一簇非线性等式约束优化问题的过滤线搜索修正正割方法 *

王祝君

(湖南工程学院理学院，湘潭 411104)

(E-mail: wangzj1998@yahoo.cn)

朱德通

(上海师范大学商学院，上海 200234)

摘　要　本文提供了一簇新的过滤线搜索修正正割方法求解非线性等式约束优化问题. 新算法簇的特点是: 用修正正割算法簇中的一个算法获得搜索方向, 回代线搜索技术得到步长, 过滤准则用来决定是否接受步长，引入二阶校正技术减少不可行性并克服 Maratos 效应. 在合理的假设条件下，分析了算法的总体收敛性. 并证明了，通过附加二阶校正步，算法簇克服了 Maratos 效应，并二步 Q- 超线性收敛到满足二阶充分最优条件的局部解. 数值结果表明了所提供的算法具有有效性.

关键词　约束优化；过滤方法；正割算法； Maratos 效应；二阶校正

MR(2000) 主题分类　65K05; 90C30; 90C51; 90C55

中图分类　O221.2

1　引言

本文研究非线性等式约束优化问题：

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c(x) = 0, \end{aligned} \tag{1.1}$$

其中 $f(x): R^n \to R^1$ 和 $c(x): R^n \to R^m$ 都是二次连续可微的，且有 $m \leq n$.

非线性模型 (1.1), 常常出现在科学，工程以及社会的各个领域，因而是相当重要的

本文 2008 年 4 月 14 日收到. 2012 年 2 月 15 日收到修改稿.

* 国家自然科学基金 (10871130), 湖南省教育厅自然科学基金 (11C0336) 资助项目.

一类问题. 目前有许多文献记载了解决这类问题的各种方法, 如序列二次规划 (SQP) 方法 [1], 正割方法 [2] 等.

解非线性约束最优化问题的传统方法是罚函数法 (价值函数法), 但选择罚函数需要确定罚参数的合适值，而寻求合适的罚参数是比较困难的. 由 Fletcher 和 Leyffer 提出的过滤方法 [3] 避免了罚函数法的缺点，是一种保证算法总体收敛性的工具. 过滤算法引入一个含约束违反度的函数，把问题 (1.1) 转化成一个双目标最优化问题：只要试验点的试探步改进了目标函数或者改进了约束违反度，就认为这是一次成功的迭代. 过滤算法很好地平衡了目标函数和约束条件. 相对于罚函数法，它有几个方面的优点：不需要罚参数的值，而这个值是很难获得的；放宽了迭代步的接受条件，起到了类似于非单调技术的加速迭代收敛的作用. 值得一提的是， Wätcher 和 Biegler[4,5] 在 Fletcher 和 Leyffer[1,3] 的过滤算法基础上，提出了过滤线搜索方法解非线性规划问题，为解 (1.1) 提供了很好的启示. 因为过滤方法可能遭遇 Maratos 效应， Wätcher 和 Biegler 引入了二阶校正步防止 Maratos 效应.

另一方面， Fontecilla 在 [2] 中提出的正割方法，是最成功的解非线性等式约束优化问题的方法之一. 它把每步迭代中搜索步 s_k 作为一个水平步 h_k 和一个垂直步 v_k 的和. 水平步 h_k 在 DFP 或 BFGS 正割校正中用来校正矩阵 B_k, 而 v_k 用来满足线性化约束性质. Fontecilla 证明了，在合适的假设下，由正割方法产生的序列 $\{x_k\}$ 局部二步 Q- 超线性收敛.

基于正割方法和过滤线搜索算法的优点，本文提出了解非线性等式约束优化问题的过滤线搜索修正正割方法. 在每步迭代，用修正正割方法产生搜索方向 s_k, 用回代线搜索程序产生步长 α_k, 用过滤准则决定试验步是否被过滤集接受. 当过滤线搜索程序拒绝了完全牛顿步，仅接受一个更小的试验步时，用带二阶校正技术的修正正割算法产生新的试验步. 本文结构如下：下节介绍一类修正正割方法和过滤线搜索方法，提出解非线性等式约束优化问题的基本算法. 第 3 节分析了算法的总体收敛性，第 4 节了讨论了算法具有的局部收敛速率. 数值实验的结果在第 5 节给出.

2 算法

本文中， g_k 表示目标函数梯度 $\nabla f(x_k)$, A_k 表示约束函数 $c(x)$ 的梯度 $\nabla c(x_k)$.

修正的正割算法簇有以下的迭代格式：

算法 I

给定 x_0, λ_0, B_0. 从 $k=0$ 进行下面的运算

$$\lambda_{k+1} = U(x_k, \lambda_k, B_k), \tag{2.1}$$

$$B_k w_k = -\nabla_x l(x_k, \lambda_{k+1}), \tag{2.2}$$

$$h_k = P(x_k) w_k, \tag{2.3}$$

$$v_k = -A^+(x_k) c_k, \tag{2.4}$$

$$y_k = \nabla_x l(x_k + h_k, \lambda_{k+1}) - \nabla_x l(x_k, \lambda_{k+1}), \tag{2.5}$$

$$B_{k+1} = \mathrm{DFP/BFGS}(h_k, y_k, B_k), \tag{2.6}$$

$$s_k = h_k + v_k, \tag{2.7}$$

$$s_k^{\mathrm{soc}} = -A^+(x_k)c(x_k + s_k), \tag{2.8}$$

$$x_{k+1} = x_k + s_k + s_k^{\mathrm{soc}}, \tag{2.9}$$

其中 l 是 (1.1) 的 Lagrange 函数

$$l(x, \lambda) = f(x) + \lambda^T c(x), \tag{2.10}$$

$x \in R^n$, $\lambda \in R^m$. (2.1) 中最常用的乘子校正公式有下面三种选择：

(i) 投影校正：

$$\lambda_{k+1}^P = -(A_k^T A_k)^{-1} A_k^T g_k; \tag{2.11}$$

(ii) 零空间校正：

$$\lambda_{k+1}^S = -(A_k^T B_k^{-1} A_k)^{-1} A_k^T B_k^{-1} g_k; \tag{2.12}$$

(iii) 牛顿校正：

$$\lambda_{k+1}^{\mathrm{N}} = (A_k^T B_k^{-1} A_k)^{-1}(c_k - A_k^T B_k^{-1} g_k). \tag{2.13}$$

(2.3) 中的投影 $P(x_k)$ 满足

$$A_k^T P(x_k) = 0, \tag{2.14}$$

它可以是斜投影

$$P_{B_k} = I - B_k^{-1} A_k (A_k^T B_k^{-1} A_k)^{-1} A_k^T, \tag{2.15}$$

当 $B_k = I$ 时，它就是正交投影：

$$P_k = I - A_k (A_k^T A_k)^{-1} A_k^T. \tag{2.16}$$

A_k^T 的广义逆 $A^+(x_k)$, 可写成

$$A_{B_k}^+ = B_k^{-1} A_k (A_k^T B_k^{-1} A_k)^{-1}, \tag{2.17}$$

当 B_k 是单位阵 I 时，就成为

$$A_{I_k}^+ = A_k (A_k^T A_k)^{-1}. \tag{2.18}$$

相应于 A_k^T 广义逆的选择， v_k 可写成

$$v_{B_k} = -B_k^{-1} A_k (A_k^T B_k^{-1} A_k)^{-1} c(x_k) \tag{2.19}$$

或

$$v_{I_k} = -A_k (A_k^T A_k)^{-1} c(x_k), \tag{2.20}$$

从而 $A_k^T v_k + c_k = 0$. 满足

$$c(x_k + s_k + s_k^{\text{soc}}) = o(\|s_k\|^2) \tag{2.21}$$

和

$$s_k^{\text{soc}} = o(\|s_k\|) \tag{2.22}$$

的 s_k^{soc}, 称为二阶校正步 [6]. 即在点 $x_k + s_k$ 给约束增加一个附加步 s_k^{soc}, 以减少不可行性，使迭代克服遭遇 Maratos 效应的困难，并获得一个改进的搜索方向. 这样的校正项有很大的选择范围，具体可以由下面公式得到.

给定 s_k, 二阶校正步 s_k^{soc} 由 (2.8) 给出，结合 (2.17)–(2.18), s_k^{soc} 可写成

$$s_{B_k}^{\text{soc}} = -B_k^{-1} A_k (A_k^T B_k^{-1} A_k)^{-1} c(x_k + s_k), \tag{2.23}$$

若 $B_k = I$, 它就变成

$$s_{I_k}^{\text{soc}} = -A_k (A_k^T A_k)^{-1} c(x_k + s_k). \tag{2.24}$$

s_k^{soc} 满足

$$A_k^T s_k^{\text{soc}} + c(x_k + s_k) = 0. \tag{2.25}$$

若 s_k^{soc} 是一个二阶校正步，$\widetilde{s}_k^{\text{soc}}$ 是附加的二阶校正步 (即在 (2.10) 中用 "$c(x_k+s_k+s_k^{\text{soc}})$" 代替 "$c(x_k+s_k)$"), 则 $s_k^{\text{soc}}+\widetilde{s}_k^{\text{soc}}$ 对 s_k 可当作一个单一的 (single) 二阶校正步. 类似地，几个连续的校正步也可作为单一校正步.

