第 34 卷 第 4 期　应 用 数 学 学 报　Vol. 34 No. 4
2011 年 7 月　ACTA MATHEMATICAE APPLICATAE SINICA　July, 2011

不等式约束优化的非单调可行信赖域-SQP 算法*

孙中波

(东北师范大学人文学院数学系，长春 130117)

(E-mail: szb21971@yahoo.com.cn)

段复建

(桂林电子科技大学数学与计算科学学院，桂林 541004)

摘　要　本文讨论不等式约束优化问题，给出一个信赖域方法与 SQP 方法相结合的新的可行算法，算法中采用了 "压缩技术", 使得 QP 子问题产生的搜索方向尽可能为可行方向，并且采用了高阶校正的方法来克服算法产生的 Maratos 效应现象．在适当的条件下，证明了算法的全局收敛性和超线性收敛性．数值结果表明算法是有效的．

关键词　非单调线搜索；信赖域算法；SQP 算法；全局收敛性；超线性收敛性

MR(2000) 主题分类　90C30

中图分类　O224.2

1　引言

考虑如下不等式约束优化问题

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t.}\ g_j(x)\le 0, \qquad j=1,2,\cdots,m, \end{aligned} \tag{1.1}$$

其中 $f(x), g_j(x): R^n\to R^1\ (j=1,2,\cdots,m)$ 是一个连续可微函数．Levenberg 和 Marguart 首次提出无约束优化问题的信赖域算法，由于这类算法具有很好的收敛性质和很强的鲁棒性，一直受到许多学者重视．随后，又有许多学者把这种无约束优化的信赖域方法推广到约束优化问题中，使得这类算法得到了迅速发展 [1−7]．针对约束优化问题，许多学者提出了不同的有效算法，由于 SQP 算法具有超线性收敛的优点，因此，多年以来

本文 2010 年 4 月 24 日收到．2010 年 12 月 20 日收到修改稿.

* 国家自然科学基金 (10861005), 广西自然科学基金 (0728206, 0991238) 资助项目.

都是一个十分活跃的领域. 近年来, SQP 算法得到了迅速的发展 [8−16], 使 SQP 算法理论日臻完善. 本文吸取了信赖域方法, SQP 方法, 以及非单调线搜索的思想, 研究和建立了非线性不等式约束优化问题的一个可行信赖域算法, 算法只在可行条件处采用了压缩技术, 不对平衡条件进行处理, 使 QP 子问题生成的搜索方向尽可能为可行方向, 如果条件不满足, 就进行非单调线搜索, 产生下一次迭代点, 不需要重新求解 QP 子问题. 在适当的条件下, 证明了算法的全局收敛性和超线性收敛性. 数值实验说明算法是有效的. 符号 $\|\cdot\|$ 是 Euclidean 范数.

2 信赖域 -SQP 算法

首先引入一些记号:

$\nabla f(x)$ 为目标函数 $f(x)$ 的梯度, $\nabla g_j(x)$ 为约束函数 $g_j(x)$ 的梯度, Δ_k 为信赖域半径, B_k 为对称正定矩阵, $\widehat{\theta}_{k_j}$ 为压缩因子, d_0^k, $\widetilde{d}_k, \widehat{d}_k$ 分别为下降方向, 可行方向, 校正方向, λ 为 Lagrange 乘子, $\digamma=\{x|g_j(x)\le 0\}$ 为可行集. $\Omega=\{j|g_j(x^*)+\nabla g_j(x^*)^T d_0^*=0\}$, $\Omega_k=\{j|g_j(x^k)+\nabla g_j(x^k)^T d_0^k=0\}$ 为积极约束集, $e=(1,\cdots,1)$, r_k 实际下降量与预优下降量的比率.

非单调信赖域 -SQP 算法

步骤 1: 初始化, $x^0\in\digamma$ 为任意初始点, $\varepsilon>0$, $0<\eta_1\le\eta_2<1$, $0<\theta_1<1<\theta_2$, $i=1$, M 是一个常数, $\nu\in\left(0,\frac{1}{2}\right)$, $\tau\in(2,3)$, $\widehat{\theta}_0\in(0,1)$, $\alpha_0\in(0,1)$, $\vartheta\in(0,1)$, $B_0=I$ 为单位阵, 令 $m(0)=0$, $k:=1$;

步骤 2: 计算 QP 子问题得到下降方向 d_0^k,

$$\begin{aligned}&\min\ \nabla f(x^k)^T d+\frac{1}{2}d^T B_k d\\&\text{s.t. } g_j(x^k)+\nabla g_j(x^k)^T d\le 0,\qquad j=1,2,\cdots,m,\\&\qquad \|d\|\le\Delta_k.\end{aligned}\tag{2.1}$$

令 (d_0^k,λ_0^k) 为 KKT 点对. 如果 $\|d_0^k\|=0$, 停. 若 $\|d_0^k\|\ne 0\le\Delta_k$, 则令 $d_1^k=d_0^k$, 转入步骤 4, 否则进入步骤 3;

步骤 3: 通过计算如下的方程组得到计算组合可行下降方向 d_1^k,

$$\begin{cases}B_k d+\nabla g_j(x^k)\lambda=-\nabla f(x^k),\\ \widehat{\theta}_{k_j}\nabla g_j(x^k)^T d+\lambda_j g_j(x^k)=\nabla g_j(x^k)^T d_0^k-\|d_0^k\|^\delta,\qquad j=1,2,\cdots,m,\\ \|d\|\le\Delta_k.\end{cases}\tag{2.2}$$

设其解为 $(\widetilde{d}^k,\widetilde{\lambda}_1^k)$, 令 $d_1^k=(1-\rho_k)d_0^k+\rho_k\widetilde{d}^k$, 其中 ρ_k 定义如下:

$$\rho_k=\begin{cases}1, & \text{如果 } \nabla f(x^k)^T\widetilde{d}^k\le\vartheta\nabla f(x^k)^T d_0^k,\\ \dfrac{(1-\vartheta)\nabla f(x^k)^T d_0^k}{\nabla f(x^k)^T(d_0^k-\widetilde{d}^k)}, & \text{否则 } \nabla f(x^k)^T\widetilde{d}^k>\vartheta\nabla f(x^k)^T d_0^k.\end{cases}$$

如果 $\|d_1^k\|\le\Delta_k$, 则转入步骤 4, 否则进入步骤 5.

步骤 4: 信赖域试探搜索，计算预优下降量 Ared_k 和实际下降量 Pred_k, 判定是否接受试探步，

$$\text{Ared}_k = f(x^k + d_1^k) - f(x^k)$$

和

$$\text{Pred}_k = \nabla f(x^k)^T d_1^k + \frac{1}{2} d_1^{kT} B_k d_1^k,$$

比率

$$r_k = \frac{\text{Ared}_k}{\text{Pred}_k}.$$

如果 $r_k > \eta_1$ 和 $g_j(x^k + d_1^k) \le 0$ 成立，则试探步被接受，令 $x^{k+1} = x^k + d_1^k$ 和 $\widehat{d}^k = 0$ 转步 7; 否则转入步骤 5.

步骤 5: 通过求解 $\widetilde{QP}$ 得到高阶校正方向 $\widehat{d}^k$,

$$\begin{aligned} &\min\ \nabla f(x^k)^T d + \frac{1}{2} d^T B_k d \\ &\text{s.t.}\ \ g_j(x^k + d_1^k) + \nabla g_j(x^k)^T d = -\|d_1^k\|^\tau, \qquad j = 1, 2, \cdots, m, \\ &\qquad \|d\| \le \Delta_k. \end{aligned} \tag{2.3}$$

如果 $\widetilde{QP}$ 无解或者 $\|\widehat{d}^k\| > \|d_1^k\|$, 则令 $\widehat{d}^k = 0$, 转入步骤 6.

