线性无关假设如何影响非线性约束最优化算法

刘新为

河北工业大学理学院, 天津 300401
E-mail: mathlxw@hebut.edu.cn

收稿日期: 2011-12-26; 接受日期: 2012-10-10

国家自然科学基金 (批准号: 10971047 和 11271107) 和河北省自然科学基金 (批准号: A2010000011) 资助项目

摘要 首先综述非线性约束最优化最近的一些进展. 首次定义了约束最优化算法的全局收敛性. 注意到最优性条件的精确性和算法近似性之间的差异, 并回顾等式约束最优化的原始的 Newton 型算法框架, 即可理解为什么约束梯度的线性无关假设应该而且可以被弱化. 这些讨论被扩展到不等式约束最优化问题. 然后在没有线性无关假设条件下, 证明了一个使用精确罚函数和二阶校正技术的算法可具有超线性收敛性. 这些认知有助于接下来开发求解包括非线性半定规划和锥规划等约束最优化问题的更加有效的新算法.

关键词 逐步二次规划 内点方法 线性无关性 局部收敛性 二阶校正技术

MSC (2010) 主题分类 90C30, 90C31, 65K05, 65K10

1 引言

为简单起见, 首先考虑等式约束最优化问题

$$\min f(x) \quad \text{s.t.} \quad c(x) = 0, \quad (1.1)$$

其中 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 和 $c: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 是连续可微函数, 且假定 c 中至少有一个分量函数 c_i 是非线性函数. 问题 (1.1) 的一阶最优性条件可以在任何一本非线性规划教科书上找到 (例如, [1–3]), 为了本文的完整性, 重述如下:

定理 1 假定 x^* 是问题 (1.1) 的一个局部极小点. 如果 $\nabla c(x^*) \in \mathbb{R}^{n \times m}$ 列满秩 (即向量组 $\nabla c_1(x^*), \dots, \nabla c_m(x^*)$ 线性无关), 那么存在唯一的 Lagrange 乘子向量 $\lambda^* \in \mathbb{R}^m$ 使得

$$\nabla f(x^*) + \nabla c(x^*)\lambda^* = 0. \quad (1.2)$$

定理 1 是非线性规划的一个基本结果, 它是非线性规划迭代方法的基础, 并提供了迭代终止的一个度量. 问题 (1.1) 的一个满足条件 (1.2) 的可行点称为它的一个 KKT (Karush-Kuhn-Tucker) 点. Lagrange 乘子向量的第 i 个分量也称为与约束函数 c_i 对应的在点 x^* 处的 Lagrange 乘子.

然而, 在求解问题 (1.1) 的所有迭代方法的实际运行中, 一般找不到一个精确满足方程 (1.2) 的解. 这意味着所有求解问题 (1.1) 的算法常常终止于某个被认为是 x^* 的近似的迭代点 x_k , 且满足

$$\|\nabla f(x_k) + \nabla c(x_k)\lambda_k\| \leq \epsilon, \quad (1.3)$$

英文引用格式: Liu X W. How does the linear independence assumption affect algorithms of nonlinear constrained optimization (in Chinese). Sci Sin Math, 2013, 43: 105–120, doi: 10.1360/012011-1037

其中 λ_k 可看成是 Lagrange 乘子向量 λ^* 的估计, $\epsilon > 0$ 是终止参数, $\|\cdot\|$ 可以是任何向量范数. 因此, 基于下列认知, 一个算法常常认为是全局收敛的.

定义 1 假定 $\{x_k\}$ 是由算法产生的一个迭代序列. 如果

$$\lim_{k \rightarrow \infty} \|c(x_k)\| = 0,$$

且对于每个 x_k , 存在一个向量 λ_k 使得

$$\liminf_{k \rightarrow \infty} \|\nabla f(x_k) + \nabla c(x_k)\lambda_k\| = 0, \quad (1.4)$$

那么对任意给定的 $\epsilon > 0$, 条件 (1.3) 将在有限次迭代内满足. 这时称算法是全局收敛的, 且称序列 $\{x_k\}$ 全局收敛到一个 KKT 点.

一个更强的全局收敛结果是由下列极限取代 (1.4):

$$\lim_{k \rightarrow \infty} \|\nabla f(x_k) + \nabla c(x_k)\lambda_k\| = 0. \quad (1.5)$$

尽管在文献中有大量的关于约束最优化的理论和算法的研究工作 (例如, [4] 在弱约束规范条件下证明了一个增广 Lagrange 函数法的全局收敛性, [5] 考虑了没有迭代序列有界性假设下的逐步二次规划的全局收敛性, [6–10] 研究了等式约束最优化的新算法及其全局收敛性, [11–18] 及 [19–23] 分别证明了论文中所提出的线搜索和信赖域内点方法的全局收敛性, [24–26] 证明了他们的方法具有强全局收敛性质), 但是却找不到一个关于约束最优化算法全局收敛性的明确的定义, 这是一个有趣的现象. 而无约束最优化是有这样的定义的, 可见 [2, 第 45 页]. 定义 1 的重要性在于它与迭代点直接相关. 它表明在约束最优化中可行性优先于稳定性, 以及在迭代中原始变量与对偶乘子估计之间的关系. 除了在定义 1 中描述的全局收敛性以外, 在文献中关于约束最优化还有另一类全局收敛性, 是基于寻找 KKT 方程组的近似解, 如 [27].

尽管向量组 $\nabla c_1(x^*), \dots, \nabla c_m(x^*)$ 线性无关的概念是明确和清楚的, 且由 $\nabla c(x)$ 的连续性, 当 x_k 充分靠近 x^* 时, 可知向量组 $\nabla c_1(x_k), \dots, \nabla c_m(x_k)$ 也是线性无关的, 但它没有反映出迭代过程中这些约束之间的差异, 比如有些约束可能有大的残差且难以满足, 而其它的约束是近似可行的或是容易满足的 (比如线性约束). 放在迭代过程中考虑, 诸如向量 $v_1 = (1, 0)^T$ 和 $v_2 = (1, \epsilon_k)^T$ 在 $\epsilon_k \neq 0$ 时总是线性无关的, 但如果 $k \rightarrow \infty$ 时 $\epsilon_k \rightarrow 0$, 那么 v_1 和 v_2 之间的线性无关性在极限点处就不成立了. 注意到 Jacobi 矩阵 $\nabla c(x_k)$ 列满秩当且仅当它有 m 个非零奇异值, 或等价地, 方阵 $\nabla c(x_k)^T \nabla c(x_k)$ 是非奇异的且它有 m 个正特征值. 我们的问题是: 当算法仅仅终止在满足 (1.3) 的一个近似解 x_k 时, 线性无关假设是否仍然是必要的? 最近的研究工作 [8, 25] 中的技术结果表明, 答案是否定的. 而且这些研究表明, 为了找一个满足 (1.3) 的近似解 x_k , 只要 $\nabla c(x_k)^T \nabla c(x_k)$ 在约束残差向量 $c(x_k)$ 的方向上正定就可以了, 显然这是比线性无关假设更弱得多的一个条件.

不等式约束最优化的情形更加复杂, 这是因为它的线性无关假设与最优解处的积极约束有关. 在迭代靠近最优解之前, 并不知道哪个是积极约束. 这对于开发求解不等式约束最优化问题的有效算法可能造成困难 (参见 [28–34]). 对诸如非线性半定规划和锥规划等一些具有特殊结构并在实际领域具有广泛应用的约束最优化问题来说, 一个类似于 Slater 约束规范条件的假设 (这个假设不易检验) 常常被使用 (参见 [35–41], 但 [42] 是一个例外). 而平衡约束最优化问题在最优解处约束梯度线性无关假设从来就不成立 (参见 [43–46]). 通过使用内点技术, 类似于等式约束最优化的处理方法, 论文 [25] 给出了带有不等式约束的一般最优化的一个更弱的替代条件.

为了明白线性无关假设在算法中的作用, 我们来回顾求解问题 (1.1) 的原始的 Newton 型算法框架 (它也是求解问题 (1.1) 的逐步二次规划方法的原始的算法框架).