在 DFP 或 BFGS 正割校正公式中， h_k 用来校正矩阵 B_k,

$$\begin{aligned} B_{k+1}^{\text{DFP}} &= B_k + \frac{(y_k - B_k h_k) y_k^T + y_k (y_k - B_k h_k)^T}{y_k^T h_k} - \frac{h_k^T (y_k - B_k h_k) y_k y_k^T}{(y_k^T h_k)^2}, \\ B_{k+1}^{\text{BFGS}} &= B_k + \frac{y_k y_k^T}{y_k^T h_k} - \frac{(B_k h_k)(B_k h_k)^T}{h_k^T B_k h_k}. \end{aligned} \tag{2.26}$$

用 (2.11)–(2.13), (2.15)–(2.18) 中分别对应乘子校正，投影算子和 A_k^T 广义逆的不同选择，可得表 2.1 中六种不同的算法.

表 2.1 六种算法

Algorithm	λ_{k+1}	$P(x_k)$	v_k	s_k^{soc}
ALG1	(2.12)	I	(2.20)	(2.23)
ALG2	(2.12)	I	(2.19)	(2.23)
ALG3	(2.11)	P_{B_k}	(2.20)	(2.24)
ALG4	(2.11)	P_{B_k}	(2.19)	(2.24)
ALG5	(2.13)	P_k	(2.20)	(2.23)
ALG6	(2.13)	P_k	(2.19)	(2.23)

当 $s_k^{\text{soc}} = 0$ 时，上面的六种算法就是 Fontecilla 在 [2] 中提出的二步算法. 本文将证明修正的正割算法可以结合过滤线搜索方法，产生具有良好总体收敛性和快速局部收敛速率的新算法.

在搜索方向确定后，可用过滤方法来确定步长 $\alpha_k \in (0,1)$ 以获得下一步迭代．过滤方法的基本思想是把 (1.1) 当作一个双目标优化问题：最小化约束违反度 $\theta(x)$ 以及最小化目标函数 $f(x)$, 其中 $\theta(x)$ 定义为 $\|c(x)\|_1$. 因而需要求试验步长 $\alpha_{k,l}$ 满足一个简单条件：它使得 f 或 θ 的值有所下降．当然，这样的条件不足以保证总体收敛性，而且，也不能接受使得 $\theta(x_k(\alpha_{k,l}))$ 或 $f(x_k(\alpha_{k,l}))$ 各自充分接近 $\theta(x_k)$ 或 $f(x_k)$ 的新试验点 $x_k(\alpha_{k,l})$. 为了避免这种情况，需加强新试验点必须满足的条件．更准确地说，当满足条件

$$\theta(x_k(\alpha_{k,l})) \le (1-\gamma_\theta)\theta(x_k) \tag{2.27}$$

或

$$f(x_k(\alpha_{k,l})) \le f(x_k) - \gamma_f \theta(x_k) \tag{2.28}$$

时，就接受试验点 $x_k(\alpha_{k,l}) := x_k + \alpha_{k,l} s_k$, 其中 $\gamma_\theta, \gamma_f \in (0,1)$. 用这个过滤准则，可能会出现试验点每次总是满足约束违反度条件 (2.27) 的情况，这样就可能收敛到一个可行点，而不是最优点．为避免这种情况，用一个 f- 型切换条件代替上述条件

$$m_k(\alpha_{k,l}) < 0, \qquad [-m_k(\alpha_{k,l})]^{s_f}[\alpha_{k,l}]^{1-s_f} > \delta[\theta(x_k)]^{s_\theta}, \tag{2.29}$$

其中 $\delta > 0$, $s_\theta > 1$, $s_f \ge 1$, $m_k(\alpha) := \alpha g_k^T s_k$ 是目标函数 f 在方向 s_k 的线性模型．若条件 (2.29) 满足， s_k 是目标函数的下降方向，要使 $x_k(\alpha_{k,l})$ 被接受，试验点还必须满足 Armijo 条件

$$f(x_k(\alpha_{k,l})) \le f(x_k) + \eta_f m_k(\alpha_{k,l}), \tag{2.30}$$

其中 $\eta_f \in \left(0, \frac{1}{2}\right)$.

算法用到一个含 (θ, f)- 对的过滤集 $\mathcal{F}_k$. 若

$$(\theta(x_k(\alpha_{k,l})), f(x_k(\alpha_{k,l}))) \in \mathcal{F}_k, \tag{2.31}$$

就说试验点不被当前过滤集接受．初始过滤集定义为

$$\mathcal{F}_0 := \{(\theta, f) \in R^2 : \theta \ge \theta_{\max}\}, \tag{2.32}$$

其中 $\theta_{\max} > \theta(x_0)$. 在迭代过程中，过滤集用校正公式

$$\mathcal{F}_{k+1} := \mathcal{F}_k \cup \left\{(\theta, f) \in R^2 : \theta \ge (1-\gamma_\theta)\theta(x_k) \text{ 和 } f \ge f(x_k) - \gamma_f\theta(x_k)\right\} \tag{2.33}$$

进行扩张．当被接受的步长或者不满足 f- 型切换条件 (2.29), 或者不满足 Armijo 条件 (2.30) 时，就用 (2.33) 扩张过滤集．若被接受的步长既满足 (2.29), 又满足 (2.30), 过滤集保持不变．这种方式保证了算法不会发生循环.

在试验步不满足上面所有准则的情况下，用所涉及函数的线性模型逼近一个最小步长 $\alpha_k^{\min}$,

$$\alpha_k^{\min} := \gamma_\alpha \cdot \begin{cases} \min\left\{\gamma_\theta, \dfrac{\gamma_f\theta(x_k)}{-g_k^T s_k}, \dfrac{\delta[\theta(x_k)]^{s_\theta}}{[-g_k^T s_k]^{s_f}}\right\}, & g_k^T s_k < 0, \\ \gamma_\theta, & g_k^T s_k \ge 0, \end{cases} \tag{2.34}$$

其中 $\gamma_\alpha \in (0,1]$. 当 $\alpha_{k,l}$ 比 $\alpha_k^{\min}$ 更小时，算法转向可行性恢复阶段．恢复阶段的目的是当回代线搜索程序不能使目标函数充分下降，步长太小时，通过减少不可行性，产生一个扩张的过滤集 $\mathcal{F}_{k+1}$ 接受的新迭代.

下面给出解问题 (1.1) 的过滤线搜索修正正割方法.

算法 II 给定初始点 x_0; 常数 $\theta_{\max} \in (\theta(x_0), \infty]$; $\gamma_\theta, \gamma_f \in (0,1)$; $0 < \tau_1 \le \tau_2 < 1$; $\delta > 0$; $\gamma_\alpha \in (0,1]; s_\theta > 1$; $s_f \ge 1$; $\eta_f \in (0, \frac{1}{2}); \varepsilon$; λ_0; $n \times n$ 对称正定矩阵 B_0.

步 1. 由 (2.32) 初始化过滤集 $\mathcal{F}_0$. 迭代计数 $k \leftarrow 0$.

步 2. 计算 f_0, g_0, c_0, A_0, 以及 f_k, g_k, c_k, w_k, A_k.

步 3. 检验收敛性．若 $\|w_k\| + \|c_k\| \le \varepsilon$, 停止．否则，转下一步.

步 4. 计算搜索方向.

步 4.1. 用 (2.11)–(2.13) 计算乘子.

步 4.2. 解线性方程 (2.2)–(2.4) 得到 h_k 和 v_k, 计算 $s_k = h_k + v_k$.

若线性系统是病态的，转可行性恢复阶段.

步 5. 回代线搜索.

步 5.1. 初始化线搜索．设 $\alpha_{k,0} = 1$ 且 $l \leftarrow 0$.

步 5.2. 计算新的试验点．若试验步长太小，即 $\alpha_{k,l} < \alpha_k^{\min}$, $\alpha_k^{\min}$ 由 (2.34) 给出，转第 10 步可行性恢复阶段．否则，计算试验点 $x_k(\alpha_{k,l}) := x_k + \alpha_{k,l} s_k$.

步 5.3. 若 $x_k(\alpha_{k,l}) \in \mathcal{F}_k$, 拒绝试验步长，转第 5.5 步.

步 5.4. 检验当前迭代的充分下降性.

步 5.4.1 情况 I: $\alpha_{k,l}$ 是 f- 型步长，即 (2.29) 满足：若 Armijo 条件 (2.30) 满足，接受试验步，转第 6 步，否则转第 5.5 步.

步 5.4.2. 情况 II: $\alpha_{k,l}$ 不是 f- 型步长，即 (2.29) 不满足：若 (2.27) 和 (2.28) 满足，接受试验步，转第 6 步，否则转第 5.5 步.

步 5.5. 计算二阶校正步．如果 $l \ne 0$, 转 5.8 步．否则，用 (2.8) 计算二阶校正步，得到新的试验点 $\widetilde{x}_{k+1} = x_k + s_k + s_k^{\mathrm{soc}}$.

步 5.6. 检验过滤集的可接受性．若 $\widetilde{x}_{k+1} \in \mathcal{F}_k$, 拒绝二阶校正步，转第 5.8 步.

步 5.7. 检验当前迭代的充分下降性.

步 5.7.1. 情况 I: 切换条件 (2.29) 满足：若 Armijo 条件

$$f(\widetilde{x}_{k+1}) \le f(x_k) + \eta_f m_k(\alpha_{k,l})$$

满足，接受 $x_{k+1} := \widetilde{x}_{k+1}$, 转第 6 步．否则，转第 5.8 步.

步 5.7.2. 情况 II: 切换条件 (2.29) 不满足：若 $\theta(\widetilde{x}_{k+1}) \le (1-\gamma_\theta)\theta(x_k)$ 或 $f(\widetilde{x}_{k+1}) \le f(x_k) - \gamma_f\theta(x_k)$ 满足，接受 $x_{k+1} := \widetilde{x}_{k+1}$, 转第 6 步．否则，转第 5.8 步．

步 5.8. 设 $\alpha_{k,l+1} \in [\tau_1\alpha_{k,l},\ \tau_2\alpha_{k,l}]$, $l \leftarrow l+1$, 回到第 5.2 步．

步 6. 接受试验点．设 $\alpha_k := \alpha_{k,l}$ 且 $x_{k+1} := x_k(\alpha_k) = x_k + \alpha_k s_k$.

