摘　要

本文研究非线性约束优化问题的求解. 我们提出几种序列二次规划 (SQP) 算法, 建立相应算法的收敛性, 并对所给算法进行数值实验.

第 2 章结合积极集估计技术, 提出一个求解非线性约束优化问题的积极集 SQP 算法, 且该算法所产生的点列均为可行点列. 该算法的主要优点在于: 算法的主搜索方向由一个低维的凸二次规划确定, 为克服 Maratos 效应, 我们通过求解一个低维的最小二乘问题得到高阶修正方向. 在适当的条件下, 我们证明算法具有全局收敛性和超线性收敛性.

第 3 章提出一个求解极大极小问题修正的 SQP 算法. 该算法的主搜索方向通过求解一个二次规划得到. 为克服 Maratos 效应, 不同于以往的方法, 我们通过求解一个线性方程组得到高阶修正方向, 无需求解二次规划子问题. 在较弱的条件下, 我们证明该算法是全局收敛和一步超线性收敛的.

第 4 章提出一个求解非线性约束优化问题的可行点 SQP 算法. 在该算法的每一个迭代步, 分别通过求解一个低维的二次规划子问题和一个低维的线性方程组得到一个下降方向和一个可行方向, 在此基础上, 我们构造一个可行下降方向. 为避免 Maratos 效应, 我们通过解一个低维的线性方程组得到高阶修正方向. 在适当的条件下, 该算法被证明是全局收敛和超线性收敛的. 与已有算法相比, 本章提出的算法的优点是: 算法中线性方程组不涉及乘子估计, 所求解的两个线性方程组的系数矩阵相同且比已有算法中系数矩阵的结构简单, 因而可减少计算量.

第 5 章提出一个求解非线性不等式约束优化问题的非内点型可行点 QP-free 算法, 这个算法不要求迭代点必须是可行域的内点. 在算法的每一个迭代步, 为得到搜索方向, 只需求解四个系数相同的线性方程组. 在适当的条件下, 我们建立该算法的全局收敛性和超线性收敛定理.

第 6 章提出一个内点型可行点 QP-free 算法, 在该算法的每一个迭代步, 为得到搜索方向, 只需求解三个系数矩阵相同的线性方程组, 而且在无严格互补条件下得到系数矩阵的一致非奇异性和近似乘子序列的有界性, 且其全局收敛性分析不受稳定点数目有限的限制.

关键词: 非线性约束优化; 极大极小问题; SQP **算法;** QP-free **算法; 全局收敛性; 超线性收敛性**

Abstract

The purpose of the thesis is to study the numerical method for solving nonlinear constrained optimization. We propose several sequential quadratic programming (SQP) algorithms, and establish their global convergence and superlinear convergence. We do numerical experiments to test the proposed algorithms.

In Chapter 2, by the use of an active set estimate technique, we propose an active set SQP method for nonlinear constrained optimization. The method generates a sequence of feasible points. Major advantage of the proposed method lies in that the main search direction is determined by a lower dimension quadratic programs. To overcome Maratos effect, we calculate a higher-order correction direction by solving a reduced least squares problem. Under appropriate conditions, we show that the proposed method is globally and superlinearly convergent.

In Chapter 3, we present a modified SQP method for the minimax problem. In the algorithm, the main search direction is obtained by solving a quadratic program which always has a solution. In order to avoid Maratos effect, different from the previous technique where a quadratic programs is solved, we solve a system of linear equations to obtain a higher-order correction direction. Under some mild conditions, we obtain the global and superlinear convergence.

In Chapter 4, we have proposed a feasible point SQP algorithm for nonlinear inequality constrained optimization problems. At each iteration, we determined a descent and feasible direction by solving a reduced quadratic programming subproblem and a reduced system of linear equations, respectively. We then device a feasible descent direction through a suitable combination of the descent direction and the feasible direction. To overcome Maratos effect, a higher-order correction direction is obtained by solving another reduced system of linear equations. The algorithm is proved to be globally and superlinearly convergent under some mild conditions. A good feature of the proposed algorithm is that the coefficient matrix for the system of linear equations do not involve multiplier estimate. Furthermore, the structure of coefficient matrix is simpler than the one in the previous algorithms.

In Chapter 5, we propose a noninterior type feasible point QP-free algorithm for nonlinear inequality constrained optimization problems. The generated iterates are feasible but not necessary interior points of the feasible region. At each iteration, a search direction is obtained by solving four systems of linear equations with the same coefficient matrix. The algorithm is proved to be globally and superlinearly convergent under some mild conditions.

In Chapter 6, we propose an interior point type feasible QP-free algorithm for nonlinear inequality constrained optimization problems. At each iteration, by solving three systems of linear equations with the same coefficient matrix, a search direction is generated. The algorithm is proved to be globally and superlinearly convergent under some mild conditions. Advantages of the algorithm include: the uniformly nonsingularity of the coefficient matrices and the boundedness of the approximate Lagrange multipliers without the strictly complementarity are obtained. Moreover, the global convergence is achieved even if the number of the stationary points is infinite.