算法 A1 (一个逐步二次规划方法的算法框架)

给定初始点 $x_0 \in \mathbb{R}^n$, 初始乘子向量估计 $\lambda_0 \in \mathbb{R}^m$, 和初始正定 Hesse 矩阵近似 $B_0 \in \mathbb{R}^{n \times n}$. 计算 $g_0 = \nabla f(x_0)$, $c_0 = c(x_0)$, $A_0 = \nabla c(x_0)$. 令 $k := 0$;

如果 $\|c_k\| > \epsilon$ 或者 $\|g_k + A_k \lambda_k\| > \epsilon$, 进入循环

求解二次规划子问题

$$\min g_k^T d + \frac{1}{2} d^T B_k d \quad \text{s.t.} \quad c_k + A_k^T d = 0, \quad (1.6)$$

(或等价地, 求解下列线性方程组, $B_k d + A_k v = -(g_k + A_k \lambda_k)$, $A_k^T d = -c_k$).

令 d_k 和 v_k 是它的解, $\lambda_{k+1} = \lambda_k + v_k$ 是新的乘子向量估计.

计算一个适当的步长 $\alpha_k \in (0, 1]$ 并令 $x_{k+1} = x_k + \alpha_k d_k$.

计算

$$g_{k+1} = \nabla f(x_{k+1}), \quad c_{k+1} = c(x_{k+1}), \quad A_{k+1} = \nabla c(x_{k+1}),$$

并修正 B_k 为 B_{k+1} . 令 $k := k + 1$;

结束循环;

返回 $x^* = x_k$, 算法终止.

在文献中有大量的关于在非线性规划算法中如何选取步长的工作, 许多证明有效的线搜索方法已经被提出. 这些方法可以粗略地分成三类, 即罚方法, 滤方法, 和最近开发的不使用罚函数或滤技术的方法. 本文并不打算提出一个新的实用的方法, 因此算法 A1 中没有给出具体的线搜索方法. 本文关注的是约束梯度线性无关假设如何影响子问题 (1.6) 以及算法的全局和局部收敛性质. 对算法 A1 来说, 没有约束梯度线性无关假设的一个直接的结果是子问题 (1.6) 可能不可行. 对于罚方法, 它还可以引起罚参数趋于无穷大, 以及序列 $\{\|d_k\|\}$ 和 $\{\|\lambda_k\|\}$ 可能无界, 接下来引起 α_k 变得任意小. 滤方法精巧地把所有难的情况归结于一个恢复阶段, 这一做法使得滤方法可以避免约束梯度线性无关假设, 参见 [47–53].

最近的研究 [25] 揭示了罚参数与序列 $\{\|d_k\|\}$ 和 $\{\|\lambda_k\|\}$ 的有界性之间的关系, 并证实约束梯度线性无关假设可以显著地弱化. [25] 证明了, 只要矩阵 $A_k^T A_k$ 在约束残差 c_k 的方向上一致正定, 那么罚参数在有限步迭代之后将保持常数, 而且序列 $\{A_k \lambda_k\}$ 是有界的. 由于 d_k 在 A_k^T 的零空间的投影 p_k 的有界性不依赖于约束梯度线性无关假设, $d_k = q_k + p_k$ 的有界性将可以保证, 倘若 q_k 是有界的. 此外, q_k 在减少约束残差 $\|c_k\|$ 获得迭代可行性中起着重要的作用. 结合内点技术, [25] 得到了带有等式和不等式约束最优化的相关结果.

没有假定约束梯度 $\nabla c_1(x^*), \dots, \nabla c_m(x^*)$ 的线性无关性, [8, 25] 已证明, 如果 Lagrange 乘子向量估计 λ_k 满足适当的条件, 则 d_k 仍然是一个局部超线性收敛步. 本文对这些条件进行进一步的提炼, 目的是希望能够找出关于 Lagrange 乘子向量估计 λ_k 的可以实现的条件来得到局部快速收敛性. 关于弱条件下的局部收敛性的研究可参见 [54–57]. 没有假定约束梯度线性无关, 通过引进二阶校正技术, [8] 证明了该文提出的不使用罚函数或滤技术的方法是局部超线性收敛的. 本文证明, 通过引进二阶校正技术, 对于最优化问题 (1.1), [25] 中的超线性收敛步能够被精确罚函数所接受, 因此罚方法也是超线性收敛的.

本文结构如下. 第 2 节给出了 $d_k = q_k + p_k$ 在没有任何线性无关假设下的关于 q_k 和 p_k 的一些技术结果, 描述了计算 q_k 的 Steihaug 共轭梯度方法, 给出了相关的全局收敛定理并把有关讨论扩展到

不等式约束最优化. 在第 3 节中, 对于问题 (1.1), 引进二阶校正技术并证明 [25] 中的超线性收敛步将被精确罚函数所接受. 最后一节给出本文的结论.

2 一些技术结果与全局收敛性

假定 x_k 是当前迭代点. 考虑子问题

$$\min \|c_k + A_k^T d\| \quad \text{s.t.} \quad \|d\| \leq \xi \|A_k c_k\|, \quad (2.1)$$

其中 $\xi > 0$ 是一个常数.

引理 1 假定 $q_k \in \mathbb{R}^n$ 是子问题 (2.1) 的一个满足下列条件的近似解:

$$\|c_k\| - \|c_k + A_k^T q_k\| \geq \gamma (\|c_k\| - \|c_k + A_k^T d_k^c\|), \quad (2.2)$$

其中 d_k^c 是子问题 (2.1) 的 Cauchy 点, $\gamma \in (0, 1]$ 是常数. 那么

$$\|c_k\| - \|c_k + A_k^T q_k\| \geq \frac{1}{2} \gamma \min \left\{ \xi, \frac{\|A_k c_k\|^2}{\|A_k^T A_k c_k\|^2} \right\} \frac{\|A_k c_k\|^2}{\|c_k\|}. \quad (2.3)$$

证明 本结论的证明类似于 [25, 引理 2.1]. 证毕. $\square$

问题 (2.1) 是一个凸最优化问题, 它可以由如下 Steihaug 共轭梯度法 (见 [2, 算法 4.3]) 近似求解, 其中为简单起见省略了迭代指标 k .

算法 A2 (求解问题 (2.1) 的 CG-Steihaug 方法)

给定 $\epsilon > 0$, 令 $v_0 = 0$, $r_0 = Ac$, $w_0 = -r_0$, $j := 0$.

如果 $\|r_0\| < \epsilon$, 返回 $q = v_0$, 终止算法;

如果 $\|A^T w_j\| \neq 0$, 进入循环

令 $\alpha_j = \frac{\|r_j\|^2}{\|A^T w_j\|^2}$, $v_{j+1} = v_j + \alpha_j w_j$;

如果 $\|v_{j+1}\| > \xi \|Ac\|$,

令 $\tau = \frac{\sqrt{(v_j^T w_j)^2 + \|w_j\|^2 (\xi^2 \|Ac\|^2 - \|v_j\|^2)} - v_j^T w_j}{\|w_j\|^2}$, 返回 $q = v_j + \tau w_j$, 终止算法;

令 $r_{j+1} = r_j + \alpha_j A A^T w_j$;

如果 $\|r_{j+1}\| < \epsilon \|r_0\|$, 返回 $q = v_{j+1}$, 终止算法;

令 $\beta_j = \frac{\|r_{j+1}\|^2}{\|r_j\|^2}$, $w_{j+1} = -r_{j+1} + \beta_j w_j$;

令 $j := j + 1$;

结束循环;

返回 $q = v_j$, 算法终止.

引理 2 假定序列 $\{v_j\}$ 由算法 A2 得到, $\text{rank}(A) = \ell \leq m$.

(1) 序列终止于第 j 次迭代, $j \leq \ell$, 且

$$0 = \|v_0\| < \cdots < \|v_j\| < \|v_{j+1}\| < \cdots < \|q\| \leq \xi \|Ac\|, \quad (2.4)$$

$$\|c\| = \|c + A^T v_0\| > \cdots > \|c + A^T v_j\| > \|c + A^T v_{j+1}\| > \cdots > \|c + A^T q\|; \quad (2.5)$$

(2) $\|q\| = \xi \|Ac\|$ 或者 $A(c + A^T q) = 0$.