步 7. 如有必要扩张过滤集．如果 k 不是 f- 型迭代，扩张过滤集．否则，设 $\mathcal{F}_{k+1} := \mathcal{F}_k$.

步 8. 校正 B_{k+1}. 计算 h_k, y_k, 用 DFP 或 BFGS 正割校正公式通过校正 B_k 获得 B_{k+1}.

步 9. 设 $k = k+1$, 回到第 3 步．

步 10. 可行性恢复阶段．通过下降不可行量 θ 计算新的迭代 x_{k+1}, 使得 x_{k+1} 满足 (2.27)–(2.28), 从而可被过滤集接受，即 $(\theta(x_{k+1}), f(x_{k+1})) \notin \mathcal{F}_k$. 用 (2.33) 扩张过滤集，从第 9 步继续迭代．

注 1 过滤机制保证了对所有的 k, $(\theta(x), f(x)) \notin \mathcal{F}_k$. 第 1 步的初始过滤集和校正准则 (2.33) 说明了对所有的 k, 若 $\theta \le \overline{\theta}$ 且 $f \le \overline{f}$, 则 $(\overline{\theta}, \overline{f}) \notin \mathcal{F}_k \Rightarrow (\theta, f) \notin \mathcal{F}_k$.

注 2 在总体收敛性分析中，仅要求 $s_f > 1$, 但在局部收敛速率的分析中，要求 $s_f > 2s_\theta$.

3 总体收敛性分析

本文用 $\mathcal{A} \subseteq N$ 表示过滤集被扩张时相应迭代的指标集，即 $\mathcal{F}_k \subset \mathcal{F}_{k+1} \Leftrightarrow k \in \mathcal{A}$. $\mathcal{R} \subseteq N$ 表示进入可行性恢复阶段的所有迭代指标集．第 10 步保证了 $\mathcal{R} \subseteq \mathcal{A}$. $\mathcal{R}_{\rm inc} \subseteq \mathcal{R}$ 表示从第 4 步进入可行性恢复阶段的所有迭代指标集．显然，这些集合满足 $\mathcal{R}_{\rm inc} \subseteq \mathcal{R} \subseteq \mathcal{A} \subseteq N$.

要得到总体收敛性结果，需假设第 10 步的可行性恢复阶段总是成功终止，且算法在第 3 步不停止．另外还需用到下面的一些假设：

(G1) 对所有的 $k \notin \mathcal{R}_{\rm inc}$ 存在一个满足 $[x_k, x_k + s_k] \in C$ 的开集 $C \subseteq R^n$, 使得 $f(x)$, $c(x)$ 在 C 上是可微的，它们的函数值以及它们的一阶导数，在 C 上是有界的且是 Lipschitz 连续的．

(G2) 矩阵 B_k 在零子空间 $\mathcal{N}(A(x)^T)$ 上的斜投影是一致正定的，即存在常数 $M_B > 0$ 使得对所有的 $k \notin \mathcal{R}_{\rm inc}$,

$$h^T[P(x_k)B_kP(x_k)]h \ge M_B\|h\|^2, \qquad \forall\, h \tag{3.1}$$

成立．而且， B_k 是非奇异且是范数有界的．

(G3) 矩阵 A_k 在 C 上列满秩．对所有的 k, 范数列 $\{\|A^+(x_k)\|\}$ 是有界的．

(G4) 从第 4 步进入可行性恢复阶段的迭代不会任意逼近可行域．换句话说，存在常数 $\theta_{\rm inc} > 0$, 每当 $\theta(x_k) \le \theta_{\rm inc}$, 就有 $k \notin \mathcal{R}_{\rm inc}$.

定义一个临界度量

$$\chi(x_k) := \begin{cases} \|w_k\|_2, & k \notin \mathcal{R}_{\text{inc}}, \\ \infty, & k \in \mathcal{R}_{\text{inc}}, \end{cases} \tag{3.2}$$

其中 w_k 取自 (2.2). 考虑满足 $\lim\limits_i \chi(x_{k_i}) = 0$ 和 $\lim\limits_i x_{k_i} = x_*$ 的迭代子列 $\{x_{k_i}\}$, 其中 x_* 是可行极限点. 由 $\chi(x_k)$ 的定义，可得对充分大的 i, $k_i \notin \mathcal{R}_{\text{inc}}$. 从 $\lim\limits_i \chi(x_{k_i}) = 0$, (3.2), (2.2) 和假设 (G2), 易得 $\lim\limits_i \|g_{k_i} + A_{k_i}\lambda_{k_i}\| = 0$. 因此，在上述假设下，用这种方式定义的 $\chi(x_k)$ 是一个临界度量.

第一个结论引用 [2] 的引理 4.1–4.2.

引理 3.1 设 h_k, v_k 和 s_k 由算法 II 产生，则 $s_k = h_k + v_k$ 满足

$$s_k = -B_k^{-1}\nabla_x l(x_k, \lambda_{k+1}^N) + \tau_k(A_{Bk}^+ - A_{I_k}^+)c_k, \tag{3.3}$$

其中 λ_{k+1}^N 由 (2.13) 给出，对算法 ALG2, ALG4, ALG5, $\tau_k = 0$, 对算法 ALG1, ALG3 和 ALG6, τ_k 分别对应 $1, 1, -1$.

引理 3.2 设 h_k, v_k 和 s_k 由算法 II 产生，则

$$A_k^T h_k = 0, \tag{3.4}$$

$$g_k^T h_k = -h_k^T B_k h_k - v_k^T B_k h_k - \tau_k c_k^T A_{I_k}^{+T} B_k^T h_k, \tag{3.5}$$

$$g_k^T v_k = -g_k^T A^+(x_k)c_k, \tag{3.6}$$

$$g_k^T s_k = -h_k^T B_k h_k - v_k^T B_k h_k - \tau_k c_k^T A_{I_k}^{+T} B_k^T h_k - g_k^T A^+(x_k)c_k, \tag{3.7}$$

$$A_k^T s_k = -c_k. \tag{3.8}$$

证 因为 $A_k^T P(x_k) = 0$, 所以 $A_k^T h_k = A_k^T P(x_k) w_k = 0$. 把 (3.3) 变形成为

$$g_k = -B_k h_k - B_k v_k - A_k \lambda_k^N + \tau_k B_k (A_{B_k}^+ - A_{I_k}^+)c_k.$$

两边同乘以 h_k, 再由 (3.4) 知

$$\begin{aligned} g_k^T h_k &= -h_k^T B_k h_k - v_k^T B_k h_k - (\lambda_k^N)^T A_k^T h_k + \tau_k c_k^T (A_{B_k}^+ - A_{I_k}^+) B_k^T h_k \\ &= -h_k^T B_k h_k - v_k^T B_k h_k - \tau_k c_k^T A_{I_k}^{+T} B_k^T h_k. \end{aligned}$$

从 (2.7), (2.4) 和 (3.5) 可得到

$$\begin{aligned} &g_k^T v_k = -g_k^T A^+(x_k)c_k, \\ &g_k^T s_k = -h_k^T B_k h_k - v_k^T B_k h_k - \tau_k c_k^T A_{I_k}^{+T} B_k^T h_k - g_k^T A^+(x_k)c_k, \\ &A_k^T s_k + c_k = A_k^T h_k + A_k^T v_k + c_k = A_k^T v_k + c_k = 0. \end{aligned}$$

引理 3.3 假设 (G1)–(G4) 成立. 那么存在常数 $M_d, M_\lambda, M_m > 0$, 使得对所有的 $k \notin \mathcal{R}_{\text{inc}}$ 和 $\alpha \in (0,1]$, 有 $\|s_k\| \le M_d, \|\lambda_k\| \le M_\lambda$, $|m_k(\alpha)| \le M_m\alpha$.

证　假设 (G1)–(G4) 保证了 B_k, A_k, 向量 g_k 和 c_k 一致有界．因此，由 (2.11) 和 (2.12), 可得出 $\|\lambda_k\| \le M_\lambda$. 考虑 (2.13), 存在 $M_\lambda \ge 0$ 使得

$$\begin{aligned}\|\lambda_{k+1}\| \le & \|(A_k^T B_k^{-1} A_k)^{-1}\| \, \|(c_k - A_k^T B_k^{-1} g_k)\| \\ \le & \|(A_k^T B_k^{-1} A_k)^{-1}\| \left(\|c_k\| + \|A_k^T B_k^{-1}\| \|g_k\|\right) \\ \le & M_\lambda.\end{aligned}$$

(2.3)–(2.4) 说明了 h_k 和 v_k 一致有界，结合 (2.7), 对所有的 $k \notin \mathcal{R}_{\rm inc}$, 有 $\|s_k\| \le M_d$. $m_k(\alpha)$ 的定义，假设 (G1) 和 $\|s_k\| \le M_d$ 说明了存在某个常数 $M_m > 0$, $|m_k(\alpha)| \le M_m \alpha$.

引理 3.4 假设 (G1)–(G4) 成立．若 $\{x_{k_i}\}$ 是一个迭代子列，满足 $\chi(x_{k_i}) \ge \varepsilon$, $\varepsilon > 0$ 是与 i 无关的常数，则存在 $\varepsilon_1, \varepsilon_2 > 0$, 使得对所有 i,

$$\theta(x_{k_i}) \le \varepsilon_1 \Longrightarrow m_{k_i}(\alpha) \le -\varepsilon_2 \alpha, \tag{3.9}$$

其中 $\alpha \in (0,1]$.