步骤 6: 非单调线搜索，计算步长 α_k, α_k 属于 $\{1, \frac{1}{2}, \frac{1}{4} \cdots\}$,

$$\begin{cases} f(x^k + \alpha d_1^k + \alpha^2 \widehat{d^k}) \le \max\limits_{1 \le j \le m(k)} f_{[k-j]}(x^k) + \nu\alpha \nabla f(x^k)^T d_1^k, \\ g_j(x^k + \alpha d_1^k + \alpha^2 \widehat{d^k}) \le 0, \qquad j = 1, 2, \cdots, m, \end{cases} \tag{2.4}$$

令 $x^{k+1} = x^k + \alpha_k d_1^k + \alpha_k^2 \widehat{d^k}$, 进入步骤 7.

步骤 7: 校正 $B_k, \Delta_k, \widehat{\theta}_{k_j}$,

校正 Δ_k

如果 $r_k \le \eta_1$, 则 $\Delta_{k+1} = \theta_1 \Delta_k$;

如果 $r_k \in (\eta_1, \eta_2)$, 则 $\Delta_{k+1} = \Delta_k$;

如果 $r_k \ge \eta_2$, 则 $\Delta_{k+1} = \theta_2 \Delta_k$;

校正矩阵 B_k

采用 [17] B_k 的校正方法;

校正压缩因子 $\widehat{\theta}_{k_j}$

$\widehat{\theta}_{k_j} = \min\limits_j \{\max\limits_j \{\lambda_{0_j}^k, \|d_0^k\|\}, \widehat{\theta}_0\}$;

步骤 8: $m(k) = \min\limits_j \{m(k-1) + 1, M\}$;

步骤 9: $k =: k + 1$, 返回步 2.

注 在不接受试探步 d_1^k 时，采用非单调线搜索技术，无需重解 QP 子问题，从而减少计算量．当 $M = 0$ 时，非单调线搜索转化成一般的单调线搜索．步骤 3 中，采用了压缩技术和凸组合策略，使下降方向尽可能的被压缩到信赖域区域中，如果压缩不到信

赖域中，采用非单调线搜索技术，避免重新求解子问题．第五部分的数值实验说明压缩因子的引入是有效的，提高了计算效率．

3 算法的可行性和全局收敛性分析

首先给出如下假设：

假设 3.1 可行集 F 非空，即 $F \neq \emptyset$.

假设 3.2 目标函数 $f(x)$ 和约束函数 $g_j(x)$, 其中 $j=1,2,\cdots,m$ 是连续可微的．

假设 3.3 对任意 $x \in F$, 向量组 $\{\nabla g_j(x),\ j=1,2,\cdots,m\}$ 线性无关，即 LICQ 条件成立．

假设 3.4 矩阵序列 $\{B_k\}$ 一致正定，即存在常数 $M>m>0$, 使得

$$m\|y\|^2 \le y^T B_k y \le M\|y\|^2, \qquad \forall\, k, y \in R^n.$$

通过上面叙述的假设条件，我们可以得到下面几个引理和定理．

引理 3.1 任意的向量 $x \in F$, 正定矩阵 $B_k \in R^{n\times n}$ 并且 $\widehat{\theta}_{k_j} > 0$, 则系统 (2.2) 非奇异．

证 考虑系统 (2.2) 的前两个等式，可以得到如下的关于 (d,λ) 的线性方程组

$$\begin{cases} B_k d + \nabla g_j(x^k)\lambda = -\nabla f(x^k), \\ \widehat{\theta}_{k_j}\nabla g_j(x^k)^T d + \lambda_j g_j(x^k) = \nabla g_j(x^k)^T d_0^k - \|d_0^k\|^\delta, \end{cases}$$

上式转化成矩阵形式，得到如下线性系统，即

$$\begin{pmatrix} B_k & \nabla g(x^k) \\ \widehat{\theta}_k \nabla g(x^k)^T & g(x^k) \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x^k) \\ (\nabla g_j(x^k)^T d_0^k - \|d_0^k\|^\delta)e \end{pmatrix},$$

其中 $g(x^k) = (g_1(x^k), g_2(x^k), \cdots, g_m(x^k))$ 和 $\nabla g(x^k) = (\nabla g_1(x^k), \nabla g_2(x^k), \cdots, \nabla g_m(x^k))$ 和 $\widehat{\theta}_k = (\widehat{\theta}_{k_1}, \widehat{\theta}_{k_2}, \cdots, \widehat{\theta}_{k_m})$. 因此我们只需证明下面的矩阵是非奇异的，即系数矩阵

$$\begin{pmatrix} B_k & \nabla g(x^k) \\ \widehat{\theta}_k \nabla g(x^k)^T & g(x^k) \end{pmatrix}$$

非奇异．接下来的证明类似于 [11, 引理 3.1], 从而结论成立．

引理 3.2 二次规划 QP 问题的解 d_0^k 及步骤 3 和步骤 5 产生的 d_1^k 和 $\widehat{d}^k$ 满足，

1. 如果 $d_0^k = 0$, 则 x^k 是原问题的一个 KKT 点．

2. 如果 $d_0^k \neq 0$, 则 $\nabla f(x^k)^T d_0^k < 0,\ \nabla f(x^k)^T d_1^k < 0,\ \nabla f(x^k)^T \widehat{d}^k < 0$ 和 $\nabla g_j(x^k)^T d_1^k < 0,\ \nabla g_j(x^k)^T d_0^k < 0,\ \nabla g_j(x^k)^T \widetilde{d}^k < 0, j \in \Omega$, 即 d_1^k 是原问题在 x^k 处的一个可行下降方向．

证 1. 由 KKT 条件的定义，如果 $d_0^k=0$, 则

$$\begin{cases} B_k d_0^k + \nabla f(x^k) + \rightarrow \sum\limits_j \lambda_j \nabla g_j(x^k) = 0, \\ g_j(x^k) + \nabla g_j(x^k)^T d_0^k \le 0, \\ \|d_0^k\| \le \Delta_k. \end{cases} \tag{3.1}$$

从而，我们有

$$\begin{cases} \nabla f(x^k) + \sum\limits_j \lambda_j \nabla g_j(x^k) = 0, \\ g_j(x^k) \le 0, \\ \|d_0^k\| \le \Delta_k. \end{cases} \tag{3.2}$$

因此，结论成立.