证明 (1) 假定序列不在 $j \leq \ell$ 时终止. 那么

$$\|A^T w_j\| \neq 0, \quad j = 0, 1, \dots, \ell; \quad w_i^T A A^T w_j = 0, \quad i \neq j, \quad i, j = 0, 1, \dots, \ell.$$

因此, $A^T w_0, A^T w_1, \dots, A^T w_\ell$ 线性无关. 这与 $\text{rank}(A) = \ell$ 矛盾, 因为 A 的秩是 ℓ 说明在 A^T 的值空间中至多有 ℓ 个线性无关的向量.

不等式 (2.4) 能够类似于 [2, 定理 4.2 的证明] 得到.

下面证明 (2.5). 由于 $r_i^T r_0 = 0, i = 1, \dots, j$, 所以

$$r_0^T w_j = r_0^T (-r_j + \beta_{j-1} w_{j-1}) = \beta_{j-1} r_0^T w_{j-1} = \dots = \beta_{j-1} \dots \beta_0 r_0^T w_0 = -\|r_j\|^2.$$

注意到 $v_j = v_0 + \sum_{i=0}^{j-1} \alpha_i w_i$, 所以

$$\begin{aligned} \|c + A^T v_{j+1}\|^2 - \|c + A^T v_j\|^2 &= 2v_j^T (r_{j+1} - r_j) + 2\alpha_j r_0^T w_j + \alpha_j^2 \|A^T w_j\|^2 \\ &= \alpha_j (2r_0^T w_j + \alpha_j \|A^T w_j\|^2) \\ &= -\alpha_j \|r_j\|^2 < 0. \end{aligned}$$

(2) 因为 $r_j = r_{j-1} + \alpha_{j-1} A A^T w_{j-1} = r_0 + \sum_{i=0}^{j-1} \alpha_i A A^T w_i = A(c + A^T v_j)$, 故

$$r_j = 0 \quad \text{当且仅当} \quad A(c + A^T v_j) = 0. \quad (2.6)$$

由于 $r_j = A(c + A^T v_j)$, 如果 $A^T w_j = 0$, 则 $w_j^T r_j = 0$. 因此, 如果 $A^T w_j = 0$, 那么

$$\|r_j\|^2 = r_j^T (-w_j + \beta_{j-1} w_{j-1}) = 0. \quad (2.7)$$

注意到算法会在 $A^T w_j = 0$, 或 $r_j = 0$, 或 $\|v_j\| > \xi \|Ac\|$ 三种情形下终止. 若 $\|v_j\| > \xi \|Ac\|$, 则由 τ 的计算, 有 $\|q\| = \xi \|Ac\|$; 否则, 由 (2.6) 和 (2.7) 可得 $r_j = 0$ 以及 $A(c + A^T v_j) = 0$. 证毕. $\square$

引理 2 表明, 只要 ξ 足够大且 A_k 满秩, 算法 A2 将可以找到一个满足 $c_k + A_k^T q_k = 0$ 的解 q_k .

引理 3 假定 $B_k \in \mathbb{R}^{n \times n}$ 对所有的 k 在 A_k^T 的零空间一致正定. 考虑子问题

$$\min (g_k + B_k q_k)^T p + \frac{1}{2} p^T B_k p \quad \text{s.t.} \quad A_k^T p = 0. \quad (2.8)$$

(1) 问题 (2.8) 总有唯一解;

(2) 令 p_k 是问题 (2.8) 的解. Lagrange 乘子向量估计 λ_{k+1} 满足 $B_k(p_k + q_k) + A_k \lambda_{k+1} + g_k = 0$. 那么, 只要 $\{g_k\}$, $\{B_k\}$ 和 $\{q_k\}$ 有界, 则 $\{p_k\}$ 和 $\{A_k \lambda_{k+1}\}$ 有界.

证明 注意到问题 (2.8) 与下列问题等价:

$$\min g_k^T d + \frac{1}{2} d^T B_k d \quad \text{s.t.} \quad A_k^T (d - q_k) = 0. \quad (2.9)$$

本结论可由 [25, 引理 2.2 和 2.3] 立即得到. 证毕. $\square$

定理 2 (全局收敛性定理) 假定无穷序列 $\{x_k\}$ 是由 $x_{k+1} = x_k + \alpha_k d_k$ 产生, 其中 $d_k = q_k + p_k$, α_k 由 Armijo 线搜索经由精确罚函数或者不使用罚函数或者滤技术方法得到. 假定 $\{x_k\}$ 是有界序列, B_k 对所有的 k 是有界的且在 A_k^T 的零空间一致正定. 如果存在常数 $\eta > 0$ 使得

$$\|A_k c_k\| \geq \eta \|c_k\| \quad (2.10)$$

对所有 $k \geq 0$ 成立, 则 $\{x_k\}$ 全局收敛到问题 (1.1) 的 KKT 点; 否则, 存在 $\{x_k\}$ 的一个极限点 x^* 和一个单位向量 w^* , 使得 $A^*w^* = 0$, 其中 $A^* = \nabla c(x^*)$.

证明 问题 (1.1) 是 [25] 中考虑的问题在 $m_I = 0$ 和 $\mathcal{I} = \emptyset$ 时的特殊情形. 对于使用精确罚函数的算法, 结论已由 [25] 中的定理 4.9 证明. 尽管问题 (2.1) 比 [8] 中的更一般, 关于不使用罚函数或者滤技术方法的结论可以与 [8] 中的定理 3.8 一样得到. 证毕. $\square$

对于不使用罚函数或者滤技术的方法, 由 [8] 知, 序列 $\{x_k\}$ 的有界性假设可以由假定水平集 $\{x : \|c(x)\| \leq \max(\|c(x_0)\|, \nu_0)\}$ 的有界性取代, 其中 $\nu_0 > 0$ 是任意给定的常数.

本节的后面部分, 考虑不等式约束最优化问题

$$\min f(x) \quad \text{s.t.} \quad c(x) \leq 0. \quad (2.11)$$

引进松弛向量 $y \in \mathbb{R}^m$, 问题 (2.11) 可以等价地转化为具有等式约束和非负约束的最优化问题

$$\min f(x) \quad \text{s.t.} \quad c(x) + y = 0, \quad y \geq 0.$$

为了求解上述问题, 考虑下列等式约束的参数最优化问题

$$\min f(x) - \mu e^T \log y \quad \text{s.t.} \quad c(x) + y = 0, \quad (2.12)$$

其中 $\mu > 0$ 是参数, $e \in \mathbb{R}^m$ 是所有分量是 1 的向量, $\log y$ 是一个向量, 它的第 i 个分量是 $\log y_i$.

定义 $z = (x^T, y^T)^T$, $F(z) = f(x) - \mu e^T \log y$, 和 $h(z) = c(x) + y$. 那么

$$\nabla F(z) = \begin{bmatrix} \nabla f(x) \\ -\mu Y^{-1} e \end{bmatrix}, \quad \nabla h(z) = \begin{bmatrix} \nabla c(x) \\ I_m \end{bmatrix},$$

其中 $Y = \text{diag}(y)$. 对于给定的 $\mu > 0$, 假定当前迭代点 $z_k = (x_k^T, y_k^T)^T$ 是一个内点, 即 $y_k > 0$. 令

$$T_k = \begin{bmatrix} I & \\ & Y_k \end{bmatrix},$$

其中 I 是 $n \times n$ 单位矩阵, $Y_k = \text{diag}(y_k)$. 计算 q_k 的子问题被改写成

$$\min \|h_k + \nabla h_k^T d\| \quad \text{s.t.} \quad \|T_k^{-1} d\| \leq \xi \|T_k \nabla h_k h_k\|. \quad (2.13)$$

问题 (2.13) 不是等式约束最优化子问题 (2.1) 的直接应用. 它是在 [25] 的收敛性分析基础上得到的. 为了应用算法 A2, 问题 (2.13) 能够等价地转化为

$$\min \|h_k + (T_k \nabla h_k)^T v\| \quad \text{s.t.} \quad \|v\| \leq \xi \|T_k \nabla h_k h_k\|. \quad (2.14)$$

令 v_k 是它的解, 则 $q_k = T_k v_k$.