证　考虑满足 $\chi(x_{k_i}) = \|w_{k_i}\|_2 \ge \varepsilon$ 的迭代子列 $\{x_{k_i}\}$．由假设 (G4), 对所有满足 $\theta(x_{k_i}) \le \theta_{\rm inc}$ 的 x_{k_i}, 有 $k_i \notin \mathcal{R}_{\rm inc}$．而且，因为 h_k 一致有界，$v_{k_i} = O(\|c(x_{k_i})\|)$, 对 $k_i \notin \mathcal{R}_{\rm inc}$, 可得

$$\begin{aligned} m_{k_i}(\alpha)/\alpha =& g_{k_i}^T s_{k_i} \\ \overset{(3.7)}{=} & -h_{k_i}^T B_{k_i} h_{k_i} - v_{k_i}^T B_{k_i} h_{k_i} - \tau_{k_i} c_{k_i}^T (A_{k_i}^+)^T B_{k_i}^T h_{k_i} - g_{k_i}^T A^+(x_{k_i}) c_{k_i} \\ \overset{(2.3)}{=} & -w_{k_i}^T P(x_{k_i}) B_{k_i} P(x_{k_i}) w_{k_i} + O(\|c_{k_i}\|) \\ \overset{(G2)}{\le} & -c_1 \|w_{k_i}\|_2^2 + c_2 \|c_{k_i}\| \\ \overset{(3.2),(3.9)}{\le} & \chi(x_{k_i}) \Big(-c_1 \varepsilon + \frac{c_2 \theta(x_{k_i})}{\varepsilon} \Big), \end{aligned} \tag{3.10}$$

其中，$c_1, c_2 > 0$, $P(x_{k_i})$ 是上面提到的斜投影．最后一个不等式用到 $\chi(x_{k_i}) \ge \varepsilon$. 定义 $\varepsilon_1 := \min\left\{\theta_{\rm inc}, \frac{\varepsilon^2 c_1}{2c_2}\right\}$, 则对所有满足 $\theta(x_{k_i}) \le \varepsilon_1$ 的 x_{k_i}, 有

$$m_{k_i}(\alpha) \le -\alpha \frac{\varepsilon c_1}{2} \chi(x_{k_i}) \le -\alpha \frac{\varepsilon^2 c_1}{2}.$$

设 $\varepsilon_2 := \frac{\varepsilon^2 c_1}{2}$, 可知命题成立．

引理 3.5 假设 (G1) 满足．则存在常数 $C_\theta, C_f > 0$ 使得对所有的 $k \notin \mathcal{R}_{\rm inc}$ 和 $\alpha \le 1$,

$$|\theta(x_k + \alpha s_k) - (1-\alpha)\theta(x_k)| \le C_\theta \alpha^2 \|s_k\|^2, \tag{3.11}$$

$$|f(x_k + \alpha s_k) - f(x_k) - m_k(\alpha)| \le C_f \alpha^2 \|s_k\|^2. \tag{3.12}$$

证　因为

$$\begin{aligned} & |\theta(x_k + \alpha s_k) - (1-\alpha)\theta(x_k)| = \big|\, \|c(x_k + \alpha s_k)\| - (1-\alpha)\|c_k\| \,\big| \\ =& |\|c_k + \alpha A_k^T s_k + O(\alpha^2 \|s_k\|^2)\| - (1-\alpha)\|c_k\|| \end{aligned}$$

$$\le \|\|c_k - \alpha c_k\| + O(\alpha^2\|s_k\|^2) - (1-\alpha)\|c_k\|\|$$
$$= O(\alpha^2\|s_k\|^2),$$

因此 (3.11) 满足. 第二个不等式 (3.12) 直接从 Taylor 展开式得到.

以上面这些引理为基础, 与 [5] 中定理 2 的证明过程相同, 可以证明所提的算法有下面的总体收敛性结果.

定理 3.1 假设 (G1)–(G4) 成立, 则 $\lim\limits_{k\to\infty}\theta(x_k)=0$ 且 $\lim\limits_{k\to\infty}\inf\chi(x_k)=0$. 换句话说, 所有的极限点都是可行点. 若 $\{x_k\}$ 有界, 则存在 $\{x_k\}$ 的极限点 x_*, 是等式约束优化问题 (1.1) 的一阶最优点.

4 局部收敛速率分析

局部收敛速率分析除了假设算法簇产生的迭代列 $\{x_k\}$ 收敛到 (1.1) 的局部解 x_* 外, 还要用到以下假设.

(L1) 函数 $f(x)$ 和 $c(x)$ 有二阶导数, 并且在 x_* 的邻域内 Lipschitz 连续.

(L2) x_* 满足二阶充分最优条件:
- x_* 是问题 (1.1) 的 KKT 点, 即存在 $\lambda_*\in R^m$, 使得 $c_*=0,\ g_*-A_*\lambda_*=0$.
- x_* 是 (1.1) 的正则点, 即 A_* 列满秩.
- Lagrange 函数的 Hessian 阵 $W_*=\nabla^2_{xx}l(x_*,\lambda_*)$ 沿 $\mathcal{N}(A_*^T)$ 正定.

(L3) 矩阵 B_k 非奇异且范数有界.

(L4)
$$\lim_{k\to\infty}\frac{\|P_k(B_k-W_*)h_k\|}{\|h_k\|}=0, \tag{4.1}$$
其中 $W(x,\lambda)=\nabla^2_{xx}l(x,\lambda)$, $W_k=W(x_k,\lambda_k)$, P_k 是 $\mathcal{N}(A_k^T)$ 上的正交投影. 条件 (4.1) 可等价地写成
$$h_k^TB_kh_k=h_k^TW_kh_k+o(\|h_k\|^2). \tag{4.2}$$

(L5) 存在常数 $\theta_{\rm inc}>0$, 若 $\theta(x_k)\le\theta_{\rm inc}$, 则算法在第 4 步不切换到可行性恢复阶段.

引理 4.1 假设 (L1)–(L5) 成立, 则存在 x_* 的邻域 U_1, 使得对所有 $x_k\in U_1$, 有
$$\|s_k^{\rm soc}\|=O(\|s_k\|^2), \tag{4.3}$$
$$\|c(x_k+s_k+s_k^{\rm soc})\|=o(\|s_k\|^2). \tag{4.4}$$

证 因为
$$c(x_k+s_k)=c(x_k)+A_k^Ts_k+O(\|s_k\|^2)\overset{(3.8)}{=}O\left(\|s_k\|^2\right),$$

(2.8) 和假设 (L3) 证明了 (4.3) 成立．因为 $x_k \to x_*$,

$$\begin{aligned} c(x_k+s_k+s_k^{\rm soc}) =& c(x_k+s_k)+A(x_k+s_k)^T s_k^{\rm soc}+O\big(\|s_k^{\rm soc}\|^2\big) \\ \overset{(2.8)}{=}& c(x_k+s_k)-A(x_k+s_k)^T A^+(x_k)c(x_k+s_k)+O(\|s_k^{\rm soc}\|^2) \\ =& -[A(x_k+s_k)-A_k]^T A^+(x_k)c(x_k+s_k)+O(\|s_k\|^4) \\ =& O(\|s_k\|\|s_k\|^2)+O(\|s_k\|^4)=o(\|s_k\|^2). \end{aligned}$$

注意到 $A(x)$ 是 Lipschitz 连续的．

显然，这个引理证明了当 x_k 逼近局部解 x_* 时，组合步 $s_k+s_k^{\rm soc}$ 比初始步 s_k 有更好的被过滤方法接受的机会．下一个引理证明，在 x_* 的邻域内，若组合步 $s_k+s_k^{\rm soc}$ 是一个 f- 型步，算法 II 的第 5.7.1 步是成功的．

引理 4.2 假设 (L1)–(L5) 成立．则存在 x_* 的一个邻域 $U_2 \subseteq U_1$, 使得只要当 $\alpha_{k,l}=1$ 时 (2.29) 成立，就有 Armijo 条件 (2.30) 满足．

证　首先确定 h_k, v_k 和 s_k 之间的关系．设 U_1 由引理 4.1 确定．由 (2.29), 知

$$\|c(x_k)\|=\theta(x_k)<\delta^{-\frac{1}{s_\theta}}[-g_k^T s_k]^{\frac{s_f}{s_\theta}}=o(\|s_k\|^2). \tag{4.5}$$

由 (2.4) 和 (4.5) 可知

$$v_k=O(\|c_k\|)=o(\|s_k\|^2), \qquad s_k=O(\|h_k\|)+O(\|v_k\|)=O(\|h_k\|)+o(\|s_k\|^2),$$

故 $s_k=O(\|h_k\|)$ 且 $v_k=o(\|h_k\|^2)$.