2. 如果 $d_0^k \neq 0$, 则由 QP 问题及 KKT 条件的定义知，

$$\begin{cases} \nabla f(x^k)^T d_0^k = -d_0^{kT} B_k d_0^k - \sum\limits_{j=1}^m \lambda_{0_j}^k \nabla g_j(x^k)^T d_0^k, \\ \nabla g_j(x^k)^T d_0^k \le -g_j(x^k) \longrightarrow 0, \qquad k \longrightarrow +\infty, \quad j \in \Omega, \end{cases} \tag{3.3}$$

故

$$\nabla f(x^k)^T d_0^k = -d_0^{kT} B_k d_0^k + \sum_{j=1}^m \lambda_{0_j}^k g_j(x^k) \le -d_0^{kT} B_k d_0^k < 0$$

和

$$\nabla f(x^k)^T d_1^k = (1-\rho_k)\nabla f(x^k)^T d_0^k + \rho_k \nabla f(x^k)^T \widetilde{d}^k.$$

根据 ρ_k 的取值不同，分两种情况进行讨论：

情况 1 当 $\rho_k = 1$ 时，有

$$\nabla f(x^k)^T d_1^k = \nabla f(x^k)^T \widetilde{d}^k \le \vartheta \nabla f(x^k)^T d_0^k < 0.$$

情况 2 当 $\rho_k = \frac{(1-\vartheta)\nabla f(x^k)^T d_0^k}{\nabla f(x^k)^T (d_0^k - \widetilde{d}^k)}$ 时，得

$$\begin{aligned} \nabla f(x^k)^T d_1^k =& \Big[1 - \frac{(1-\vartheta)\nabla f(x^k)^T d_0^k}{\nabla f(x^k)^T (d_0^k - \widetilde{d}^k)}\Big] \nabla f(x^k)^T d_0^k + \frac{(1-\vartheta)\nabla f(x^k)^T d_0^k}{\nabla f(x^k)^T (d_0^k - \widetilde{d}^k)} \nabla f(x^k)^T \widetilde{d}^k \\ =& \nabla f(x^k)^T d_0^k - \nabla f(x^k)^T d_0^k \frac{(1-\vartheta)\nabla f(x^k)^T d_0^k}{\nabla f(x^k)^T (d_0^k - \widetilde{d}^k)} + \frac{(1-\vartheta)\nabla f(x^k)^T d_0^k \nabla f(x^k)^T \widetilde{d}^k}{\nabla f(x^k)^T (d_0^k - \widetilde{d}^k)} \\ =& \vartheta \nabla f(x^k)^T d_0^k < 0. \end{aligned}$$

综上可得 $\nabla f(x^k)^T d_1^k < 0$.

如果 $j \in \Omega$ 时，则 $\widehat{\theta}_{k_j} \nabla g_j(x^k)^T \widetilde{d}^k = -\lambda_j g_j(x^k) + \nabla g_j(x^k)^T d_0^k - \|d_0^k\|^\delta$, 故有

$$\nabla g_j(x^k)^T \widetilde{d}^k = \frac{\nabla g_j(x^k)^T d_0^k - \|d_0^k\|^\delta}{\widehat{\theta}_{k_j}} < 0, \qquad k \rightarrow +\infty.$$

如果 $j\in\Omega$ 且 $\rho_k\in(0,1]$ 时，则

$$\nabla g_j(x^k)^T d_1^k=(1-\rho_k)\nabla g_j(x^k)^T d_0^k+\rho_k\nabla g_j(x^k)^T\widetilde{d}^k<0,\qquad k\longrightarrow+\infty.$$

下面证明 $\nabla f(x^k)^T\widehat{d}^k<0$.

由步骤 5 且当 $k\to+\infty$ 时，有

$$\begin{aligned}&\nabla f(x^k)^T\widehat{d}^k\\&=-\widehat{d}^{kT}B_k\widehat{d}^k-\nabla g_j(x^k)^T\widehat{d}^k\\&=-\widehat{d}^{kT}B_k\widehat{d}^k+\|d_1^k\|^\tau+g_j(x^k+d_1^k)\\&=-\widehat{d}^{kT}B_k\widehat{d}^k+g_j(x^k)+\nabla g_j(x^k)^T d_1^k+o(\|d_1^k\|)+\|d_1^k\|^\tau\ (\tau\in(2,3))\\&<0,\end{aligned}$$

从而引理 3.2 结论成立.

引理 3.3 对任意迭代指标 k, 存在某一个指标数 $j=j(k)$, 使 $\alpha_k=(\frac{1}{2})^j$ 成立.

证 由算法步骤 6 知,

$$\begin{aligned}s\triangleq&f(x^k+\alpha d_1^k+\alpha^2\widehat{d}^k)-\max_{0\le j\le m(k)}f_{[k-j]}(x^k)-\nu\alpha\nabla f(x^k)^T d_1^k\\\le&f(x^k+\alpha d_1^k+\alpha^2\widehat{d}^k)-f(x^k)-\nu\alpha\nabla f(x^k)^T d_1^k\\=&f(x^k+\alpha d_1^k)+\alpha^2\nabla f(x^k+\alpha d_1^k)^T\widehat{d}^k+o(\alpha^2)-f(x^k)-\nu\alpha\nabla f(x^k)^T d_1^k\\=&\alpha\nabla f(x^k)^T d_1^k+o(\alpha)+o(\alpha^2)+\alpha^2\nabla f(x^k+\alpha d_1^k)^T\widehat{d}^k-\nu\alpha\nabla f(x^k)^T d_1^k\\\le&(1-\nu)\alpha\nabla f(x^k)^T d_1^k+o(\alpha)+o(\alpha^2)\\=&(1-\nu)\alpha\nabla f(x^k)^T d_1^k+o(\alpha)\end{aligned}$$

成立. 当 α 充分小时，再由引理 3.2 知， $s\le0$.

由步骤 6 中的第二式知,

$$\begin{aligned}g_j(x^k+\alpha d_1^k+\alpha^2\widehat{d}^k)=&g_j(x^k+\alpha d_1^k)+\alpha^2\nabla g_j(x^k+\alpha d_1^k)^T\widehat{d}^k+o(\alpha^2)\\=&g_j(x^k)+\alpha\nabla g_j(x^k)^T d_1^k+o(\alpha)+o(\alpha^2),\end{aligned}$$

当 α 充分小时，由引理 3.2 知， $g_j(x^k+\alpha d_1^k+\alpha^2\widehat{d}^k)\le0$.

引理 3.1, 引理 3.2 和引理 3.3 说明了算法是良定的，可行的.

下面，将证明 $\{x^k\}$ 的任意聚点 x^* 一定是问题 (1.1) 的 KKT 点. 因为序列 $\{x^k\}$, $\{B_k\},\{d_0^k\},\{d_1^k\},\{\lambda_0^k\},\{\lambda_1^k\}$ 均有界，不妨设存在一个无穷子集 K_1 使得

$$x^k\to x^*,\quad B_k\to B_*,\quad d_0^k\to d_0^*,\quad d_1^k\to d_1^*,\quad \lambda_0^k\to\lambda_0^*,\quad \lambda_1^k\to\lambda_1^*,\quad k\in K_1.$$

证 见 [18].

引理 3.4 如果 $x^k\to x^*,\ B_k\to B_*,\ k\in K_1$, 则 $d_0^k\to0,\ d_1^k\to0,\ k\in K_1,\ k\to+\infty$.

证 由引理 3.2, 知算法产生的迭代序列 $\{f(x^k)\}$ 是下降序列.

下面根据算法的结构分两种情况讨论：

1. 当 x^{k+1} 为信赖域试探搜索接受的迭代点时，即迭代点 x^{k+1} 分别由步骤 2, 4 和步骤 2, 3, 4 生成.

如果迭代点 x^{k+1} 是由步骤 2, 4 生成，由于迭代序列 $\{f(x^k)\}$ 是下降序列，则必存在一个无穷子集 $\widehat{K}_1 \subset K_1$ $(|\widehat{K}_1| = +\infty)$, 使得 $x^{k+1} = x^k + d_1^k$, $k \in \widehat{K}_1$, 结合步骤 4 中 $\mathrm{Ared}_k = f(x^k + d_0^k) - f(x^k)$, 可以得到

$$
\begin{aligned}
0 &= \lim_{k\in\widehat{K}_1} \mathrm{Ared}_k = \lim_{k\in\widehat{K}_1} [f(x^k + d_0^k) - f(x^k)] \\
&= \lim_{k\in\widehat{K}_1} [f(x^k) + \nabla f(\widehat{\xi})^T d_0^k - f(x^k)] = \lim_{k\in\widehat{K}_1} \nabla f(\widehat{\xi})^T d_0^k \le 0,
\end{aligned}
$$

其中 $\widehat{\xi} \in (x^k, x^k + d_0^k)$, 所以 $d_0^k \to 0$, $k \in \widehat{K}_1$. 又 $d_0^k \to d_0^*$, $k \in \widehat{K}_1$, 故 $d_0^* = 0$, 即 $d_0^k \to 0, k \in K_1$.