类似于引理 1 和 3, 对于不等式约束最优化可以证明下列结论:

引理 4 假定 $q_k \in \mathbb{R}^{n+m}$ 是问题 (2.13) 的满足下列条件的近似解:

$$\|h_k\| - \|h_k + \nabla h_k^T q_k\| \geq \gamma (\|h_k\| - \|h_k + \nabla h_k^T d_k^c\|), \quad (2.15)$$

其中 $d_k^c = T_k v_k^c$, v_k^c 是问题 (2.14) 的 Cauchy 点, $\gamma \in (0, 1]$ 是一个常数. 那么

$$\|h_k\| - \|h_k + \nabla h_k^T q_k\| \geq \frac{1}{2} \gamma \min \left\{ \xi, \frac{\|T_k \nabla h_k h_k\|^2}{\|\nabla h_k^T T_k^2 \nabla h_k h_k\|^2} \right\} \frac{\|T_k \nabla h_k h_k\|^2}{\|h_k\|}. \quad (2.16)$$

引理 5 假设 $B_k \in \mathbb{R}^{(n+m) \times (n+m)}$ 对所有的 k 在 ∇h_k^T 的零空间一致正定. 考虑子问题

$$\min (\nabla F_k + B_k q_k)^T p + \frac{1}{2} p^T B_k p \quad \text{s.t.} \quad \nabla h_k^T p = 0. \quad (2.17)$$

(1) 问题 (2.17) 总有唯一解;

(2) 令 p_k 是问题 (2.17) 的解, Lagrange 乘子向量估计 λ_{k+1} 满足 $B_k(p_k + q_k) + \nabla h_k \lambda_{k+1} + \nabla F_k = 0$. 如果 $\{T_k \nabla F_k\}$, $\{T_k B_k T_k\}$ 和 $\{T_k^{-1} q_k\}$ 是有界的, 则 $\{T_k^{-1} p_k\}$ 和 $\{T_k \nabla h_k \lambda_{k+1}\}$ 有界.

下列结论是 [25, 定理 4.9] 在 $m = m_I$ 和 $\mathcal{E} = \emptyset$ 的特殊情形.

定理 3 (全局收敛性定理) 对任意给定的 $\mu > 0$, 假定 $\{z_k\}$ 是由 $z_{k+1} = z_k + \alpha_k d_k$ 得到的一个无穷序列, 其中 $d_k = q_k + p_k$, α_k 由 Armijo 线搜索经由使用精确罚函数的内点方法产生. 还假定 $\{x_k\}$ 是有界序列, B_k 对所有的 k 是有界的且在 ∇h_k^T 的零空间一致正定. 如果存在常数 $\eta > 0$ 使得

$$\|T_k \nabla h_k h_k\| \geq \eta \|h_k\| \quad (2.18)$$

对所有 $k \geq 0$ 成立, 那么 $\{z_k\}$ 全局收敛到问题 (2.12) 的 KKT 点. 否则, 存在 $\{z_k\}$ 的一个极限点 $z^* = (x^{*T}, y^{*T})^T$ 和单位向量 w^* 满足 $A^* w^* = 0$ 和 $Y^* w^* = 0$, 其中 $A^* = \nabla c(x^*)$, $Y^* = \text{diag}(y^*)$.

比较问题 (2.12) 的 KKT 条件

$$g^* + A^* \lambda^* = 0, \quad y^* \circ \lambda^* = \mu e, \quad c^* + y^* = 0, \quad y^* \geq 0, \quad \lambda^* \geq 0,$$

和不等式约束问题 (2.11) 的 KKT 条件

$$g^* + A^* \lambda^* = 0, \quad \lambda^* \circ c^* = 0, \quad c^* \leq 0, \quad \lambda^* \geq 0,$$

可以看出, 当 $\mu > 0$ 充分小时, 问题 (2.12) 的 KKT 点常常可以看成是问题 (2.11) 的一个近似 KKT 点.

3 精确罚函数与局部收敛性

针对问题 (1.1), 定义精确罚函数

$$\phi(x; \rho) = f(x) + \rho \|c(x)\|, \quad (3.1)$$

其中 $\rho > 0$ 是罚参数. 在不假定约束梯度 $\nabla c_1(x^*), \dots, \nabla c_m(x^*)$ 线性无关的情况下, 本节考虑以 $d_k = q_k + p_k$ 为搜索方向, 以 $\phi(x; \rho)$ 为效益函数, 并由 Armijo 线搜索方法选取步长的算法的局部收敛性. 没有线性无关假设, [25] 已证明在适当条件下 $d_k = q_k + p_k$ 是一个超线性收敛步, 但它没有证明 d_k 可以被 $\phi(x; \rho)$ 接受, 因此没有证明算法具有局部超线性收敛性. 通过引进二阶校正技术, [8] 证明了所提出的不使用罚函数或滤技术的算法具有局部超线性收敛性. 本文的证明是在比 [8, 25] 更一般的局部假设基础上完成的.

为了叙述方便, 首先详细给出局部算法框架如下:

算法 A3 (使用精确罚函数的算法框架)

给定 $x_0 \in \mathbb{R}^n$, $B_0 \in \mathbb{R}^{n \times n}$, 和 $\rho > 0$, $\sigma > 0$, $\delta \in (0, 1/2)$, $\xi > 0$. 计算 $g_0 = \nabla f(x_0)$, $c_0 = c(x_0)$, $A_0 = \nabla c(x_0)$.

For $k = 0, 1, \dots$,

计算 $d_k = q_k + p_k$, 其中 q_k 和 p_k 分别由子问题 (2.1) 和 (2.8) 产生;

修正 ρ 使得

$$g_k^T d_k + (\rho/2)(\|c_k + A_k^T d_k\| - \|c_k\|) \leq -\sigma \|d_k\|^2;$$

如果 $\phi(x_k + d_k; \rho) - \phi(x_k; \rho) \leq \delta\{g_k^T d_k + \rho(\|c_k + A_k^T d_k\| - \|c_k\|)\}$, 令 $x_{k+1} = x_k + d_k$,

否则由下列子问题计算二阶校正步 $s_k = \hat{q}_k + \hat{p}_k$: 求解子问题

$$\min \quad \|\hat{c}_k + A_k^T d\| \quad \text{s.t.} \quad \|d\| \leq \xi \|A_k c_k\|, \quad (3.2)$$

得到 $\hat{q}_k$ (其中 $\hat{c}_k = c(x_k + d_k)$), 再由子问题

$$\min \quad (g_k + B_k d_k + B_k \hat{q}_k)^T p + \frac{1}{2} p^T B_k p \quad \text{s.t.} \quad A_k^T p = 0 \quad (3.3)$$

得到 $\hat{p}_k$, 若

$$\phi(x_k + d_k + s_k; \rho) - \phi(x_k; \rho) \leq \delta\{g_k^T d_k + \rho(\|c_k + A_k^T d_k\| - \|c_k\|)\}, \quad (3.4)$$

则令 $x_{k+1} = x_k + d_k + s_k$, 否则由 Armijo 线搜索选择步长 $\alpha_k \in (0, 1]$ 使得

$$\phi(x_k + \alpha_k d_k; \rho) - \phi(x_k; \rho) \leq \delta\alpha_k\{g_k^T d_k + \rho(\|c_k + A_k^T d_k\| - \|c_k\|)\}, \quad (3.5)$$

且令 $x_{k+1} = x_k + \alpha_k d_k$;

计算 $g_{k+1} = \nabla f(x_{k+1})$, $c_{k+1} = c(x_{k+1})$, $A_{k+1} = \nabla c(x_{k+1})$, 且修正 B_k 为 B_{k+1} ;

end (For)

为了证明算法 A3 的局部超线性收敛性, 需要下列局部条件:

假设 1 (a) $x_k \rightarrow x^*$ 当 $k \rightarrow \infty$, 其中 x^* 是问题 (1.1) 的一个 KKT 点, $\lambda^* \in \mathbb{R}^m$ 是相应的 Lagrange 乘子向量;

(b) $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 和 $c: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 在 $\mathbb{R}^n$ 上二阶连续可微, 且二阶导数在 x^* 处 Lipschitz 连续;

(c) 梯度集 $\{\nabla c_i(x^*) : i \in J_*\}$ 是由 $\{\nabla c_i(x^*) : i \in \{1, 2, \dots, m\}\}$ 中所有线性无关向量组成的集合. $(\lambda_{k+1})_i \rightarrow \lambda_i^*$ 当 $k \rightarrow \infty$, 其中 $i \notin J_*$ 且 $i = 1, \dots, m$, λ_{k+1} 是由问题 (2.8) 得到的 Lagrange 乘子向量估计;

(d) $d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \geq \beta \|d\|^2$, $\forall d \in \{d \neq 0 : \nabla c_i(x^*)^T d = 0, i \in J_*\}$, 其中 $L(x, \lambda) = f(x) + \lambda^T c(x)$, $\beta > 0$ 是一个常数.