下面的证明中，向量 $v\in R^n$ 的第 i 个分量用 v_i 表示．对 ALG1 和 ALG2 有

$$\begin{aligned} g_k^T s_k^{\rm soc} \overset{(2.23)}{=}& -g_k^T[B_k^{-1}A_k(A_k^T B_k^{-1}A_k)^{-1}]c(x_k+s_k) \\ \overset{(2.12)}{=}& (\lambda_{k+1}^S)^T c(x_k+s_k) \\ =& \sum_{i=1}^m \lambda_{k+1,i}^S\Big[c_i(x_k)+\nabla c_i(x_k)^T s_k+\frac{1}{2}s_k^T\nabla^2 c_i(x_k)s_k+o(\|s_k\|^2)\Big] \\ =& \frac{1}{2}s_k^T\Big(\sum_{i=1}^m \lambda_{k+1,i}^S\nabla^2 c_i(x_k)\Big)s_k+o(\|s_k\|^2). \end{aligned}$$

对 ALG3 和 ALG4 有

$$\begin{aligned} & g_k^T s_k^{\rm soc} \overset{(2.24)}{=} -g_k^T A_k(A_k^T A_k)^{-1}c(x_k+s_k) \overset{(2.11)}{=} (\lambda_{k+1}^P)^T c(x_k+s_k) \\ =& \frac{1}{2}s_k^T\Big(\sum_{i=1}^m \lambda_{k+1,i}^P\nabla^2 c_i(x_k)\Big)s_k+o(\|s_k\|^2). \end{aligned}$$

对 ALG5 和 ALG6 有

$$\begin{aligned} g_k^T s_k^{\rm soc} \overset{(2.23)}{=}& -g_k^T[B_k^{-1}A_k(A_k^T B_k^{-1}A_k)^{-1}]c(x_k+s_k) \\ =& [(A_k^T B_k^{-1}A_k)^{-1}(c_k-A_k^T B_k^{-1}g_k)]^T c(x_k+s_k)-[(A_k^T B_k^{-1}A_k)^{-1}c_k]^T \end{aligned}$$

$$\begin{aligned}
&\overset{(2.13)}{=}(\lambda_{k+1}^N)^T c(x_k+s_k)-c_k^T(A_k^T B_k^{-1}A_k)^{-1}\\
&=\frac{1}{2}s_k^T\Big(\sum_{i=1}^m \lambda_{k+1,i}^N\nabla^2 c_i(x_k)\Big)s_k+o(\|s_k\|^2).
\end{aligned}$$

注意到上面关于六个算法的关系式可统一写成：

$$g_k^T s_k^{\rm soc}=\frac{1}{2}s_k^T\Big(\sum_{i=1}^m \lambda_{k+1,i}\nabla^2 c_i(x_k)\Big)s_k+o(\|s_k\|^2), \tag{4.6}$$

其中， λ_{k+1} 是 $\lambda_{k+1}^P,\lambda_{k+1}^S$ 或 λ_{k+1}^N 中的一个．因而

$$\begin{aligned}
&f(x_k+s_k+s_k^{\rm soc})-f(x_k)=g_k^T s_k+g_k^T s_k^{\rm soc}+\frac{1}{2}s_k^T\nabla^2 f_k s_k+o(\|s_k\|^2)\\
&\overset{(4.6)}{=}g_k^T s_k+\frac{1}{2}s_k\Big(\sum_{i=1}^m \lambda_{k+1,i}\nabla^2 c_i(x_k)\Big)s_k+\frac{1}{2}s_k^T\nabla^2 f_k s_k+o(\|s_k\|^2)\\
&=g_k^T s_k+\frac{1}{2}s_k^T W_k s_k+o(\|s_k\|^2).
\end{aligned}$$

进一步，

$$\begin{aligned}
&f(x_k+s_k+s_k^{\rm soc})-f(x_k)-\eta {g_k}^T s_k\\
&=g_k^T(s_k+s_k^{\rm soc})+\frac{1}{2}s_k^T\nabla^2 f_k s_k+o(\|s_k\|^2)-\eta {g_k}^T s_k\\
&=(1-\eta)g_k^T s_k+\frac{1}{2}s_k^T W_k s_k+o(\|s_k\|^2)\\
&\overset{(3.7)}{=}(1-\eta)[-h_k^T B_k h_k-v_k^T B_k h_k-\tau_k c_k^T A_{I_k}^{+T}B_k^T h_k-g_k^T A^+(x_k)c_k]\\
&\quad+\frac{1}{2}s_k^T W_k s_k+o(\|s_k\|^2)\\
&\overset{(4.2)}{=}\Big(\eta-\frac{1}{2}\Big)h_k^T B_k h_k+o(\|h_k\|^2)-(1-\eta)v_k^T B_k h_k-(1-\eta)\tau_k c_k^T A_{I_k}^{+T}B_k^T h_k\\
&\quad-(1-\eta)g_k^T A^+(x_k)c_k+\frac{1}{2}h_k^T W_k v_k+\frac{1}{2}v_k^T W_k v_k+o(\|s_k\|^2)\\
&\overset{(4.5)}{=}\Big(\eta-\frac{1}{2}\Big)h_k^T B_k h_k+o(\|h_k\|^2)+O(\|v_k\|\|h_k\|)\\
&\quad+O(\|v_k\|^2)+O(\|c_k\|)+O(\|c_k\|\|h_k\|)+o(\|s_k\|^2)\\
&\le\Big(\eta-\frac{1}{2}\Big)h_k^T B_k h_k+o(\|h_k\|^2).
\end{aligned}$$

因为当 $x_k\to x_*$ 时， $h_k\to 0$, 所以对所提的六个算法， Armijo 条件 (2.30) 在 $\alpha=1$ 时满足，其中用到假设 (L2) 和 $\eta_f<\frac{1}{2}$.

下面用精确罚函数 $\phi_\rho(x)=f(x)+\rho\theta(x)$ 和罚函数的模型

$$q_\rho(x_k,h,v)=f(x_k)+g_k^T h+g_k^T v+\frac{1}{2}h^T B_k h+v^T B_k h+\rho\|A_k^T h+A_k^T v+c_k\| \tag{4.7}$$

作为技术手段处理二阶校正步的影响，而算法中并没有提及它们.

引理 4.3 假设 (L1)–(L5) 成立，则存在邻域 $U_3 \subseteq U_2$ 和正常数 ρ, 使得对所有的 $x_k \in U_3$, 有

$$\phi_\rho(x_k) - \phi_\rho(x_k + s_k + s_k^{\mathrm{soc}}) \ge \frac{1+\gamma_\theta}{2}(q_\rho(x_k,0,0) - q_\rho(x_k,h_k,v_k)) \ge 0.$$

证　对某个正常数 δ_1,

$$\begin{aligned}
&q_\rho(x_k,0,0) - q_\rho(x_k,h_k,v_k)\\
\overset{(4.7)}{=}&\rho\|c_k\| - g_k^T h_k - g_k^T v_k - \frac{1}{2}h_k^T B_k h_k - v_k^T B_k h_k - \rho\|A_k^T h_k + A_k^T v_k + c_k\|\\
\overset{(3.8)}{=}&\rho\|c_k\| - g_k^T h_k - g_k^T v_k - \frac{1}{2}h_k^T B_k h_k - v_k^T B_k h_k\\
\overset{(3.7)}{=}&\rho\|c_k\| + h_k^T B_k h_k + v_k^T B_k h_k + \tau_k c_k^T A_{I_k}^{+T} B_k^T h_k + g_k^T A^+(x_k) c_k\\
&- \frac{1}{2}h_k^T B_k h_k - v_k^T B_k h_k\\
=&\rho\|c_k\| - \lambda_{k+1}^T c_k + \frac{1}{2}h_k^T B_k h_k + \tau_k c_k^T A_{I_k}^{+T} B_k^T h_k\\
\ge&(\rho - \|\lambda_{k+1}\|_\infty - \delta_1)\|c_k\| + \frac{1}{2}h_k^T B_k h_k,
\end{aligned} \tag{4.8}$$

这里用到了 $\|h_k\|$ 的一致有界性和 Cauchy-Schwartz 不等式以导出最后的不等式. 选择 $\rho \ge \|\lambda_{k+1}\|_\infty + \delta_1$, 可得 $q_\rho(x_k,0,0) - q_\rho(x_k,h_k,v_k) \ge 0$, 这样证明了命题中第二个不等式. 进一步，可得

$$\begin{aligned}
&\phi_\rho(x_k) - \phi_\rho(x_k + s_k + s_k^{\mathrm{soc}}) - \frac{1+\gamma_\theta}{2}(q_\rho(x_k,0,0) - q_\rho(x_k,h_k,v_k))\\
\overset{(4.8)}{=}&f(x_k) + \rho\theta(x_k) - f(x_k + s_k + s_k^{\mathrm{soc}}) - \rho\theta(x_k + s_k + s_k^{\mathrm{soc}})\\
&- \frac{1+\gamma_\theta}{2}\Big(\rho\|c_k\| - \lambda_{k+1}^T c_k + \frac{1}{2}h_k^T B_k h_k + \tau_k c_k^T A_{I_k}^{+T} B_k^T h_k\Big)\\
=&-g_k^T s_k - \frac{1}{2}s_k^T W_k s_k + o(\|s_k\|^2) + \frac{1-\gamma_\theta}{2}\rho\|c_k\|\\
&- \frac{1+\gamma_\theta}{2}\Big(-\lambda_{k+1}^T c_k + \frac{1}{2}h_k^T B_k h_k + \tau_k c_k^T A_{I_k}^{+T} B_k^T h_k\Big)\\
\overset{(4.2)}{=}&\frac{1-\gamma_\theta}{4}h_k^T B_k h_k + v_k^T B_k h_k + \frac{1-\gamma_\theta}{2}\tau_k c_k^T A_{I_k}^{+T} B_k^T h_k + \frac{\gamma_\theta - 1}{2}\lambda_{k+1}^T c_k\\
&- \frac{1}{2}h_k^T W_k v_k - \frac{1}{2}v_k^T W_k v_k + \frac{1-\gamma_\theta}{2}\rho\|c_k\| + o(\|s_k\|^2) + o(\|h_k\|^2)\\
=&\frac{\gamma_\theta - 1}{2}\lambda_{k+1}^T c_k + \frac{1-\gamma_\theta}{4}h_k^T B_k h_k + \frac{1-\gamma_\theta}{2}\rho\|c_k\|\\
&+ O(\|v_k\|\|h_k\|) + O(\|v_k\|^2) + o(\|s_k\|^2) + o(\|h_k\|^2)\\
=&\frac{\gamma_\theta - 1}{2}\lambda_{k+1}^T c_k + \frac{1-\gamma_\theta}{4}h_k^T B_k h_k + \frac{1-\gamma_\theta}{2}\rho\|c_k\| + o(\|s_k\|^2) + o(\|h_k\|^2)\\
\overset{(4.5)}{\ge}&\frac{1-\gamma_\theta}{2}(\rho - \|\lambda_{k+1}\|_\infty)\|c_k\| + \frac{1-\gamma_\theta}{4}h_k^T B_k h_k + o(\|h_k\|^2) + o(\|s_k\|^2)\\
\ge&\frac{1-\gamma_\theta}{2}(\rho - \|\lambda_{k+1}\|_\infty)\|c_k\| + o(\|h_k\|^2).
\end{aligned}$$

最后的不等式可由 $\gamma_\theta < 1$ 和假设 (L3) 得出. 定义 $\rho \ge \|\lambda_{k+1}\|_\infty$, 可知命题成立.