如果代点 x^{k+1} 是由步骤 2, 3, 4 生成，证明方法同上述证明.

2. 当 x^{k+1} 不为信赖域试探搜索接受的迭代点时, 即迭代点 x^{k+1} 由步骤 2–6 生成, 也即是 $x^{k+1} = x^k + \alpha_k d_1^k + \alpha_k^2 \widehat{d}^k$. 迭代序列 $\{f(x^k)\}$ 采用了非单调线搜索技术，因此，由 $x^k \to x^*$, 知

$$
f(x^k) \to f(x^*), \qquad k \in K_1, \ \ k \longrightarrow +\infty.
$$

在 x^* 处，我们有

$$
\begin{aligned}
&\min \ \nabla f(x^*)^T d + \frac{1}{2} d^T B_* d \\
&\text{s.t.} \ \ g_j(x^*) + \nabla g_j(x^*)^T d \le 0, \qquad j = 1, 2, \cdots, m, \\
&\qquad \|d\| \le \Delta_k.
\end{aligned} \tag{3.4}
$$

显然， d_0^* 是 QP 问题的可行解，又由于 B_* 的正定性知 d_0^* 是唯一解，结合引理 3.1, 知 $\widetilde{d}^*$ 也是唯一解，所以 d_1^* 是唯一解. 步骤 5 中矩阵 $B_k \to B_*$ $(k \to +\infty)$ 是对称正定矩阵，因此 $\widehat{d}^*$ 是唯一解.

现在反设 $d_0^* \ne 0$, 由引理 3.2 和 $f(x)$ 与 $g_j(x)$ 的连续可微性知,

$$
\nabla f(x^*)^T d_0^* < 0, \qquad \nabla f(x^*)^T d_1^* < 0, \qquad \nabla g_j(x^*)^T d_1^* < 0, \qquad j \in \Omega.
$$

从而，在 x^* 处，存在步长 α_k 使 $\alpha_k \ge \alpha^* = \inf_k \{\alpha_k, k \in K_1\} > 0$ 恒成立，所以

$$
\begin{aligned}
0 = &\lim_{k\in K_1, k\to+\infty} (f(x^{k+1}) - \max_{0<j<m(k)} f_{[k-j]}(x^k)) \le \lim_{k\in K_1, k\to+\infty} \nu \alpha_k \nabla f(x^k)^T d_1^k \\
\le &\frac{1}{2} \alpha^* \nabla f(x^*)^T d_1^* < 0,
\end{aligned}
$$

因此 $d_1^k \to 0$, $k \in K_1$, $k \to +\infty$.

由算法的结构，考虑到引理 3.2, 引理 3.4 和 d_0^* 为 (3.4) 的解，可以得到如下的全局收敛性定理 3.1.

定理 3.1 假设条件 3.1–3.3 成立，则算法或者有限步终止于 QP 问题的 KKT 点，或者产生无穷点列，使得序列 $\{x^k\}$ 的每一个聚点 x^* 均满足 QP 问题的 KKT 点.

4 算法的超线性收敛率

假设 4.1 函数 $f(x)$ 和 $g_j(x),\ j=1,2,\cdots,m$ 二阶连续可微.

假设 4.2 在 KKT 点对 (x^*,λ^*) 处，二阶充分条件和严格互补条件成立，即

$$\lambda_j^*-g_j(x^*)>0,\qquad j=1,2,\cdots,m$$

和

$$d^T\nabla_{xx}^2L(x^*,\lambda^*)d>0,\qquad \forall\, 0\neq d\in\{d|\nabla g_j(x^*)^Td=0,\ j\in\Omega\}.$$

通过假设条件 4.1 和 4.2, 可以得到下面几个引理和定理.

引理 4.1 $\{x^k\}$ 整列收敛到 x^*, 即 $x^k\to x^*,\ k\to+\infty$.

证 一方面，由步骤 6 中的第一个式子，引理 3.2, 我们有

$$\begin{aligned}0=&\lim_{k\to+\infty}\Big(f(x^{k+1})-\max_{0<j<m(k)}f_{[k-j]}(x^k)\Big)\le\lim_{k\to+\infty}(f(x^{k+1})-f(x^k))\\ \le&\lim_{k\to+\infty}\nu\alpha_k\nabla f(x^k)^Td_1^k=\lim_{k\to+\infty}\nu(1-\rho_k)\alpha_k\nabla f(x^k)^Td_0^k+\lim_{k\to+\infty}\nu\rho_k\alpha_k\nabla f(x^k)^T\widetilde{d}^k\\ \le&\lim_{k\to+\infty}\frac{1}{2}\nu\alpha(1-\widehat{\rho})\nabla f(x^k)^Td_0^k+\lim_{k\to+\infty}\frac{1}{2}\nu\alpha\widehat{\rho}\nabla f(x^k)^T\widetilde{d}^k,\end{aligned}$$

根据 ρ_k 的不同取值，下面分两种情况讨论：

情况 1 当 $\rho_k=1$ 时，有

$$0\le\lim_{k\to+\infty}\nu\alpha_k\nabla f(x^k)^T\widetilde{d}^k\le-\frac{1}{2}\nu\alpha\omega\|\widetilde{d}^k\|<0.$$

情况 2 当 $\rho_k=\frac{(1-\vartheta)\nabla f(x^k)^Td_0^k}{\nabla f(x^k)^T(d_0^k-\widetilde{d}^k)}$ 时，有

$$\begin{aligned}0\le&\lim_{k\to+\infty}\frac{1}{2}\nu\alpha(1-\widehat{\rho})\nabla f(x^k)^Td_0^k+\lim_{k\to+\infty}\frac{1}{2}\nu\alpha\widehat{\rho}\nabla f(x^k)^T\widetilde{d}^k\\ \le&\lim_{k\to+\infty}\frac{1}{2}\nu\alpha(1-\widehat{\rho})\nabla f(x^k)^Td_0^k\\ \le&\lim_{k\to+\infty}-\frac{1}{2}\nu\alpha(1-\widehat{\rho})\omega\|d_0^k\|\\ <&0,\end{aligned}$$

其中 $\lim\limits_{k\to+\infty}\rho_k=\widehat{\rho}\in(0,1],\ \alpha=\inf\limits_k\{\alpha_k,\ k\in K\}>0,\ \|\nabla f(x)\|\le\omega$. 因此，$\alpha_k\nabla f(x^k)^Td_1^k\to 0,\ \alpha_k\nabla f(x^k)^Td_0^k\to0,\ \alpha_k\nabla f(x^k)^T\widetilde{d}^k\to0,\ \alpha_k\|d_0^k\|\to0,\ \alpha_k\|d_1\|\to0,\ k\to+\infty$ 恒成立.