(e) $\|c_k + A_k^T d_k\| \leq \tau_k \|c_k\|$ 和 $\|q_k\| \leq \kappa_0 \|c_k\|$, 其中 $\tau_k \in [0, 1]$, $\kappa_0 > 0$ 是一个常数.

假设 1(c) 对于 $J_* = \{1, 2, \dots, m\}$ 自然成立, 倘若 $\nabla c_i(x^*)$, $i = 1, 2, \dots, m$ 线性无关. 因为 $\|A_k\|$ 有界, 如果 $\|q_k\| \leq \xi \|A_k c_k\|$, 那么 $\|q_k\| \leq \kappa_0 \|c_k\|$.

引理 6 假定假设 1 成立. 如果 $\|(B_k - \nabla_{xx}^2 L(x^*, \lambda^*))d\| = o(\|d\|)$ 对每个 $d \in \mathbb{R}^n$ 成立, $\tau_k = o(1)$, 那么存在正的常数 ρ 和 σ 使得对充分大的 k ,

$$g_k^T d_k + (\rho/2)(\|c_k + A_k^T d_k\| - \|c_k\|) \leq -\sigma \|d_k\|^2.$$

证明 令 $P_k = I - A_{kJ_*}(A_{kJ_*}^T A_{kJ_*})^{-1} A_{kJ_*}^T$, 其中 A_{kJ_*} 是一个列是 $\nabla c_i(x_k)$, $i \in J_*$ 的矩阵. 则 $A_{kJ_*}^T P_k d_k = 0$ 和 $A_{kJ_*}^T (I - P_k) d_k = A_{kJ_*}^T d_k$. 因此,

$$\begin{aligned} d_k^T \nabla_{xx}^2 L(x^*, \lambda^*) d_k &= d_k^T P_k^T \nabla_{xx}^2 L(x^*, \lambda^*) P_k d_k + d_k^T (I - P_k)^T \nabla_{xx}^2 L(x^*, \lambda^*) (I + P_k) d_k \\ &\geq \beta \|P_k d_k\|^2 - \kappa \|(I - P_k) d_k\| \\ &\geq \beta \|d_k\|^2 - \kappa' \|c_k\|, \end{aligned}$$

其中 β 由假设 1(d) 给出, κ 和 κ' 是两个正常数且 $\kappa < \kappa'$.

如果 $\|(B_k - \nabla_{xx}^2 L(x^*, \lambda^*))d\| = o(\|d\|)$, 那么

$$d_k^T B_k d_k = d_k^T \nabla_{xx}^2 L(x^*, \lambda^*) d_k + o(\|d_k\|^2).$$

由于 d_k 是最优化问题 (2.8) 的解, 所以

$$g_k^T d_k = -d_k^T B_k d_k - \lambda_{k+1}^T A_k^T d_k.$$

因此,

$$\begin{aligned} g_k^T d_k + (\rho/2)(\|c_k + A_k^T d_k\| - \|c_k\|) + \sigma \|d_k\|^2 \\ &= -d_k^T B_k d_k - \lambda_{k+1}^T A_k^T d_k + (\rho/2)(\|c_k + A_k^T d_k\| - \|c_k\|) + \sigma \|d_k\|^2 \\ &\leq -\beta \|d_k\|^2 + (\kappa' + \kappa_0 \|A_k \lambda_{k+1}\|) \|c_k\| + (\rho/2)(\tau_k - 1) \|c_k\| + \sigma \|d_k\|^2 + o(\|d_k\|^2) \\ &\leq (\sigma - \beta) \|d_k\|^2 + \{\kappa'' - (\rho/2)(1 - o(1))\} \|c_k\| + o(\|d_k\|^2), \end{aligned}$$

其中由引理 3(2) 得 $\kappa' + \kappa_0 \|A_k \lambda_{k+1}\| \leq \kappa''$, κ'' 是一个常数. 因此, 通过选取 $\sigma < \beta$ 和 $\rho > 2\kappa''$ 可得结论. 证毕. $\square$

因为 $x_k \rightarrow x^*$, 引理 6 和 (3.5) 说明 $\|d_k\| \rightarrow 0$ 当 $k \rightarrow \infty$. 下列引理 7 和定理 4 可以类似地由 [8] 中的方法证明.

引理 7 假定假设 1 成立. 令 $P_k = I - A_{kJ_*}(A_{kJ_*}^T A_{kJ_*})^{-1} A_{kJ_*}^T$, 其中 A_{kJ_*} 是以 $\nabla c_i(x_k)$, $i \in J_*$ 为列的矩阵. 如果 $\|(B_k - \nabla_{xx}^2 L(x^*, \lambda^*))d\| = o(\|d\|)$ 对每个 $d \in \mathbb{R}^n$ 成立, 那么存在常数 $\eta > 0$ 使得对每个 $d \in \mathbb{R}^n$ 和所有充分大的 k ,

$$\left\| \begin{bmatrix} P_k \nabla_{xx}^2 L(x^*, \lambda^*) \\ A_{kJ_*}^T \end{bmatrix} d \right\| \geq \eta \|d\|. \quad (3.6)$$

证明 它可以类似于 [8] 中定理 4.3 的第一部分得到证明. 证毕. $\square$

定理 4 假定假设 1 成立. 如果 $\|(B_k - \nabla_{xx}^2 L(x^*, \lambda^*))d\| = o(\|d\|)$, 且 $\tau_k = o(1)$, 则

$$\lim_{k \rightarrow \infty} \|x_k + d_k - x^*\| / \|x_k - x^*\| = 0. \quad (3.7)$$

证明 令 $P_* = I - A_{*J_*}(A_{*J_*}^T A_{*J_*})^{-1} A_{*J_*}^T$, 其中 A_{*J_*} 是一个矩阵, 它的列是 $\nabla c_i(x^*)$, $i \in J_*$. 定义 $A^* = \nabla c(x^*)$, 则 $P_* A^* = 0$. 因此, 由于假设 1(c),

$$\begin{aligned} P_k A_k (\lambda_{k+1} - \lambda^*) \\ &= (P_k A_k - P_* A^*) (\lambda_{k+1} - \lambda^*) \end{aligned}$$

$$\begin{aligned}
 &= P_k \sum_{i \notin J_*} (\nabla c_i(x_k) - \nabla c_i(x^*))((\lambda_{k+1})_i - \lambda_i^*) + (P_k - P_*) \sum_{i \notin J_*} \nabla c_i(x^*)((\lambda_{k+1})_i - \lambda_i^*) \\
 &= o(\|x_k - x^*\|).
 \end{aligned}$$

注意到 $g_k + B_k d_k + A_k \lambda_{k+1} = 0$, 可得

$$\begin{aligned}
 P_k B_k d_k &= -P_k(g_k + A_k \lambda^*) - P_k A_k(\lambda_{k+1} - \lambda^*) \\
 &= -P_k(g_k + A_k \lambda^*) + o(\|x_k - x^*\|) \\
 &= -P_k \nabla_{xx}^2 L(x^*, \lambda^*)(x_k - x^*) + O(\|x_k - x^*\|^2) + o(\|x_k - x^*\|).
 \end{aligned}$$

因此,

$$P_k(B_k - \nabla_{xx}^2 L(x^*, \lambda^*))d_k = -P_k \nabla_{xx}^2 L(x^*, \lambda^*)(x_k + d_k - x^*) + O(\|x_k - x^*\|^2) + o(\|x_k - x^*\|). \quad (3.8)$$

由于 $c_{J_*}^* = 0$ 和

$$c_{kJ_*} = c_{kJ_*} - c_{J_*}^* = A_{kJ_*}^T(x_k - x^*) + O(\|x_k - x^*\|^2),$$

所以

$$A_{kJ_*}^T(x_k + d_k - x^*) = c_{kJ_*} + A_{kJ_*}^T d_k + O(\|x_k - x^*\|^2). \quad (3.9)$$

把 (3.8) 和 (3.9) 放在一起, 可得出

$$\begin{aligned}
 &\begin{bmatrix} P_k \nabla_{xx}^2 L(x^*, \lambda^*) \\ A_{kJ_*}^T \end{bmatrix} (x_k + d_k - x^*) \\
 &= \begin{bmatrix} -P_k(B_k - \nabla_{xx}^2 L(x^*, \lambda^*))d_k + o(\|x_k - x^*\|) \\ c_{kJ_*} + A_{kJ_*}^T d_k \end{bmatrix} + O(\|x_k - x^*\|^2). \quad (3.10)
 \end{aligned}$$

由引理 7 知, 对所有充分大的 k , 上述表达式中左边的系数矩阵是列满秩的.