定义三个常数 ρ_1,ρ_2,ρ_3 满足：

$$
\begin{aligned}
&\rho_1=\rho>\|\lambda_{k+1}\|_\infty,\\
&\rho_2=\frac{1+\gamma_\theta}{1-\gamma_\theta}\rho_1+\frac{3\gamma_f}{1-\gamma_\theta},\\
&\rho_3=(1-\gamma_\theta)\rho_2-\gamma_f.
\end{aligned}
$$

引理 4.4 假设 (L1)–(L5) 成立，则存在邻域 $U_3\subseteq U_2$ 和常数 $\rho_1,\rho_2,\rho_3>0$, 使得对所有的 $x_k\in U_3$, 有

$$
\begin{aligned}
&2\gamma_\theta\rho_2<(1+\gamma_\theta)(\rho_2-\rho_1)-2\gamma_f,\\
&2\rho_3\geq(1+\gamma_\theta)\rho_1+(1-\gamma_\theta)\rho_2,\\
&\rho_i>\|\lambda_{k+1}\|_\infty,\qquad i=1,2,3.
\end{aligned}
$$

证 由 ρ_i $(i=1,2,3)$ 的定义，可得

$$
\begin{aligned}
&(1+\gamma_\theta)(\rho_2-\rho_1)-2\gamma_f-2\gamma_\theta\rho_2=(1-\gamma_\theta)\rho_2-(1+\gamma_\theta)\rho_1-2\gamma_f\\
=&(1+\gamma_\theta)\rho_1+3\gamma_f-(1+\gamma_\theta)\rho_1-2\gamma_f=\gamma_f>0.
\end{aligned}
$$

从而有 $2\gamma_\theta\rho_2<(1+\gamma_\theta)(\rho_2-\rho_1)-2\gamma_f$. 从

$$
\begin{aligned}
(1+\gamma_\theta)\rho_1+(1-\gamma_\theta)\rho_2-2\rho_3=&(1+\gamma_\theta)\rho_1+(1-\gamma_\theta)\rho_2-2(1-\gamma_\theta)\rho_2+2\gamma_f\\
=&(1+\gamma_\theta)\rho_1-(1-\gamma_\theta)\rho_2+2\gamma_f\\
=&(1+\gamma_\theta)\rho_1-(1+\gamma_\theta)\rho_1-3\gamma_f+2\gamma_f\\
=&-\gamma_f<0,
\end{aligned}
$$

知第二个不等式成立. 注意到 $\rho_2-\rho_1=\frac{2\gamma_\theta\rho_1}{1-\gamma_\theta}\rho_1+\frac{3\gamma_f}{1-\gamma_\theta}$ 且 $\rho_3-\rho_1=\gamma_\theta\rho_1+2\gamma_f>0$. 因此， $\rho_2,\rho_3>\rho_1$. 结合 ρ_1 的定义可得 $\rho_i>\|\lambda_{k+1}\|_\infty$, $i=1,2,3$.

下面这个引理是引理 4.3 的推广. 主要的不同在于二阶校正步的选择. 按照前面的分析，有其它的二阶校正步 $\widetilde{s}_k^{\rm soc}$ 的定义，把所有的 $\widetilde{s}_k^{\rm soc}$ 的选择列在下面：

$$
\begin{aligned}
&\widetilde{s}_k^{\rm soc}=s_k^{\rm soc},\qquad \widetilde{s}_k^{\rm soc}=\sigma_k s_k^{\rm soc}+s_{k+1}+\sigma_{k+1}s_{k+1}^{\rm soc},\\
&\widetilde{s}_k^{\rm soc}=\sigma_k s_k^{\rm soc}+s_{k+1}+\sigma_{k+1}s_{k+1}^{\rm soc}+s_{k+2}+\sigma_{k+2}s_{k+2}^{\rm soc},\\
&\widetilde{s}_k^{\rm soc}=\sigma_k s_k^{\rm soc}+s_{k+1}+\sigma_{k+1}s_{k+1}^{\rm soc}+s_{k+2}+\sigma_{k+2}s_{k+2}^{\rm soc}+s_{k+3}+\sigma_{k+3}s_{k+3}^{\rm soc},
\end{aligned}
$$

其中 $\sigma_k,\sigma_{k+1},\sigma_{k+2},\sigma_{k+3}\in\{0,1\}$.

引理 4.5 假设 (L1)–(L5) 成立，则存在邻域 $U_3\subseteq U_2$, 对所有的 $x_k\in U_3$, 只要 $x_{l+1}=x_l+s_l+\sigma_l s_l^{\rm soc}$, $l\in\{k,\cdots,k+j\}$, $j\in\{-1,0,1,2\}$, 就有

$$
\phi_{\rho_i}(x_k)-\phi_{\rho_i}(x_k+s_k+\widetilde{s}_k^{\rm soc})\geq\frac{1+\gamma_\theta}{2}(q_{\rho_i}(x_k,0)-q_{\rho_i}(x_k,s_k))\geq 0,
$$

其中 $i=2,3$, $\widetilde{s}_k^{\rm soc}$ 是上面列出的选择中任一个.

用 [4] 中定理 4.7 相同的证明技巧和方法，可证明下面的定理，它在局部收敛分析中扮演了重要的角色.

定理 4.1 假设 (L1)–(L5) 成立．那么，对充分大的 k, 可取得 $x_{k+1}=x_k+s_k$ 或 $x_{k+1}=x_k+s_k+s_k^{\rm soc}$ 形式的完全步长.

为了分析局部收敛速率，需引入下面的定义和引理.

因为 W_* 沿着 $\mathcal{N}(A_*^T)$ 正定，存在正常数 $\overline{c}$, 使得对所有的 $c\geq\overline{c}$, $W_*^c=W_*+cA_*A_*^T$ 是正定的．考虑 $U:R^n\times R^m\times R^{n\times n}\rightarrow R^m$. 如果存在非负常数 $\overline{\phi}<1$, 使得对包含在 $(x_*,\lambda_*,W_*^{\overline{c}})$ 的开邻域 N_1 中的每个 (x,λ,B), $\lambda_+=U(x,\lambda,B)$ 满足 $\|(W_*^{\overline{c}})^{-1}A_*(\lambda_+-\lambda_*)\|\leq\overline{\phi}\|x-x_*\|$, 其中 $\overline{\phi}$ 依赖于 $\overline{c}$, 就说乘子校正 U 是 x- 占优的 (在 $W_*^{\overline{c}}$). 注意到若 U 是 x- 占优的，那么存在非负常数 ϕ, 使得 $\|A_*(\lambda_+-\lambda_*)\|\leq\phi\|x-x_*\|$, 其中 ϕ 依赖于 $\overline{c}$.

以下三个引理由 Fontecilla 在 [2] 中给出.

引理 4.6 设 $F:R^n\rightarrow R^n$ 在包含 x_* 的开凸集 $D\subset R^n$ 里是连续可微的．假设 F' 在 D 上是 Lipschitz 连续的，而且 $[F^{'}(x_*)]^{-1}$ 存在．那么存在常数 $\varepsilon>0$, $\rho>0$, 使得对所有的 $u,v\in D$, $\max\{\|v-x_*\|,\|u-x_*\|\}\leq\varepsilon$, $\frac{1}{\rho}\|v-u\|\leq\|F(v)-F(u)\|\leq\rho\|v-u\|$.

引理 4.7 假设 (L1)–(L2) 成立．那么存在正常数 C_1,C_2, 和 $\varepsilon>0$, 使得对任意的 $\lambda\in R^m$ 和 $\sigma(x,x_+)\leq\varepsilon$, 有

$$\|\nabla_x l(x_+,\lambda)-\nabla_x l(x,\lambda)-W_*(x_+-x)\|\leq[C_1\sigma(x,x_+)+C_2\|\lambda-\lambda_*\|]\|x_+-x\|.$$

引理 4.8 假设 (L1)–(L4) 成立，且 $\{x_k\}$ 由 [2] 中二步算法产生，则序列 $\{x_k\}$ 二步超线性收敛到 x_*.

定理 4.2 假设 (L1)–(L5) 成立，则由算法 II 产生的序列 $\{x_k\}$ 二步超线性收敛到 x_*, 即

$$\lim_{k\rightarrow\infty}\frac{\|x_{k+1}-x_*\|}{\|x_{k-1}-x_*\|}=0.$$

证　按照定理 4.1, 对充分大的 k, 算法 II 取得 $x_{k+1}=x_k+s_k$ 或 $x_{k+1}=x_k+s_k+s_k^{\rm soc}$ 形式的完全步长．在取得 $x_{k+1}=x_k+s_k$ 形式的完全步长的情况下，由引理 4.8 可得结论．下面考虑完全步长为 $x_{k+1}=x_k+s_k+s_k^{\rm soc}$ 形式的情况.