另一方面，由步骤 3 知，

$$\nabla f(x^k)^T\widetilde{d}^k=-\widetilde{d}^{kT}B_k\widetilde{d}^k-\sum_{j=1}^m\lambda_j\nabla g_j(x^k)^T\widetilde{d}^k$$

$$
\begin{aligned}
&= -\widetilde{d}^{kT}B_k\widetilde{d}^k + \sum_{j=1}^{m}\frac{\lambda_j^2}{\widehat{\theta}_{k_j}}g_j(x^k) + \sum_{j=1}^{m}\frac{\lambda_j\|d_0^k\|^{\delta}}{\widehat{\theta}_{k_j}} - \sum_{j=1}^{m}\frac{\lambda_j\nabla g_j(x^k)^Td_0^k}{\widehat{\theta}_{k_j}}\\
&= -\widetilde{d}^{kT}B_k\widetilde{d}^k + \sum_{j=1}^{m}\frac{(\lambda_j+\lambda_j^2)}{\widehat{\theta}_{k_j}}g_j(x^k) - \sum_{j=1}^{m}\frac{\lambda_j[g_j(x^k)+\nabla g_j(x^k)^Td_0^k+o(\|d_0^k\|)]}{\widehat{\theta}_{k_j}}\\
&\le -m\|\widetilde{d}^k\|^2 + \sum_{j=1}^{m}\frac{(\lambda_j+\lambda_j^2)}{\widehat{\theta}_{k_j}}g_j(x^k) - \sum_{j=1}^{m}\frac{\lambda_j[g_j(x^k)+\nabla g_j(x^k)^Td_0^k+o(\|d_0^k\|)]}{\widehat{\theta}_{k_j}},
\end{aligned}
$$

从而,

$$
\lim_{k\to+\infty}\alpha_k\nabla f(x^k)^Td_1^k = \lim_{k\to+\infty}(1-\rho_k)\alpha_k\nabla f(x^k)^Td_0^k + \lim_{k\to+\infty}\rho_k\alpha_k\nabla f(x^k)^T\widetilde{d}^k.
$$

根据 ρ_k 的不同取值情况，分下面两种情况讨论：

情况 1 当 $\rho_k=1$ 时，

$$
\begin{aligned}
0 &= \lim_{k\to+\infty}\alpha_k\nabla f(x^k)^Td_1^k\\
&= \lim_{k\to+\infty}(1-\rho_k)\alpha_k\nabla f(x^k)^Td_0^k + \lim_{k\to+\infty}\rho_k\alpha_k\nabla f(x^k)^T\widetilde{d}^k\\
&\le \lim_{k\to+\infty}\alpha_k\nabla f(x^k)^T\widetilde{d}^k\\
&= \lim_{k\to+\infty}\alpha_k\Big[-\widetilde{d}^{kT}B_k\widetilde{d}^k + \frac{\lambda_j^2g_j(x^k)}{\widehat{\theta}_{k_j}} - \frac{\lambda_j\nabla g_j(x^k)^Td_0^k}{\widehat{\theta}_{k_j}} + \frac{\lambda_j\|d_0^k\|^{\delta}}{\widehat{\theta}_{k_j}}\Big]\\
&\le \lim_{k\to+\infty}\alpha_k\Big\{-m\|\widetilde{d}^k\|^2 + \sum_{j=1}^{m}\frac{(\lambda_j+\lambda_j^2)}{\widehat{\theta}_{k_j}}g_j(x^k) - \sum_{j=1}^{m}\frac{\lambda_j[g_j(x^k)+\nabla g_j(x^k)^Td_0^k+o(\|d_0^k\|)]}{\widehat{\theta}_{k_j}}\Big\}\\
&\le 0.
\end{aligned}
$$

情况 2 当 $\rho_k=\frac{(1-\vartheta)\nabla f(x^k)^Td_0^k}{\nabla f(x^k)^T(d_0^k-\widetilde{d}^k)}$ 时，我们有

$$
\begin{aligned}
0 &= \lim_{k\to+\infty}\alpha_k\nabla f(x^k)^Td_1^k\\
&= \lim_{k\to+\infty}(1-\rho_k)\alpha_k\nabla f(x^k)^Td_0^k + \lim_{k\to+\infty}\rho_k\alpha_k\nabla f(x^k)^T\widetilde{d}^k\\
&\le \lim_{k\to+\infty}\alpha_k\vartheta\nabla f(x^k)^Td_0^k < 0.
\end{aligned}
$$

由 $x^{k+1}=x^k+\alpha_kd_1^k+\alpha_k^2\widehat{d}^k$ 并且 $\|\widehat{d}^k\|\le\|d_1^k\|$ 有：

$$
\begin{aligned}
\|x^{k+1}-x^k\| &= \|\alpha_kd_1^k+\alpha_k^2\widehat{d}^k\| = \|\alpha_k(1-\rho_k)d_0^k+\alpha_k^2\rho_kd_1^k+\alpha_k^2\widehat{d}^k\|\\
&\le \alpha_k\|d_0^k\|+\alpha_k\|d_1^k\|+\alpha_k^2\|\widehat{d}^k\|\\
&\le \alpha_k\|d_0^k\|+\alpha_k\|d_1^k\|+2\alpha_k\|d_1^k\|\to 0,
\end{aligned}
$$

故知 $x^k\to x^*$, $k\to+\infty$.

引理 4.2 记集合 $\Omega_k=\{j|g_j(x^k)+\nabla g_j(x^k)^Td_0^k=0\}$, 则当 k 充分大时，有

$$
\Omega_k=\Omega,\quad \lim_{k\to\infty}d_0^k=0,\quad \lim_{k\to\infty}\lambda_0^k=\lambda_0^*,\quad \lim_{k\to\infty}\widetilde{d}^k=0,\quad \lim_{k\to\infty}\widetilde{\lambda}^k=\widetilde{\lambda}^*.
$$

证 见 [18].

引理 4.3 当 k 充分大时，有

$$\|d_0^k\| \sim \|\widetilde{d}^k\| \sim \|d_1^k\|,$$

而且

$$\|\widehat{d}^k\| = O(\|d_1^k\|).$$

证 根据算法分两种情况来考虑.

一方面，由算法的步骤 2, 4, 7 知， $\|d_0^k\| \sim \|d_1^k\|$ 显然.

另一方面，由算法的步骤 3, 4, 7 知，

$$g_j(x^k + d_0^k) = g_j(x^k) + \nabla g_j(x^k)^T d_0^k + O(\|d_0^k\|^2) = O(\|d_0^k\|^2), \qquad j \in \Omega,$$

并且，

$$\widehat{\theta}_{k_j} \nabla g_j(x^k)^T \widetilde{d}^k = -\|d_0^k\|^\delta - g_j(x^k + d_0^k) = O(\|d_0^k\|^2) = \|\widetilde{d}^k\|,$$

则有 $\|\widetilde{d}^k\| \sim \|d_0^k\|$. 又由于 $d_1^k = (1-\rho_k)d_0^k + \rho_k \widetilde{d}^k$, 从而， $\|d_0^k\| \sim \|\widetilde{d}^k\| \sim \|d_1^k\|$ 成立.

下面证明 $\|\widehat{d}^k\| = O(\|d_1^k\|)$. 类似于第二种情况，

$$\begin{aligned} g_j(x^k + d_1^k) =& g_j(x^k) + \nabla g_j(x^k)^T d_1^k + O(\|d_1^k\|^2) \\ =& g_j(x^k) + (1-\rho_k)\nabla g_j(x^k)^T d_0^k + \rho_k \nabla g_j(x^k)^T \widetilde{d}^k + O(\|d_1^k\|), \end{aligned}$$

从而，

$$\|g_j(x^k + d_1^k)\| = O(\|d_0^k\|) + O(\|d_1^k\|) + o(\|d_1^k\|^2) = O(\|d_1^k\|).$$

又 $\nabla g_j(x^k)^T \widehat{d}^k = -\|d_1^k\|^\tau - g_j(x^k + d_1^k)$, 其中 $\tau \in (2,3)$, 故有 $\|\widehat{d}^k\| = -\|d_1^k\|^\tau + O(\|d_1^k\|)$, 结论成立.

