{3.28}$$

光滑精确罚函数使得在解 x^* 处

$$P(x) = P(x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 P(x^*)(x - x^*) + o(||x - x^*||^2), \tag{3.29}$$

故当 s_k 是超线性收敛步时必有

$$\begin{aligned} P(x_k+s_k) &= P(x^*)+O(||x_k+s_k-x^*||^2)=P(x^*)+o(||x_k-x^*||^2) \\ &< P(x^*)+\frac{1}{4}(x_k-x^*)^T\nabla^2P(x^*)(x_k-x^*) \\ &< P(x_k). \end{aligned} \tag{3.30}$$

所以 s_k 可被接受. Powell 和 Yuan[34] 最先利用光滑精确罚函数作为价值函数构造信赖域方法，他们的方法在每次迭代的 $P_k(x)$ 是 Fletcher(1975) 光滑精确罚函数. 他们在一定假定下证明了当 k 充分大时 $P_k(x)$ 将保持不变，从而建立起算法的超线性收敛结果.

当约束条件具有特殊形式时，信赖域方法常常可特别地构造，这方面的讨论可见于 [3], [7] 和 [19] 等.

参考文献

[1] J.V. Burke, A robust trust region method for constrained nonlinear programming problems, Report MCS-P131-0190, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois, USA, 1990.

[2] J.V. Burke, J.J. Moré, On the identification of active constraints, *SIAM J. Numer. Anal.*, 25(1988), 1197-1211.

[3] J.V. Burke, J.J. Moré, G. Toraldo, Convergence properties of trust region methods for linear and convex constraints, *Math. Prog.*, 47(1990), 305–336.

[4] R. Byrd, R.B. Schnabel, G.A. Shultz, A trust region algorithm for nonlinearly constrained optimization, *SIAM Journal of Numer. Anal.*, 24(1987), 1152–1170.

[5] M.R. Celis, J.E. Dennis, R.A. Tapia, A trust region algorithm for nonlinear equality constrained optimization, in: P.T. Boggs, R.H. Byrd and R.B. Schnabel, eds., *Numerical Optimization* (SIAM, Philadelphia, 1985), 71–82.

[6] T.F. Coleman, A.R. Conn, Nonlinear programming via an exact penalty function: asymptotic analysis, *Math. Prog.*, 24(1982), 123–136.

[7] A.R. Conn, N.I.M. Gould, Ph.L. Toint, Global convergence of a class of trust region algorithms for optimization with simple bounds, *SIAM J. Numer. Anal.*, 25(1988), 433–460.

[8] J.E. Dennis, H.H.W. Mei, Two new unconstrained optimization algorithms which use function and gradient values, *J. Opt. Theory and Applns.*, 28(1979), 453–482.

[9] J.E. Dennis, J.J. Moré. A characterization of superlinear convergence and its application to quasi-Newton methods, *Math. Comp.*, 28 (1974), 549–560.

[10] I.S. Duff, J. Nocedal, J.K. Reid, The use of linear programming for the solution of sparse sets of nonlinear equations, *SIAM J. Sci. Stat. Comput.*, 8(1987), 99–108.

[11] M. El-Alem, A global convergence theory for the Celis-Dennis-Tapia trust region algorithm for constrained optimization, Report, 88–10, Math. Sciences Dept., Rice University, Texas, USA, 1988.

[12] M. El-Hallabi, R.A. Tapia, A global convergence theory for arbitrary norm trust-region methods for nonlinear equations, Technical Report, 87–25, Math. Sciences Dept., Rice University, Texas, USA, 1987.

[13] R. Fletcher, Practical Methods of Optimization, Vol. 1, Unconstrained Optimization, John Wiley and Sons, Chichester, 1980.

[14] R. Fletcher, Practical Methods of Optimization, Vol. 2, Constrained Optimization, John Wiley and Sons, Chichester, 1981.

[15] R. Fletcher, A model algorithm for composite NDO problem, *Math. Prog. Study,* 17(1982), 67–76. (1982a)

[16] R. Fletcher, Second order correction for nondifferentiable optimization, in: G.A. Watson, ed., Numerical Analysis, Springer-Verlag, Berlin, 1982, 85–115. (1982b)

[17] R. Fletcher, Practical Methods of Optimization (second edition), John Wiley and Sons, Chichester, 1987.

[18] D.M. Gay, Computing optimal local constrained step, *SIAM J. Sci. Stat. Comp.,* 2(1981) 186–197.

[19] D.M. Gay, A trust region approach to linearly constrained optimization, in: D.F. Griffiths, ed., Lecture Notes in Mathematics 1066: Numerical Analysis, Springer-Verlag, Berlin, 1984, 72–105.

[20] H.F.H. Khalfan, Topics in quasi-Newton methods for unconstrained optimization. PhD thesis, University of Colorado, 1989.

[21] K. Levenberg, A method for the solution of certain nonlinear problems in least squares, *Qart. Appl. Math.,* 2(1944), 164–166.