而 $\tau_k = o(1)$, $\|c_k\| = \|c_k - c^*\| = O(\|x_k - x^*\|)$, 以及

$$\|c_{kJ_*} + A_{kJ_*}^T d_k\| \leq \|c_k + A_k^T d_k\| \leq \tau_k \|c_k\|,$$

故

$$\|c_{kJ_*} + A_{kJ_*}^T d_k\| = o(\|x_k - x^*\|).$$

如果 $\|(B_k - \nabla_{xx}^2 L(x^*, \lambda^*))d_k\| = o(\|d_k\|)$, 那么

$$\left\| \begin{bmatrix} -P_k(B_k - \nabla_{xx}^2 L(x^*, \lambda^*))d_k + o(\|x_k - x^*\|) \\ c_{kJ_*} + A_{kJ_*}^T d_k \end{bmatrix} \right\| = \sqrt{(o(\|d_k\|) + o(\|x_k - x^*\|))^2 + o(\|x_k - x^*\|^2)}.$$

因此, 由 (3.10) 且因为 $x_k \rightarrow x^*$ 当 $k \rightarrow \infty$, 可得

$$\lim_{k \rightarrow \infty} \|x_k + d_k - x^*\|/\|x_k - x^*\| = \lim_{k \rightarrow \infty} o(\|d_k\|)/\|x_k - x^*\|, \quad (3.11)$$

它表明

$$\lim_{k \rightarrow \infty} \|d_k\|/\|x_k - x^*\| = 1. \quad (3.12)$$

最后, 由 (3.11) 可得结论 (3.7). 证毕. $\square$

与假设 1(c) 和 1(e) 相对应, 还需要下列关于二阶校正子问题 (3.3) 的假设 (如果 $\nabla_{c_i}(x^*)$, $i = 1, 2, \dots, m$ 线性无关, 这些假设自然成立):

假设 2 (a) $(\lambda_{k+1})_i - (\mu_{k+1})_i \rightarrow 0$ 当 $k \rightarrow \infty$, 其中 $i \notin J_*$ 及 $i = 1, \dots, m$, $\mu_{k+1} \in \mathbb{R}^m$ 是子问题 (3.3) 的相应的 Lagrange 乘子向量;

(b) $\|c(x_k + d_k) + A_k^T s_k\| \leq \tau_k \|c(x_k + d_k)\|$, 其中 $\tau_k \in [0, 1]$.

在给出本节的主要结论之前, 还需要下列引理.

引理 8 如果假设 1 和 2 成立, 且对任何 $d \in \mathbb{R}^n$ 都有 $\|(B_k - \nabla_{xx}^2 L(x^*, \lambda^*))d\| = o(\|d\|)$, 以及 $\tau_k = o(1)$, 那么

$$\|s_k\| = o(\|d_k\|). \quad (3.13)$$

证明 由于

$$\|P_k B_k d\| \geq \|P_k \nabla_{xx}^2 L(x^*, \lambda^*)d\| - \|P_k (B_k - \nabla_{xx}^2 L(x^*, \lambda^*))d\|,$$

由引理 7 可知, 存在常数 $\hat{\eta} < \eta$ 使得

$$\left\| \begin{bmatrix} P_k B_k \\ A_{kJ_*}^T \end{bmatrix} d \right\| \geq \hat{\eta} \|d\| \quad (3.14)$$

对任何 $d \in \mathbb{R}^n$ 及所有充分大的 k 成立.

因为 s_k 是问题 (3.3) 的解, 所以

$$g_k + B_k(d_k + s_k) + A_k \mu_{k+1} = 0.$$

由于 $g_k + B_k d_k = -A_k \lambda_k$, 可得

$$B_k s_k = A_k \lambda_{k+1} - A_k \mu_{k+1}. \quad (3.15)$$

因此, 根据假设 2(a) 可得

$$P_k B_k s_k = o(\|x_k - x^*\|) = o(\|d_k\|), \quad (3.16)$$

其中第二个等式由 (3.12) 得到. 由假设 1(e) 和 2(b), 并注意到 $\|c_k\| = O(\|x_k - x^*\|) = O(\|d_k\|)$, 可得

$$\begin{aligned} \|A_{kJ_*}^T s_k\| &\leq \|c(x_k + d_k) + A_k^T s_k\| + \|c(x_k + d_k)\| \\ &\leq (1 + \tau_k) \|c(x_k + d_k)\| \\ &\leq (1 + \tau_k) (\|c_k + A_k^T d_k\| + O(\|d_k\|^2)) \\ &\leq (1 + \tau_k) \tau_k \|c_k\| + O(\|d_k\|^2). \end{aligned}$$

因此,

$$\|A_{kJ_*}^T s_k\| = o(\|d_k\|). \quad (3.17)$$

最后, 由 (3.14), (3.16) 和 (3.17) 可得所要的结论. 证毕. $\square$

下列定理表明, 对于充分大的 k , 超线性收敛步将被算法 A3 所接受.

定理 5 假定假设 1 和 2 成立. 如果对所有 $d \in \mathbb{R}^n$ 都有 $\|(B_k - \nabla_{xx}^2 L(x^*, \lambda^*))d\| = o(\|d\|)$, 且 $\delta < 1/2$, $\tau_k = o(1)$, $\rho \geq \|\lambda^*\|$, 那么对所有充分大的 k , $x_{k+1} = x_k + d_k$, 或者 $x_{k+1} = x_k + d_k + s_k$.

证明 为了证明上述结论, 只需证明 (3.4) 对充分大的 k 成立. 由

$$\nabla_x L(x_k, \lambda^*)^T d_k = (x_k - x^*)^T \nabla_{xx}^2 L(x^*, \lambda^*) d_k + O(\|x_k - x^*\|^2 \|d_k\|),$$

及关系式 (3.7) 和 (3.12), 可得

$$d_k^T \nabla_{xx}^2 L(x^*, \lambda^*) d_k = -g_k^T d_k - (\lambda^*)^T A_k^T d_k + o(\|d_k\|^2).$$

因此, 由引理 8,

$$\begin{aligned} f(x_k + d_k + s_k) - f(x_k) &= L(x_k + d_k + s_k, \lambda^*) - L(x_k, \lambda^*) - (\lambda^*)^T (c(x_k + d_k + s_k) - c(x_k)) \\ &= -\frac{1}{2} d_k^T \nabla_{xx}^2 L(x^*, \lambda^*) d_k - (\lambda^*)^T (c(x_k + d_k + s_k) - c(x_k)) + o(\|d_k\|^2) \\ &= \frac{1}{2} g_k^T d_k + \frac{1}{2} (\lambda^*)^T A_k^T d_k - (\lambda^*)^T (c(x_k + d_k + s_k) - c(x_k)) + o(\|d_k\|^2). \end{aligned}$$