由 (3.3) 可知，

$$B_kd_k=B_ks_k+B_ks_k^{\rm soc}=-\nabla_x l(x_k,\lambda_{k+1}^N)+\tau_kB_k(A_{B_k}^+-A_{I_k}^+)c_k-B_kA^+(x_k)c(x_k+s_k),$$

其中，$d_k=s_k+s_k^{\rm soc}$. 用 P_k 乘这个方程，而且因为 $P_kB_kA_{B_k}^+=0$, 可得

$$\begin{aligned}P_kB_kd_k=&-P_k\nabla_x l(x_k,\lambda_{k+1}^N)+P_k\tau_kB_k(A_{B_k}^+-A_{I_k}^+)c_k-P_kB_kA^+(x_k)c(x_k+s_k)\\=&-P_k\nabla_x l(x_k,\lambda_*)-\tau_kP_kB_kA_{I_k}^+c_k-P_kB_kA^+(x_k)c(x_k+s_k).\end{aligned}$$

两边同时加上 $-P_*\nabla_x l(x_{k+1},\lambda_*)-A_*c_{k+1}$, 整理可得

$$\begin{aligned}-P_*\nabla_x l(x_{k+1},\lambda_*)-A_*c_{k+1}=&-P_*[\nabla_x l(x_{k+1},\lambda_*)-\nabla_x l(x_k,\lambda_*)-W_*d_k]\\&+(P_k-P_*)[\nabla_x l(x_k,\lambda_*)-\nabla_x l(x_*,\lambda_*)]+(P_k-P_*)W_*d_k\end{aligned}$$

$$+ P_k(B_k - W_*)d_k + \tau_k P_k B_k A_{I_k}^+ c_k$$
$$+ P_k B_k A^+(x_k)c(x_k + s_k) - A_* c_{k+1}. \tag{4.9}$$

下面，用 K_i $(i = 1, 2, \cdots)$ 表示正常数．由引理 4.1–4.2, 有

$$\|s_k^{\mathrm{soc}}\| = O(\|s_k\|^2), \tag{4.10}$$
$$\|c_{k+1}\| = \|c(x_k + s_k + s_k^{\mathrm{soc}})\| = o(\|s_k\|^2), \tag{4.11}$$
$$\|c_k\| = o(\|s_{k-1}\|^2), \tag{4.12}$$
$$\|c(x_k + s_k)\| = O(\|s_k\|^2), \tag{4.13}$$
$$\|v_k\| = O(\|c_k\|) = o(\|s_{k-1}\|^2). \tag{4.14}$$

因为 $d_k = s_k + s_k^{\mathrm{soc}} = h_k + v_k + s_k^{\mathrm{soc}}$, 所以

$$\begin{aligned}
&\|P_k(B_k - W_*)d_k\| \\
\le&\|P_k(B_k - W_*)h_k\| + \|P_k(B_k - W_*)v_k\| + \|P_k(B_k - W_*)s_k^{\mathrm{soc}}\| \\
\overset{(4.10),(4.14)}{=}&\|P_k(B_k - W_*)h_k\| + o(\|s_{k-1}\|^2) + O(\|s_k\|^2),
\end{aligned} \tag{4.15}$$

并且 $\|d_k\| = \|s_k + O(\|s_k\|^2)\|$. 因此,

$$\|s_k\| - O(\|s_k\|^2) \le \|d_k\| \le \|s_k\| + O(\|s_k\|^2), \qquad \left|\frac{\|d_k\|}{\|s_k\|} - 1\right| = o(\|s_k\|).$$

从而

$$\|d_k\| = O(\|s_k\|). \tag{4.16}$$

由引理 4.7, 可得

$$\|\nabla_x l(x_{k+1}, \lambda_*) - \nabla_x l(x_k, \lambda_*) - W_* d_k\| \le K_1 \varepsilon \|d_k\| = o(\|d_k\|) = o(\|s_k\|). \tag{4.17}$$

考虑正交投影 P_k. 对所有 x_* 邻域内的 x_k, 有 $\|P_k - P_*\| \le K_2\|x_k - x_*\|$. 由于 $f(x)$ 和 $c(x)$ 是 Lipschitz 连续可微的，可得

$$\big\|(P_k - P_*)[\nabla_x l(x_k, \lambda_*) - \nabla_x l(x_*, \lambda_*)]\big\| \le K_3\|x_k - x_*\|^2 \tag{4.18}$$

和

$$\|(P_k - P_*)W_* d_k\| \le K_4\|x_k - x_*\|\|d_k\| = K_5\|x_k - x_*\|\|s_k\|. \tag{4.19}$$

定义算子 $H_c : H_c(x) = P_*\nabla_x l(x, \lambda_*) + cA_*c(x)$, 可以证明 $H_c(x_*) = 0$ 且 $H_c^{'}(x)$ 是非奇异的 [7]. 在 (4.9) 中取范数，用三角不等式，再把引理 4.7 用于 H_c, 可得

$$\begin{aligned}
\frac{1}{\rho}\|x_{k+1} - x_*\| \le& K_1\varepsilon\|d_k\| + K_3\|x_k - x_*\|^2 + K_5\|x_k - x_*\|\|s_k\| \\
&+ \|P_k(B_k - W_*)d_k\| + K_6\|c_k\| + K_7\|c_{k+1}\| + K_8\|c_k + s_k\| \\
=&K_3\|x_k - x_*\|^2 + K_5\|x_k - x_*\|\|s_k\| + \|P_k(B_k - W_*)h_k\| \\
&+ o(\|s_k\|) + o(\|s_{k-1}\|^2) + o(\|s_k\|^2) + O(\|s_k\|^2).
\end{aligned} \tag{4.20}$$

从 (2.9) 和 (4.16) 知

$$\|x_{k+1}-x_*\| = \|d_k + x_k - x_*\| \le O(\|s_k\|) + \|x_k - x_*\| \le K_9\|x_k - x_*\|. \tag{4.21}$$

由

$$\begin{aligned}\|s_k\| &\overset{(2.7)}{=} \|h_k + v_k\| \overset{(2.2)-(2.4)}{=} \|-P_{B_k}B_k^{-1}\nabla_x l(x_k,\lambda_{k+1}) - A(x_k)^+ c_k\|\\ &= \big\|P_{B_k}B_k^{-1}[\nabla_x l(x_k,\lambda_{k+1}) - \nabla_x l(x_*,\lambda_{k+1})]\\ &\quad + P_{B_k}B_k^{-1}[\nabla_x l(x_*,\lambda_{k+1}) - \nabla_x l(x_*,\lambda_*)] + A(x_k)^+(c_k - c_*)\big\|\end{aligned}$$

和乘子校正是 x- 占优的事实，(4.21) 以及假设 (L1), 可得

$$\|s_k\| \le K_{10}\|x_k - x_*\|, \tag{4.22}$$

$$\|s_{k-1}\| \le K_{10}\|x_{k-1} - x_*\|. \tag{4.23}$$

因为 $s_k = h_k + v_k = h_k - A^+(x_k)c_k$. 用 P_k 乘这个方程，由于 $A_k^T h_k = 0$, 从而

$$P_k s_k = P_k h_k = [I - A_k(A_k^T A_k)^{-1}A_k^T] = h_k. \tag{4.24}$$

用 P_{B_k} 乘这个方程，有同样的结果．所以，或者 $P_k s_k = h_k$ 或者 $P_{B_k} s_k = h_k$, 对这两种情况，都有

$$\|h_k\| \le \|s_k\| \overset{(4.22)}{\le} K_{10}\|x_k - x_*\|. \tag{4.25}$$

用 $\|x_{k-1} - x_*\|$ 除 (4.20), 且由 (4.22), (4.23) 和 (4.25), 可得

$$\begin{aligned}\frac{\|x_{k+1}-x_*\|}{\|x_{k-1}-x_*\|} \le& K_{11}\frac{\|P_k(B_k - W_*)h_k\|}{\|h_k\|} + K_{12}\frac{\|x_k - x_*\|^2}{\|x_{k-1}-x_*\|} + K_{13}\frac{\|x_k - x_*\|\|s_k\|}{\|s_k\|}\\ &+ \frac{o(\|s_k\|)}{\|s_k\|} + \frac{o(\|s_k\|^2)}{\|s_k\|} + \frac{o(\|s_{k-1}\|^2)}{\|s_{k-1}\|} + \frac{O(\|s_k\|^2)}{\|s_k\|}\\ =& K_{11}\frac{\|P_k(B_k - W_*)h_k\|}{\|h_k\|} + K_{14}\|x_k - x_*\|\\ &+ \frac{o(\|s_k\|)}{\|s_k\|} + \frac{o(\|s_k\|^2)}{\|s_k\|} + \frac{o(\|s_{k-1}\|^2)}{\|s_{k-1}\|}.\end{aligned} \tag{4.26}$$

对 (4.26) 两边同时取极限，由 (4.1) 和 $x_k \to x_*$, 可知命题成立.

下面定理表明，在更强的条件下，算法 ALG2, 4, 5 具有局部一步 Q- 超线性收敛速率.