下面证明当 k 充分大时，步骤 6 中非单调线搜索中的步长恒为 1. 根据上述引理，再作如下假设.

假设 4.3 矩阵序列 $\{B_k\}$ 也满足

$$\|P_k(B_k - \nabla_{xx}^2 L(x^k, \lambda^k) d_1^k\| = o(\|d_1^k\|),$$

其中

$$\begin{aligned} &\nabla_{xx}^2 L(x^k, \lambda^k) = \nabla_{xx}^2 f(x^k) + \sum_{j \in \Omega} \lambda_j^k \nabla_{xx}^2 g_j(x^k), \\ &P_k = I_n - A_k(A_k^T A_k)^{-1} A_k^T, \qquad A_k = (\nabla g_j(x^k),\ j \in \Omega). \end{aligned}$$

引理 4.4 当 k 充分大时，步长恒为 1, 即 $\alpha_k \equiv 1,\ x^{k+1} = x^k + d_1^k + \widehat{d}^k$.

证 首先证明步长 $\alpha_k \equiv 1$ 能够使得算法是适定的. 分两种情况考虑

情况 1 当 $j \in \Omega_k \backslash \Omega$ 时，由于 $\widehat{d}^k \to 0,\ d_1^k \to 0,\ k \to \infty$, 且 $x^k \to x^*,\ k \to \infty$, 故有

$$\lim_{k\to\infty} g_j(x^k + d_1^k + \widehat{d}^k) = \lim_{k\to\infty} g_j(x^*) < 0,$$

从而 $\alpha_k \equiv 1$, 在这种情况下算法总是可行的.

情况 2 当 $j \in \Omega$ 时，在 $x^k + d_1^k$ 处对 $g_j(x^k + d_1^k + \widehat{d}^k)$ 利用泰勒展开有

$$\begin{aligned} g_j(x^k + d_1^k + \widehat{d}^k) =& g_j(x^k + d_1^k) + \nabla g_j(x^k + d_1^k)^T \widehat{d}^k + O(\|\widehat{d}^k\|^2) \\ =& g_j(x^k + d_1^k) + \nabla g_j(x^k)^T \widehat{d}^k + d_1^{kT} \nabla_{xx}^2 g_j(x^k) \widehat{d}^k + O(\|\widehat{d}^k\|^2) \\ =& g_j(x^k + d_1^k) + \nabla g_j(x^k)^T \widehat{d}^k + O(\|d_1^k\|^2), \end{aligned}$$

又由于算法中 $\widetilde{QP}$ 知,

$$\nabla g_j(x^k)^T \widehat{d}^k = -g_j(x^k + d_1^k) - \|d_1^k\|^\tau, \qquad j \in \Omega,$$

代入上式有

$$g_j(x^k + d_1^k + \widehat{d}^k) = -\|d_1^k\|^\tau + O(\|d_1^k\|^2).$$

因为 $\tau \in (2,3)$ 知， $g_j(x^k + d_1^k + \widehat{d}^k) \le 0$, $j \in \Omega$ 恒成立．从情况 1 和情况 2 知，算法在 $\alpha_k \equiv 1$ 使得算法是适定的.

其次，证明单位步长对目标函数 $f(x)$ 具有充分下降性.

$$\begin{aligned} s \triangleq& f(x^k + d_1^k + \widehat{d}^k) - \max_{0 \le j \le m(k)} f_{[k-j]}(x^k) - \nu \nabla f(x^k)^T d_1^k \\ \le& f(x^k + d_1^k + \widehat{d}^k) - f(x^k) - \nu \nabla f(x^k)^T d_1^k, \\ =& f(x^k) + \nabla f(x^k)^T (d_1^k + \widehat{d}^k) + \frac{1}{2}(d_1^k + \widehat{d}^k)^T \nabla_{xx}^2 f(x^k)(d_1^k + \widehat{d}^k) - f(x^k) \\ & - \nu \nabla f(x^k)^T d_1^k + o(\|d_1^k + \widehat{d}^k\|^2) \\ =& \nabla f(x^k)^T (d_1^k + \widehat{d}^k) + \frac{1}{2} d_1^{kT} \nabla_{xx}^2 f(x^k) d_1^k + \frac{1}{2} \widehat{d}^{kT} \nabla_{xx}^2 f(x^k) \widehat{d}^k \\ & - \nu \nabla f(x^k)^T d_1^k + d_1^{kT} \nabla_{xx}^2 f(x^k) \widehat{d}^k + o(\|d_1^k\|^2) \\ =& \nabla f(x^k)^T (d_1^k + \widehat{d}^k) + \frac{1}{2} d_1^{kT} \nabla_{xx}^2 f(x^k) d_1^k - \nu \nabla f(x^k)^T d_1^k + o(\|d_1^k\|^2). \end{aligned}$$

又由于

$$\begin{aligned} &\nabla f(x^k)^T d_1^k = (1-\rho_k) \nabla f(x^k)^T d_0^k + \rho_k \nabla f(x^k)^T \widetilde{d}^k \\ =& (1-\rho_k)\Big(- d_0^{kT} B_k d_0^k - \sum_{j=1}^m \lambda_{0_j}^k \nabla g_j(x^k)^T d_0^k \Big) + \rho_k \Big(- \widetilde{d}^{kT} B_k \widetilde{d}^k - \sum_{j=1}^m \lambda_{1_j}^k \nabla g_j(x^k)^T \widetilde{d}^k \Big). \end{aligned}$$

下面根据 ρ_k 的不同取值，分两种情况讨论：

情况 1 当 $\rho_k = 1$ 时，即 $\nabla f(x^k)^T \widetilde{d}^k \le \vartheta \nabla f(x^k)^T d_0^k$ 有

$$\begin{aligned} \nabla f(x^k)^T d_1^k =& - d_0^{kT} B_k d_0^k - \widetilde{d}^{kT} B_k \widetilde{d}^k - \sum_{j=1}^m \lambda_{0_j}^k \nabla g_j(x^k)^T d_0^k - \sum_{j=1}^m \lambda_{1_j}^k \nabla g_j(x^k)^T \widetilde{d}^k + o(\|d_1^k\|^2) \\ =& - d_1^{kT} B_k d_1^k + \sum_{j \in \Omega} \lambda_j^k g_j(x^k) + o(\|d_1^k\|^2), \qquad j \in \Omega. \end{aligned}$$

情况 2 当 $\rho_k=\frac{(1-\vartheta)\nabla f(x^k)^T d_0^k}{\nabla f(x^k)^T(d_0^k-\widetilde{d}^k)}$ 时，即 $\nabla f(x^k)^T\widetilde{d}^k>\vartheta\nabla f(x^k)^T d_0^k$, 有

$$\begin{aligned}&\nabla f(x^k)^T d_1^k\\ =&-d_0^{kT}B_k d_0^k-\widetilde{d}^{kT}B_k\widetilde{d}^k+\widetilde{d}^{kT}B_k\widetilde{d}^k-\sum_{j=1}^m\lambda_{0_j}^k\nabla g_j(x^k)^T d_0^k\\ &+\frac{(1-\vartheta)\nabla f(x^k)^T d_0^k}{\nabla f(x^k)^T(d_0^k-\widetilde{d}^k)}\Big[-d_0^{kT}B_k d_0^k+\sum_{j=1}^m\lambda_{1_j}^k\nabla g_j(x^k)^T d_0^k\Big]\\ \le&-d_1^{kT}B_k d_1^k-\sum_{j=1}^m\lambda_{0_j}^k\nabla g_j(x^k)^T d_0^k-\sum_{j=1}^m\lambda_{1_j}^k\nabla g_j(x^k)^T d_0^k+o(\|d_1^k\|^2),\qquad j\in\Omega\\ =&-d_1^{kT}B_k d_1^k+\sum_{j\in\Omega}\lambda_j^k g_j(x^k)+o(\|d_1^k\|^2),\qquad j\in\Omega\end{aligned}$$