再利用引理 3, 得到

$$\begin{aligned} &\phi(x_k + d_k + s_k; \rho) - \phi(x_k; \rho) - \delta \{g_k^T d_k + \rho(\|c_k + A_k^T d_k\| - \|c_k\|)\} \\ &= (1/2 - \delta) \{g_k^T d_k + \rho(\|c_k + A_k^T d_k\| - \|c_k\|)\} - (\rho/2)(\|c_k + A_k^T d_k\| - \|c_k\|) \\ &\quad + (1/2)(\lambda^*)^T A_k^T d_k - (\lambda^*)^T (c(x_k + d_k + s_k) - c_k) \\ &\quad + \rho(\|c(x_k + d_k + s_k)\| - \|c_k\|) + o(\|d_k\|^2) \\ &\leq -(1/2 - \delta) \sigma \|d_k\|^2 - (1/4 + \delta/2) \rho (\|c_k + A_k^T d_k\| - \|c_k\|) + (1/2)(\lambda^*)^T A_k^T d_k \\ &\quad - (\lambda^*)^T (c(x_k + d_k + s_k) - c_k) + \rho(\|c(x_k + d_k + s_k)\| - \|c_k\|) + o(\|d_k\|^2). \end{aligned}$$

注意到 $\|c_k + A_k^T d_k\| \leq \tau_k \|c_k\|$, 以及

$$\begin{aligned} \|c(x_k + d_k + s_k)\| &= \|c(x_k + d_k) + \nabla c(x_k + d_k)^T s_k + O(\|s_k\|^2)\| \\ &\leq \|c(x_k + d_k) + A_k^T s_k\| + O(\|s_k\|^2) + O(\|s_k\| \|d_k\|) \\ &\leq \tau_k \|c(x_k + d_k)\| + o(\|d_k\|^2) \\ &\leq \tau_k \|c_k + A_k^T d_k\| + o(\|d_k\|^2) \\ &\leq \tau_k^2 \|c_k\| + o(\|d_k\|^2), \end{aligned}$$

可得

$$(1/2)(\lambda^*)^T A_k^T d_k - (\lambda^*)^T (c(x_k + d_k + s_k) - c_k)$$

$$\begin{aligned}
&= (1/2)(\lambda^*)^T(c_k + A_k^T d_k) - (1/2)(\lambda^*)^T c_k - (\lambda^*)^T(c(x_k + d_k + s_k) - c_k) \\
&\leq (\tau_k/2 + \tau_k^2)\|\lambda^*\| \|c_k\| + (1/2)(\lambda^*)^T c_k + o(\|d_k\|^2).
\end{aligned}$$

因此, 可推出

$$\begin{aligned}
&\phi(x_k + d_k + s_k; \rho) - \phi(x_k; \rho) - \delta\{g_k^T d_k + \rho(\|c_k + A_k^T d_k\| - \|c_k\|)\} \\
&\leq -(1/2 - \delta)\sigma\|d_k\|^2 + (\tau_k/2 + \tau_k^2)\|\lambda^*\| \|c_k\| + \tau_k^2 \rho \|c_k\| \\
&\quad + (1/2)(\lambda^*)^T c_k + (1/4 + \delta/2)\rho\|c_k\| - \rho\|c_k\| + o(\|d_k\|^2).
\end{aligned}$$

最终, 当 $\rho \geq \|\lambda^*\|$, $\delta < 1/2$, $\tau_k = o(1)$,

$$\begin{aligned}
&\phi(x_k + d_k + s_k; \rho) - \phi(x_k; \rho) - \delta\{g_k^T d_k + \rho(\|c_k + A_k^T d_k\| - \|c_k\|)\} \\
&\leq -(1/2 - \delta)\sigma\|d_k\|^2 + (1/2)(\lambda^*)^T c_k - (3/4 - \delta/2)\rho\|c_k\| + o(\|d_k\|^2) + o(\|c_k\|) \\
&\leq 0.
\end{aligned}$$

从而由算法可知, $x_{k+1} = x_k + d_k + s_k$. 证毕. $\square$

4 结论

本文首次给出了约束最优化算法全局收敛性的定义. 首先综述了关于非线性等式约束最优化的一些最近的进展, 应用 Steihaug 共轭梯度法求解线性化约束残差极小化问题并证明了它的收敛性, 这些讨论也扩展到不等式约束最优化. 然后, 在没有约束梯度线性无关假设条件下, 证明了一个使用精确罚函数和二阶校正技术的算法具有局部超线性收敛性.

没有约束梯度的线性无关性假设, 与问题 (2.8) 相对应的满足 $B_k d_k + g_k + A_k \lambda_{k+1} = 0$ 的对偶乘子估计 λ_{k+1} 和问题 (3.3) 的满足 $B_k(d_k + s_k) + g_k + A_k \mu_{k+1} = 0$ 的 μ_{k+1} 就可能不是唯一的 (它们的存在性分别由问题 (2.8) 和 (3.3) 的可行性保证). 因此, 一个可执行的算法应该明确给出 λ_{k+1} 和 μ_{k+1} 如何得到. 局部收敛性条件假设 1(c) 说明, 倘若 λ^* 是已知的, λ_{k+1} 可以由下列子问题得出:

$$\min \|\lambda - \lambda^*\| \quad \text{s.t.} \quad B_k d_k + g_k + A_k \lambda = 0. \quad (4.1)$$

根据假设 2(a), μ_{k+1} 可以由下列子问题得出:

$$\min \|\mu - \lambda_{k+1}\| \quad \text{s.t.} \quad B_k(d_k + s_k) + g_k + A_k \mu = 0. \quad (4.2)$$

因为上述两个子问题 (4.1) 和 (4.2) 都是严格凸的, 由极限 $x_k \rightarrow x^*$, $d_k \rightarrow 0$ 和 $s_k \rightarrow 0$, 可得 $\lambda_{k+1} \rightarrow \lambda^*$ 和 $\mu_{k+1} - \lambda_{k+1} \rightarrow 0$ 当 $k \rightarrow \infty$. 困难在于在求解一个问题之前 λ^* 一般是未知的. 因此, 没有约束梯度的线性无关性假设, 如何产生 λ_{k+1} 来保证算法的局部快速收敛性仍然是一个有待解决的问题. 在关于 CUTE 测试问题集 (参见 [58]) 的数值实验中, 已经观察到利用最小二乘问题

$$\min \|g_k + A_k \lambda\|$$

来产生 λ_{k+1} 的算法有很好的数值表现 (例如, 参见 [8, 59]). 如果 A_k 在 k 充分大时仍然是列满秩的, 那么由 $\nabla c(x)$ 的连续性可知, 即使约束梯度在 x^* 处是线性相关的, 使用上述最小二乘问题的算法可以保证 λ_{k+1} 和 λ_k 在 k 充分大时充分靠近. 因此, 可以认为极限 $\lambda_{k+1} \rightarrow \lambda^*$ 是满足的. 然而, 直到目前仍然没有相关的理论结果来证实这一观察的现象.