[22] N. Marotos, Exact Penalty Function Algorithms for Finite Dimensional and Control Optimization Problems, Ph. D. thesis, Imperial College Sci. Tech., University of London, 1978.

[23] D.W. Marquardt, An algorithm for least-squares estimation of nonlinear inequalities, *SIAM J. Appl. Math.,* 11(1963), 431–441.

[24] D.Q. Mayne, E. Polak, A superlinearly convergent algorithm for constrained optimization problems, *Math. Prog. Study,* 16(1982), 45–61.

[25] J.J. Moré, The Levenberg-Marquardt algorithm: implementation and theory, in: G.A. Watson, ed., Lecture Notes in Mathematics 630: Numerical Analysis, Springer-Verlag, Berlin, 1978, 105–116.

[26] J.J. Moré, Recent developments in algorithms and software for trust region methods, in: A. Bachem, M. Grötschel and B. Korte, eds., Math. Prog.: The State of the Art, Springer-Verlag, Berlin, 1983, 258–287.

[27] J.J. Moré, Trust region and projected gradients, Report ANL/MCS-TM-107, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois, USA, 1988.

[28] J.J. Moré, D.C. Sorensen, Computing a trust region step, *SIAM J. Sci. Stat. Comp.,* 4(1983), 553–572.

[29] J. Nocedal, Y. Yuan, Combining trust region and line search techniques, Technical Report, NAM06, Dept of Computer Science, Northwestern University, Illinois, USA, 1991.

[30] E. O. Omojokun, Trust Region Algorithms for Optimization with Nonlinear Equality and Inequality Constraints, Ph. D. Thesis, University of Colorado at Boulder, 1989.

[31] M.J.D. Powell, A new algorithm for unconstrained optimization, in: J.B. Rosen, O.L. Mangasarian and K. Ritter, eds., Nonlinear Programming, Academic Press, New York, 1970, 31–66.

[32] M.J.D. Powell, Convergence properties of a class of minimization algorithms, in: O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds., Nonlinear Programming 2, Academic Press, New York, 1975, 1–27.

[33] M.J.D. Powell, On the global convergence of trust region algorithms for unconstrained optimization, *Math. Prog.,* 29(1984), 297-303.

[34] M.J.D. Powell, Y. Yuan, A trust region algorithm for equality constrained optimization, *Math. Prog.,* 49(1991), 189–211.

[35] G.A. Shultz, R.B. Schnabel, R.H. Byrd, A family of trust-region-based algorithms for unconstrained minimization with strong global convergence, *SIAM J. Numer. Anal.,* 22(1985), 47–67.

[36] D.C. Sorensen, Newton's method with a model trust region modification, *SIAM J. Numer. Anal.,* 20(1982), 409–426. (1982a)

[37] D.C. Sorensen, Trust regin methods for unconstrained optimization, in: M.J.D. Powell, ed., Nonlinear Optimization 1981, Academic Press, London, 1982, 29–38. (1982b)

[38] T. Steihaug, The conjugate gradient method and trust regions inlarge scale optimization, *SIAM J. Numer. Anal.,* 20(1983), 626–637.

[39] S.W. Thomas, Sequential estimation techniques for quasi-Newton algorithms, Report, 75-227, Dept of Computer Science, Cornell University, Ithaca, NY, USA, 1975.

[40] Ph.L. Toint, Global convergence of a class of trust region methods for nonconvex minimization in Hilbert space, *IMA J. Numer. Anal.,* 8(1988), 231–252.

[41] A. Vardi, A trust region algorithm for equality constrained minimization: convergence properties and implementation, *SIAM J. Numer. Anal.,* 22(1985), 575–591.

[42] R. S. Womersley, Local properties of algorithms for minimizing nonsmooth composite functions, *Math. Prog.*, 32(1985), 69–89.

[43] Y. Yuan, An example of only linearly convergence of trust region algorithms for nonsmooth optimization, *IMA J. Numer. Anal.,* 4(1984), 327–335.

[44] Y. Yuan, Conditions for convergence of trust region algorithms for nonsmooth optimization, *Math. Prog.*, 31(1985), 220–228. (1985a)

[45] Y. Yuan, On the superlinear convergence of a trust region algorithm for nonsmooth optimization, *Math. Prog.*, 31(1985), 269–285. (1985b)

[46] Y. Yuan, On a subproblem of trust region algorithms for constrained optimization, *Math. Prog.*, 47(1990), 53–63.

[47] Y. Yuan, A dual algorithm for minimizing a quadratic function with two quadratic constraints, *J. Comput. Math.*, 9(1991), 348–359.

[48] Y. Yuan, A new trust region algorithm for nonlinear optimization in: D. Bainov and V. Covachev, eds. Proc. First Int. Col. Numer. Anal., VSP, Zeist, 1993, 141–152.

[49] Y. Yuan, Local convergence and numerical results of a trust region algorithm, Report 435, Institute für Angew. Math., University of Würzburg, Germany 1993.

[50] Y. Zhang, Computing a Celis-Dennis-Tapia trust region step equality constrained optimization, *Math. Prog.*, 55(1992), 109–124.