和 $\nabla f(x^k)^T\widehat{d}^k=-\widehat{d}^{kT}B_k\widehat{d}^k+\sum\limits_{j=1}^m\widehat{\lambda}_j^k g_j(x^k+d_1^k)+o(\|d_1^k\|^2), j\in\Omega$. 又因为

$$\begin{aligned}&\nabla f(x^k)^T(d_1^k+\widehat{d}^k)\\ =&-d_1^{kT}B_k d_1^k+\sum_{j\in\Omega}\lambda_j^k g_j(x^k)-\widehat{d}^{kT}B_k\widehat{d}^k+\sum_{j\in\Omega}\widehat{\lambda}_j^k g_j(x^k+d_1^k)+o(\|d_1^k\|^2)\\ \le&-d_1^{kT}B_k d_1^k+\sum_{j\in\Omega}\widehat{\lambda}_j^k\nabla g_j(x^k)^T(d_1^k+\widehat{d}^k)+o(\|d_1^k\|^2),\qquad j\in\Omega.\end{aligned}$$

从而

$$g_j(x^k+d_1^k+\widehat{d}^k)=g_j(x^k)+\nabla g_j(x^k)^T(d_1^k+\widehat{d}^k)+\frac{1}{2}d_1^{kT}\nabla_{xx}^2 g_j(x^k)d_1^k+o(\|d_1^k\|).$$

故

$$-\sum_{j\in\Omega}\lambda_j^k\nabla g_j(x^k)^T(d_1^k+\widehat{d}^k)=\sum_{j\in\Omega}\lambda_j^k g_j(x^k)+\frac{1}{2}d_1^{kT}\Big(\sum_{j\in\Omega}\lambda_j^k\nabla_{xx}^2 g_j(x^k)\Big)d_1^k+o\big(\|d_1^k\|^2\big).$$

我们有

$$\begin{aligned}s\triangleq&f(x^k+d_1^k+\widehat{d}^k)-\max_{0\le j\le m(k)}f_{[k-j]}(x^k)-\nu\nabla f(x^k)^T d_1^k\\ \le&f(x^k+d_1^k+\widehat{d}^k)-f(x^k)-\nu\nabla f(x^k)^T d_1^k,\\ \le&-d_1^{kT}B_k d_1^k-\sum_{j\in\Omega}\lambda_j^k\nabla g_j(x^k)^T(d_1^k+\widehat{d}^k)+\frac{1}{2}d_1^{kT}\nabla_{xx}^2 f(x^k)d_1^k-\nu\nabla f(x^k)^T d_1^k\\ =&-d_1^{kT}B_k d_1^k+\sum_{j\in\Omega}\lambda_j^k g_j(x^k)+\frac{1}{2}d_1^{kT}\Big(\sum_{j\in\Omega}\lambda_j^k\nabla_{xx}^2 g_j(x^k)\Big)d_1^k\\ &+\frac{1}{2}d_1^{kT}\nabla_{xx}^2 f(x^k)d_1^k+\nu d_1^{kT}B_k d_1^k-\nu\sum_{j\in\Omega}\lambda_j^k g_j(x^k)+o(\|d_1^k\|^2)\\ =&(\nu-1)d_1^{kT}B_k d_1^k+(1-\nu)\sum_{j\in\Omega}\lambda_j^k g_j(x^k)\end{aligned}$$

$$+\frac{1}{2}d_1^{kT}\Big(\sum_{j\in\Omega}\lambda_j^k\nabla_{xx}^2 g_j(x^k)+\nabla_{xx}^2 f(x^k)\Big)d_1^k+o(\|d_1^k\|^2)$$
$$=(\nu-1)d_1^{kT}B_kd_1^k+\frac{1}{2}d_1^{kT}\nabla_{xx}^2L(x^k,\lambda_k)d_1^k+(1-\nu)\sum_{j\in\Omega}\lambda_j^kg_j(x^k)+o(\|d_1^k\|^2)$$
$$=\Big(\nu-\frac{1}{2}\Big)d_1^{kT}B_kd_1^k-\frac{1}{2}d_1^{kT}B_kd_1^k+\frac{1}{2}d_1^{kT}\nabla_{xx}^2L(x^k,\lambda_k)d_1^k$$
$$+(1-\nu)\sum_{j\in\Omega}\lambda_j^kg_j(x^k)+o(\|d_1^k\|^2).$$

记 $A_*=(\nabla g_j(x^*),\ j\in\Omega),\ P_*=I_n-A_*(A_*^TA_*)^{-1}A_*^T$, 则 $P_k\to P_*$.
令 $d_1^k=P_*d_1^k+y_k,\ y_k=A_*(A_*^TA_*)^{-1}A_*^Td_1^k$, 从而 $\Omega_k=\Omega$, 有

$$y_k=A_*(A_*^TA_*)^{-1}(A_*-A_k)^Td_1^k+A_*(A_*^TA_*)^{-1}A_*^Td_1^k,$$

故有 $\|y_k\|=o(\|d_1^k\|)+O(\|d_1^k\|^2)$, 因此

$$s\le\Big(\nu-\frac{1}{2}\Big)m\|d_1^k\|^2+\frac{1}{2}(d_1^{kT}P_*+y_k^T)(\nabla_{xx}^2L(x^k,\lambda^k)-B_k)d_1^k$$
$$+(1-\nu)\sum_{j\in\Omega}\lambda_j^kg_j(x^k)+o(\|d_1^k\|^2)$$
$$=\Big(\nu-\frac{1}{2}\Big)m\|d_1^k\|^2+(1-\nu)\sum_{j\in\Omega}\lambda_j^kg_j(x^k)+o(\|d_1^k\|^2)+o(\|d_1^k\|)+O(\|d_1^k\|^2)\le 0.$$

结合上述各种情况，结论成立．再由引理 4.2 和引理 4.3 可得到如下收敛性定理.

定理 4.4 算法是超线性收敛的，即

$$\|x^{k+1}-x^*\|=o(\|x^k-x^*\|).$$

5 数值试验

利用数值测试函数如下

1. 该问题是用来说明 Maratos 效应产生的例子

$$\min f(x)=x_1^2+x_2^2$$
$$\text{s.t. } g(x)=-(x_1+1)^2-x_2^2+4\le 0.$$

初始点 $x^0=(4,2)$，最优值为 $f(x^*)=1$.

2. 该问题是用来测试算法的例子

$$\min f_0(x)=(1-x_1)^2-10(x_2-x_1)^2+x_1^2-2x_1x_2+e^{-x_1-x_2}$$
$$\text{s.t. } g_1(x)=x_1^2+x_2^2-16\le 0,$$
$$g_2(x)=2-x_1-x_2\le 0,$$
$$g_3(x)=(x_2-x_1)^2+x_1-6\le 0.$$

初始点 $x^0=(2,3)$, 最优值为 $f(x^*)=-947.5255$.

3. 测试函数为

$$\min f(x)=\sum_{j=1}^{10}(\ln^2(x_j-2)+\ln^2(10-x_j))-\Big(\prod_{j=1}^{10}x_j\Big)^{0.2}$$
$$\text{s.t. } 2.001\le x_j\le 9.999,\qquad j=1,2,\cdots,10.$$

最优值为 $f(x^*)=-45.77846971$.