参考文献

- 1 Conn A R, Gould N I M, Toint Ph L. Trust-Region Methods. MPS-SIAM Series on Optimization. Philadelphia: SIAM, 2000
- 2 Nocedal J, Wright S J. Numerical Optimization. New York: Springer, 1999
- 3 Sun W Y, Yuan Y X. Optimization Theory and Methods: Nonlinear Programming. New York: Springer, 2006
- 4 Andreani R, Birgin E G, Martinez J M, et al. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. Math Program, 2008, 111: 5–32
- 5 Solodov M. Global convergence of an SQP method without boundedness assumptions on any of the iterative sequences. Math Program, 2009, 118: 1–12
- 6 Byrd R H, Curtis F E, Nocedal J. An inexact SQP method for equality constrained optimization. SIAM J Optim, 2008, 19: 351–369
- 7 Gould N I M, Toint Ph L. Nonlinear programming without a penalty function or a filter. Math Program, 2009, 122: 155–196
- 8 Liu X W, Yuan Y X. A sequential quadratic programming method without a penalty function or a filter for nonlinear equality constrained optimization. SIAM J Optim, 2011, 21: 545–571
- 9 Powell M J D, Yuan Y X. A trust region algorithm for equality constrained optimization. Math Program, 1991, 49: 189–211
- 10 Ulbrich M, Ulbrich S. Non-monotone trust region methods for nonlinear equality constrained optimization without a penalty function. Math Program, 2003, 95: 103–135
- 11 Benson H Y, Shanno D F, Vanderbei R J. Interior-point methods for nonconvex nonlinear programming: Filter methods and merit functions. Comput Optim Appl, 2002, 23: 257–272
- 12 Forsgren A, Gill Ph E. Primal-dual interior methods for nonconvex nonlinear programming. SIAM J Optim, 1998, 8: 1132–1152
- 13 Forsgren A, Gill Ph E, Wright M H. Interior methods for nonlinear optimization. SIAM Review, 2002, 44: 525–597
- 14 Gould N I M, Orban D, Toint Ph L. An interior-point l_1 -penalty method for nonlinear optimization. Technical report. Rutherford Appleton Laboratory, Chilton, Oxfordshire, England, 2003, RAL-TR-2003-022
- 15 Shanno D F, Vanderbei R J. Interior-point methods for nonconvex nonlinear programming: Orderings and higher-order methods. Math Program, 2000, 87: 303–316
- 16 Ulbrich M, Ulbrich S, Vicente L N. A globally convergent primal-dual interior-point filter method for nonlinear programming. Math Program, 2004, 100: 379–410
- 17 Vanderbei R J, Shanno D F. An interior-point algorithm for nonconvex nonlinear programming. Comput Optim Appl, 1999, 13: 231–252
- 18 Yamashita H. A globally convergent primal-dual interior point method for constrained optimization. Optim Methods Softw, 1998, 10: 448–469
- 19 Byrd R H, Gilbert J C, Nocedal J. A trust region method based on interior point techniques for nonlinear programming. Math Program, 2000, 89: 149–185
- 20 Dennis J E, Heinkenschloss M, Vicente L N. Trust-region interior-point SQP algorithms for a class of nonlinear programming problems. SIAM J Control Optim, 1998, 36: 1750–1794
- 21 Gertz E M, Gill Ph E. A primal-dual trust region algorithm for nonlinear optimization. Math Program, 2004, 100: 49–94
- 22 Tseng P. Convergent infeasible interior-point trust-region methods for constrained minimization. SIAM J Optim, 2002, 13: 432–469
- 23 Yamashita H, Yabe H, Tanabe T. A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization. Math Program, 2004, 102: 111–151
- 24 Chen L F, Goldfarb D. Interior-point ℓ_2 -penalty methods for nonlinear programming with strong global convergence properties. Math Program, 2006, 108: 1–36
- 25 Liu X W, Yuan Y X. A null-space primal-dual interior-point algorithm for nonlinear optimization with nice convergence properties. Math Program, 2010, 125: 163–193
- 26 Tits A L, Wächter A, Bakhtiari S, et al. A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties. SIAM J Optim, 2003, 14: 173–199
- 27 El-Bakry A S, Tapia R A, Tsuchiya T, et al. On the formulation and theory of the Newton interior-point method for nonlinear programming. J Optim Theo Appl, 1996, 89: 507–541
- 28 Burke J V, Han S P. A robust sequential quadratic programming method. Math Program, 1989, 43: 277–303

- 29 Byrd R H. Robust trust-region method for constrained optimization. Houston, TX: Paper presented at the SIAM Conference on Optimization, 1987
- 30 Byrd R H, Marazzi M, Nocedal J. On the convergence of Newton iterations to non-stationary points. *Math Program*, 2004, 99: 127–148
- 31 Liu X W, Sun J. A robust primal-dual interior point algorithm for nonlinear programs. *SIAM J Optim*, 2004, 14: 1163–1186
- 32 Liu X W, Yuan Y X. A robust algorithm for optimization with general equality and inequality constraints. *SIAM J Sci Comput*, 2000, 22: 517–534
- 33 Wächter A, Biegler L T. Failure of global convergence for a class of interior point methods for nonlinear programming. *Math Program*, 2000, 88: 565–574
- 34 Yuan Y X. On the convergence of a new trust region algorithm. *Numer Math*, 1995, 70: 515–539
- 35 Burer S, Monteiro R D C, Zhang Y. Solving a class of semidefinite programs via nonlinear programming. *Math Program*, 2002, 93: 97–122
- 36 Correa R, Ramirez H C. A global algorithm for nonlinear semidefinite programming. *SIAM J Optim*, 2005, 15: 303–318
- 37 Fares B, Noll D, Apkarian P. Robust control via sequential semidefinite programming. *SIAM J Control Optim*, 2002, 40: 1791–1820
- 38 Freund R W, Jarre F, Vogelbusch C H. Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling. *Math Program*, 2007, 109: 581–611
- 39 Jarre F. An interior method for nonconvex semidefinite programs. *Optim Eng*, 2000, 1: 347–372
- 40 Qi H. Local duality of nonlinear semidefinite programming. *Math Oper Res*, 2009, 34: 124–141
- 41 Sun D. The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Math Oper Res*, 2006, 31: 761–776
- 42 Alizadeh F, Haeblerly J P A, Overton M L. Complementarity and nondegeneracy in semidefinite programming. *Math Program*, 1997, 77: 111–128
- 43 Leyffer S. <http://wiki.mcs.anl.gov/leyffer/index.php/MacMPEC>, 2012
- 44 Liu X W, Perakis G, Sun J. A robust SQP method for mathematical programs with linear complementarity constraints. *Comput Optim Appl*, 2006, 34: 5–33
- 45 Liu X W, Sun J. Generalized stationary points and an interior point method for mathematical programs with equilibrium constraints. *Math Program*, 2004, 101: 231–261
- 46 Luo Z Q, Pang J S, Ralph D. *Mathematical Programs with Equilibrium Constraints*. Cambridge: Cambridge University Press, 1996
- 47 Chin C M, Fletcher R. On the global convergence of an SLP-filter algorithm that takes EQP steps. Report NA/199, Department of Mathematics. Dundee, Scotland: Dundee University, 2001
- 48 Fletcher R, Gould N, Leyffer S, et al. Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming. *SIAM J Optim*, 2002, 13: 635–659
- 49 Fletcher R, Leyffer S. Nonlinear programming without a penalty function. *Math Program*, 2002, 91: 239–269
- 50 Fletcher R, Leyffer S, Toint Ph L. On the global convergence of a filter-SQP algorithm. *SIAM J Optim*, 2002, 13: 44–59
- 51 Ribeiro A A, Karas E W, Gonzaga C C. Global convergence of filter methods for nonlinear programming. *SIAM J Optim*, 2008, 19: 1231–1249
- 52 Wächter A, Biegler L T. Line search filter methods for nonlinear programming: Motivation and global convergence. *SIAM J Optim*, 2005, 16: 1–31
- 53 Wächter A, Biegler L T. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math Program*, 2006, 106: 25–57
- 54 Byrd R H, Liu G, Nocedal J. On the local behaviour of an interior point method for nonlinear programming. In: Griffiths D F, Higham D J, Watson G A, eds. *Numerical Analysis*. Reading, MA: Addison-Wesley Longman 1997, 37–56
- 55 Izmailov A F, Solodov M V. Newton-type methods for optimization problems without constraint qualifications. *SIAM J Optim*, 2004, 15: 210–228
- 56 Wright S J. Constraint identification and algorithm stabilization for degenerate nonlinear programs. *Math Program*, 2003, 95: 137–160
- 57 Armand P, Benoist J. A local convergence property of primal-dual methods for nonlinear programming. *Math Program*, 2008, 115: 199–222
- 58 Bongartz I, Conn A R, Gould N I M. CUTE: Constrained and unconstrained testing environment. *ACM Trans Math*

Softw, 1995, 21: 123–160

59 Byrd R H, Hribar M E, Nocedal J. An interior-point algorithm for large-scale nonlinear programming. SIAM J Optim, 1999, 9: 877–900

How does the linear independence assumption affect algorithms of nonlinear constrained optimization

LIU XinWei

Abstract Firstly, some recent progress in nonlinear constrained optimization is surveyed in this paper. The terminology on the global convergence of algorithms for constrained optimization is officially defined for the first time. By noticing the gap between the exactness of optimality conditions and the approximation of algorithms, and looking into the initial Newton-type algorithmic framework for equality constrained optimization, one will understand why the assumption on the linear independence of gradients of the constraints should be and can be weakened. The discussions are also extended to the optimization with inequality constraints. Without the linear independence assumption, the local convergence of an algorithm using the exact penalty function and the second-order correction is then proved. These recognitions may help to develop more efficient new algorithms for constrained optimization including nonlinear semidefinite and conic programming in the future.

Keywords sequential quadratic programming, interior-point method, linear independence, local convergence, second-order correction

MSC(2010) 90C30, 90C31, 65K05, 65K10

doi: 10.1360/012011-1037