定理 4.3　假设 (L1)–(L3) 成立，且 $\{x_k\}$ 由 ALG2, ALG4 以及 ALG5 中的一个产生．如果

$$\lim_{k\to\infty}\frac{\|P_k(B_k - W_*)s_k\|}{\|s_k\|} = 0, \tag{4.27}$$

那么 $\{x_k\}$Q- 超线性收敛到 x_*, 即

$$\lim_{k\to\infty}\frac{\|x_{k+1}-x_*\|}{\|x_k - x_*\|} = 0.$$

证 由定理 4.1, 对充分大的 k, 算法 II 取得 $x_{k+1}=x_k+s_k$ 或 $x_{k+1}=x_k+s_k+s_k^{\rm soc}$ 形式的完全步长. 首先考虑完全步长为 $x_{k+1}=x_k+s_k+s_k^{\rm soc}$ 形式的情况. 当 $\{x_k\}$ 由 ALG2, ALG4 以及 ALG5 中的一个产生时，由 (3.3) 有 $B_ks_k=-\nabla_xl(x_k,\lambda_{k+1}^N)$. 用 P_k 乘以这个等式两边，可得 $P_kB_ks_k=-P_k\nabla_xl(x_k,\lambda_{k+1}^N)=-P_k\nabla_xl(x_k,\lambda_*)$. 把 $-P_*\nabla_xl(x_{k+1},\lambda_*)-A_*c_{k+1}$ 加到等式两边并整理可得

$$\begin{aligned}&-P_*\nabla_xl(x_{k+1},\lambda_*)-A_*c_{k+1}\\=&-P_*[\nabla_xl(x_{k+1},\lambda_*)-\nabla_xl(x_k,\lambda_*)-W_*(s_k+s_k^{\rm soc})]\\&+(P_k-P_*)[\nabla_xl(x_k,\lambda_*)-\nabla_xl(x_*,\lambda_*)]\\&+P_k(B_k-W_*)s_k+(P_k-P_*)W_*s_k-P_*W_*s_k^{\rm soc}-A_*c_{k+1}.\end{aligned}\tag{4.28}$$

在下面的推导过程中，用 C_i $(i=3,4,\cdots)$ 表示正常数. 用 (4.20) 的方法，对 (4.28) 取范数可得

$$\begin{aligned}\|x_{k+1}-x_*\|\le&C_3\varepsilon\|s_k+s_k^{\rm soc}\|+C_4\|x_k-x_*\|^2+\|P_k(B_k-W_*)s_k\|\\&+C_5\|x_k-x_*\|\|s_k\|+C_6\|s_k^{\rm soc}\|+C_7\|c_{k+1}\|\\\le&C_4\|x_k-x_*\|^2+\|P_k(B_k-W_*)s_k\|\\&+C_5\|x_k-x_*\|\|s_k\|+o(\|s_k\|)+o(\|s_k\|^2),\end{aligned}\tag{4.29}$$

其中用到 (4.3), (4.4). 用 $\|x_k-x_*\|$ 除 (4.29) 可得

$$\frac{\|x_{k+1}-x_*\|}{\|x_k-x_*\|}\overset{(4.22)}{\le}C_8\frac{\|P_k(B_k-W_*)s_k\|}{\|s_k\|}+C_9\|x_k-x_*\|+\frac{o(\|s_k\|)}{\|s_k\|}+\frac{o(\|s_k\|^2)}{\|s_k\|}.\tag{4.30}$$

对 (4.30) 两边同时取极限，由 (4.27) 和 $x_k\to x_*$, 可知 $\{x_k\}$Q- 超线性收敛到 x_*.

当 $s_k^{\rm soc}=0$ 时，就是 $x_{k+1}=x_k+s_k$ 的情况，易证 $\{x_k\}$Q- 超线性收敛到 x_*.

注 3 目前还不能证明在相同条件下 ALG1,3,6 也具有一步 Q- 超线性收敛速率. 考虑证明超线性收敛速率的关键一步 (4.9) 的右边，式中有含 $\|c_k\|$ 的项，在 ALG2,4,5 中 $\tau_k=0$, 从而 (4.9) 式可消去 $\|c_k\|$ 项，而 ALG1,3,6 中 $\tau_k\ne0$, 故 (4.9) 中总含有 $\|c_k\|$ 项，由于 $\|c_k\|=o(\|s_{k-1}\|^2)$, 这些算法的一步 Q- 超线性收敛速率难以得到证明.

5 数值结果

对算法 II 进行了初步的数值实验和比较. 所有的程序都用 Matlab 7.0 编写，在 PentiumIV 计算机上运行. 本文中常数的取值情况为 $\tau_1=\tau_2=0.5$, $\eta_f=0.35$, $\gamma_\theta=\gamma_f=0.5$, $s_\theta=1.5$, $s_f=3.2$, B_0 取单位阵. 中止条件为 $\|w_k\|+\|c_k\|_1\le10^{-6}$. 测试题取自 [8] 中 Example 1–Example 4.

表 5.1 比较了修正的六种算法关于四个算题的运行结果，其中 N_t, N_a, N_{fg} 和 f_* 分别表示迭代次数，过滤集扩张次数， f 和 g 的计算次数，以及目标函数在 x_* 的值. 表 5.1 表明所提出的六个算法都是有效的.

表 5.1　六个正割过滤线搜索方法的数值结果

Algorithm	Example 1				Example 2			
	N_t	N_a	N_{fg}	f_*	N_t	N_a	N_{fg}	f_*
ALG1	6	4	13	0.015172	15	13	36	3.4774e-012
ALG2	6	4	13	0.015172	18	4	51	1.6411e-025
ALG3	19	15	33	0.015172	15	8	36	3.4774e-012
ALG4	12	9	26	0.015172	18	14	51	1.6411e-025
ALG5	12	9	26	0.015172	18	14	51	1.6411e-025
ALG6	7	5	14	0.015172	13	18	31	9.3551e-013
Algorithm	Example 3				Example 4			
	N_t	N_a	N_{fg}	f_*	N_t	N_a	N_{fg}	f_*
ALG1	17	15	84	0.05395	31	26	114	0.029311
ALG2	11	11	24	0.05395	31	23	141	0.029311
ALG3	17	14	84	0.05395	32	26	122	0.029311
ALG4	11	11	24	0.05395	33	22	118	0.029311
ALG5	16	15	52	0.05395	31	23	141	0.029311
ALG6	57	53	355	0.05395	26	23	86	0.029311

用算法簇中 ALG3 计算算题中的 Example 2 (初始点 $x_0=[1.65,-0.8165]^T$, 最优点 $x_*=[\sqrt{2},0]^T$), 可得表 5.2 中的数值结果，数据显示了该算法具有明显的二步 Q- 超线性收敛速率.

表 5.2　算法 ALG3 的数值结果

N_t	$\|x_k-x_*\|$	f_k	N_t	$\|x_k-x_*\|$	f_k	N_t	$\|x_k-x_*\|$	f_k
1	8.011e-001	5.215e-001	6	8139e-004	3.006e-008	11	5.878e-005	4.88e-009
2	1.668e-001	1.59e-002	7	1.452e-004	2.98e-008	12	1.792e-005	4.543e-010
3	1.034e-001	1.515e-002	8	5.326e-004	4.011e-007	13	2.578e-005	9.404e-010
4	4.742e-002	3.181e-003	9	6.393e-005	5.780e-009	14	1.925e-006	5.242e-012
5	4.743e-002	3.179e-003	10	7.869e-005	8.758e-009	15	1.568e-006	3.477e-012

参　考　文　献

[1] Fletcher R, Gould N I M, Leyffer S, Toint Ph L, Wächter A. Global Convergence of a Trust-region SQP-filter for General Nonlinear Programming. *SIAM J. Optim.*, 2002, 13(3): 635–659

[2] Fontecilla R. Local Convergence of Secant Method for Nonlinear Constrained Optimization. *SIAM J. Numer. Anal.*, 1988, 25(2): 692–712

[3] Fletcher R, Leyffer S. Nonlinear Programming without a Penaty Function. *Math. Program.*, 2002, 91(2): 44–59

[4] Wächter A, Biegler L T. Line Search Filter Methods for Nonlinear Programming: Local Convergence. *SIAM J. Optim.*, 2005, 6(1): 32–48

[5] Wächter A, Biegler L T. Line Search Filter Methods for Nonlinear Programming: Motivation and Global Convergence. *SIAM J. Optim.*, 2005, 16(1): 1–31

[6] Conn A R, Gould N I M, Toint Ph L. Trust Region Methods. MPS/SIAM Ser. Optim.1, SIAM, Philadelphia, 2000

[7] Fontecilla R, Steihaug T, Tapia R A. A Convergence Theory for a Class of Quasi-Newton Methods for Constrained Optimization. *SIAM J. Numer. Anal.*, 1987, 24(4): 1133–1151

[8] Wang Z J, Zhu D T. A Line Search Filter Secant Method for Nonlinear Equality Constrained Optimization. *J. Systems Sci. & Comp.*, 2010, 23: 343–361

A Class of Line Search Filter Improved Secant Methods for Nonlinear Equality Constrained Optimization

WANG Zhujun

(*Science College, Hunan Institute of Engineering, Xiangtan* 411104)

(*E-mail: wangzj1998@yahoo.cn*)

ZHU Detong

(*Business College, Shanghai Normal University, Shanghai* 200234)

Abstract This paper proposes a new class of line search filter improved secant methods for general nonlinear equality constrained optimization. The feature of these new algorithms is that one of the improved secant algorithms is used to produce a search direction, a backtracking line search procedure to generate step size, some filtered rules to determine step acceptance, second order correction technique to reduce infeasibility and overcome the Maratos effects. Under mild assumptions the global convergence is established. Moreover, it is also established that the Maratos effect are overcome in our new approaches by adding second order correction steps so that two-step Q-superlinear convergence to second order sufficient local solution is achieved. The results of numerical experiments are reported to show the effectiveness of these proposed algorithms.

Key words constrained optimization; filter method; secant algorithm; Maratos effect; second order correction

MR(2000) Subject Classification 65K05; 90C30; 90C51; 90C55

Chinese Library Classification O221.2