上面的函数在 Matlab 7.1 环境下运行实现， CPU 为奔腾 (R), 2.19GHz, 1G 内存. 程序代码为 Matlab 语言，测试函数 1 的初始点 x^0 为 $(4,2)$ 和 $(20,10)$, 测试函数 2 的初始点 x^0 为 $(2,3)$ 和 $(2.5,2.5)$ 测试函数 3 的初始点 x^0 与 [18] 相同，所有测试函数的精度要求为 $e=10^{-8}$, 相应的参数选取与文献相同，用 P 代表测试函数， NTR 代表算法 2.2 的迭代次数， NRk 代表 QP 子问题迭代次数， Ndk 表示不满足子问题的次数， S 代表迭代时间， CPU 的运行时间单位为秒， E 代表误差， MINF 表目标函数的最后函数值， x^* 为最优点， $-$ 代表文献中没有给出具体的数据. 程序运行结果见下表 1.

对数值结果进行说明，与 P 相对应的数字中， 1, 2 代表第一个测试函数所取不同的测试初始点得到的结果， 3, 4 代表第二个测试函数所取不同的测试初始点得到的结果， 5 代表测试函数三的测试结果， 1[13], 2[13], 3[18] 分别代表文献中的数值测试结果. 从下面数值结果比较表 1 中，我们可以得出，算法 2.2 是可行的，有效的. 从 CPU 的运行时间来看，所有结果都在我们可接受的时间内实现. 算法针对相同的函数，相同的初始点，通过迭代次数和 CPU 时间可以看出，算法 2.2 是优于文献中的算法的. 表 1 说明文中对可行性条件通过压缩技术效果很好，减少了 $\widetilde{QP}$ 和 QP 子问题的计算次数，提高了计算效率.

表 1 算法 2.2 与文献算法比较的数值结果

P	NTR	NRk	Ndk	S	x^*	$MINF$	E
1	5	4	1	2	$(1.0000000031,0)$	1.000000	$7.251334e-016$
2	54	50	4	8	$(1.0000000033,0)$	1.000000	$7.450580e-009$
3	580	578	2	38	$(3.50365456,1.92125879)$	-947.5255	10^{-8}
4	957	948	9	44	$(3.50365965,1.92125478)$	-947.5255	10^{-8}
5	377	369	8	20	$9.3502659921*(1,1,\cdots,1)$	-45.778469	10^{-8}
1[13]	–	–	–	–	$(1.000000,0.000000)$	1.000000	10^{-6}
2[13]	–	–	–	–	$(3.505484,1.926496)$	-947.5255	10^{-6}
3[18]	500	474	26	40	$9.3502658401*(1,1,\cdots,1)$	-45.778469	10^{-8}

参 考 文 献

[1] Powell M J D. On the Global Convergence of Trust Region Algorithm for Unconstrained Minimization. *Mathematical Programming*, 1984, 29: 297–303

[2] Shultz G A, Schbaoel R B, Byrd R H. A Family of Trust-region-based Algorithms for Unconstrained Minimization with Strong Global Convergence Properties. *SIAM Journal on Numerical Analysis*,

1985, 22: 47–67

[3] Yuan Y. On a Subproblem of Trust Region Algorithms for Constrained Optimization. *Mathematical Programming*, 1990, 47: 53–63

[4] Sun W Y. On Nonquadratic Model for Unconstrained Optimization. *Asia and Pacific Journal of Operational Research*, 1996, 13: 43–63

[5] Grippo L, Lampariello F, Lucidi S. A Truncated Newton Method with Nonmonotone Line Search for Unconstrained Optimization. *Journal of Optimization Theory and Application*, 1989, 60: 401–419

[6] Yang Y, Sun W Y. A New Nonmonotonic Self-adaptive Trust Region Algorithm with Line Search. *Chinese Journal of Engineering Mathematics*, 2007, 24: 788–794

[7] Fu J H, Sun W Y. Nonmonotone Adaptive Trust-region Method for Unconstrained Optimization Problems. *Applied and Mathematical Computation*, 2005, 163: 489–504

[8] Han S P. A Globally Convergent Method for Nonlinear Programming. *Journal of Optimization Theory and Applications*, 1977, 22: 297–309

[9] Panier E R, Tits A L, Zhou J. Avoiding the Marotos Effect by Means of a Nonmonotone Line Search I: General Constrained Problems. *SIAM Journal on Numerical Analysis*, 1991, 28: 1183–1195

[10] Bonnans J F, Panier E R, Tits A L, Zhou J. Avoiding the Marotos Effect by Means of a Nonmonotone Line Search II: Inequality Constrained Problems-feasible Iterates. *SIAM Journal on Numerical Analysis*, 1992, 29: 1187–1202

[11] Panier E R, Tits A L, Hersovits J N. A QP-free, Globally Convergent, Locally Superlinearly Convergent Algorithm for Inequality Constrained Optimization. *SIAM Journal on Control and Optimization*, 1988, 26: 788–811

[12] Zhang J L, Zhang X S. A Modified SQP Method with Nonmonotone Linesearch Technique. *Jounal of Global Optimization*, 2001, 21: 201–218

[13] 简金宝. 非线性不等式约束最优化快速收敛的可行信赖域算法. 计算数学, 2002, 24(3): 273–282 (Jian J B. A Feasible Trust Region Algorithm with Fast Convergence for Nonlinear Inequality Constrained Opitmization. *Jounal of Mathematica Numerica Sinica*, 2002, 24: 273–282)

[14] Zhu Z B, Zhang K C, Jian J B. A Improved SQP Algorithm for Inequality Constrained Optimization. *Jounal of Mathematical Method of Operations Research*, 2003, 58: 271–282

[15] Zhu Z B. A Interior Point Type QP-gree Algorithm with Superlinear Convergence for Inequality Constrained Optimization. *Jounal of Applied Mathematical Modelling*, 2007, 31: 1201–1212

[16] Zhu Z B. A Sequential Constrained Quadratic Programming Algorithm for Inequality Constrained Optimization. *Jounal of Computational and Applied Mathematics*, 2008, 212: 112–125

[17] Yuan Y, Sun W Y. Optimization Theory and Methods Nonlinear Programming, Vol.1. New York: Springer-Verlag, 2006

[18] 朱志斌. 非线性最优化问题可行序列二次规划算法的研究. 博士学位论文，西安，西安交通大学， 2004 (Zhu Z B. Research on Feasible Sequential Quadratic Programming Algorithms for Nonlinear Optimization. Ph.D.Thesis, Xi’an Jiaotong University, Xi’an, 2004)

A Feasible Trust Region SQP Method with Nonmonotone Line Search for Inequality Constrained Optimization

SUN Zhongbo

(*Department of Mathematical Education,*

College of Humanities and Sciences of Northeast Normal University, Changchun 130117)

(*E-mail: szb21971@yahoo.com.cn*)

DUAN Fujian

(*School of Computational Science and Mathematics,*

Guilin University of Electronic Technology, Guilin 541004)

Abstract In this paper, inequality constrained programming problems are discussed, based on a combination technique of a trust region method and an SQP method, a new feasible algorithm is proposed. A "compression" technique is used such that search direction is feasible for QP subproblem. We use high order revised direction to avoid Marotos effect. Under some suitable conditions, the global and superlinear convergence can be induced.

Key words nonmonotone line search; trust region method; SQP method; global convergence; superlinear convergence

MR(2000) Subject Classification 90C30

Chinese Library Classification O224.2