Key Words: Nonliear constrained optimization; Minimax problems; SQP algorithm; QP-free algorithm; Global convergence; Superlinear convergence

湖 南 大 学

学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：　　　　　　　　日期：2008年 5月 23日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权湖南大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于

1、保密□，在____年解密后适用本授权书。

2、不保密□。

（请在以上相应方框内打"√"）

作者签名：　　　　　　　　日期：2008年 5月22 日

导师签名：　　　　　　　　日期：2008年 5 月23 日

第 1 章　绪论

1.1　课题的发展概况与研究意义

本文中我们主要考虑求解如下非线性不等式约束优化问题:

$$
\begin{array}{lll}
 & \min & f(x) \\
(P) & \text{s.t.} & g_j(x) \le 0,\ \ j \in I \triangleq \{1, 2, ..., m\},
\end{array}
\tag{1.1}
$$

其中 m 是正整数, 函数 f, $g_j(j \in I): R^n \to R$ 连续可微.

序列二次规划 (SQP) 方法最早由 Wilson[5] 于 1963 年在其博士论文中首次提出. 该算法的主要步骤如下, 设当前迭代点为 x^k, 通过求解如下二次规划子问题得 d^k 及相应 Lagrange 乘子 λ^k.

$$
\begin{array}{lll}
 & \min & \nabla f(x^k)^T d + \frac{1}{2} d^T H_k d \\
(QP) & \text{s.t.} & g_j(x^k) + \nabla g_j(x^k)^T d \le 0, \quad j \in I,
\end{array}
\tag{1.2}
$$

其中 H_k 为问题 (1.1) 的 Lagrange 函数关于 x 的二阶导数 $\nabla^2_{xx} L(x^k, \lambda^k)$, 令下一个迭代点为 $x^{k+1} = x^k + d^k$. 设问题 (1.2) 的最优解为 x^* 且 λ^* 为相应的 Lagrange 乘子. Wilson[5] 证明了, 当 (x^k, λ^k) 充分接近 (x^*, λ^*) 时, 该算法收敛且具有二阶收敛速度. 但是由于该算法仅具有局部收敛性质, 并且由于 $\nabla^2_{xx} L(x^k, \lambda^k)$ 通常是不正定的, 因而 (1.2) 的解可能不存在, 加之需要计算目标函数和约束函数的二阶导数, 使用起来不很方便, 所以在其后一段时间内没引起足够重视. 直到上个世纪 70 年代中期, 一些学者才对其进行了进一步的研究. 1976 年, Han[6] 提出了用类似无约束优化的 DFP 公式修正 H_k, 并证明了该算法的局部超线性收敛性. 1977 年, Han[7] 进一步提出用 l_1 精确罚函数作为效益函数确定步长, 从而建立了该算法的全局收敛性. 其后 Powell[8] 于 1977 年对 Han 的方法作了进一步修改, 具体给出 l_1 精确罚函数中参数的选取及矩阵 H_k 的修正方法, 建立了一个快速有效的算法, 通常称作 Wilson-Han-Powell 方法. 此后, 序列二次规划方法引起了许多学者的极大兴趣, 得到了进一步的研究和发展, 并已取得大量的研究成果 [9]. 目前该方法已成为求解中小规模约束优化问题最受欢迎的方法之一. 随着 SQP 算法自身理论的不断完善, 许多学者将该算法应用到工程与经济等各种实际优化模型上, 例如均衡问题 ([10]-[12]), 最优控制问题 ([13],[14]), 极大极小问题 ([90]-[99]), 半无限规划问题 [65] 等, 并取得了显著的效果. 在研究 SQP 类算法的同时, 有学者平行地提出了另一类算法: QP-free 算法 (或称序列线性方程组算法)([103]-[109]).QP-free 算法通过求解一个或多个线性方程组确定 d^k, 无需求解类似于 (1.2) 的二次规划子问题. 由于解线性方程组比求解含不等式约束的二次规

划要简单得多，因而 QP-free 类算法比 SQP 算法计算量少. 因此对 QP-free 算法的研究具有重要的理论意义和应用价值并已引起国内外学者的广泛关注.

我们根据 SQP 算法的初始点与迭代点的性质，将 SQP 算法分为不可行点 SQP 算法和可行点 SQP 算法. 下面简单介绍 SQP 算法和 QP-free 算法的产生与发展.

1.1.1 不可行点序列二次规划算法

SQP 算法的出现与发展基于以下事实：设 (d,λ) 为 (1.2) 的 KKT 点对，则

$$\begin{aligned}&\nabla f(x)+Hd+\sum_{j\in I}\lambda_j\nabla g_j(x)=0,\\&\lambda_j\geq 0,\ g_j(x)+\nabla g_j(x)^Td\leq 0,\\&\lambda_j(g_j(x)+\nabla g_j(x)^Td)=0.\end{aligned}\tag{1.3}$$

如果 $d=0$, 则 (1.3) 等价于 (1.1) 的 KKT 条件.

为表述方便，下面我们给出 SQP 算法的一般步骤：

步 0. 给出初始点 x^0, 初始对称正定矩阵 H_0, 令 $k:=0$.

步 1. 在 x^k 处，求解二次规划 (1.2) 得解 d^k.

步 2. 令 $x^{k+1}=x^k+\alpha_kd^k$, 其中步长 α_k 由某种线搜索得到.

步 3. 修正 H_k 得 H_{k+1}, 使 H_{k+1} 保持对称正定.

步 4. 令 $k:=k+1$, 返回步 1.

当 SQP 算法的初始点与迭代点不要求是问题 (1.1) 的可行点时，我们把 SQP 算法称作不可行点 SQP 算法. 显然 Wilson-Han-Powell 方法是不可行点 SQP 算法. 由于迭代点不一定是可行点，因此二次规划 (1.2) 的可行域可能是空集，从而导致二次规划 (1.2) 不相容. 为克服这个缺陷，许多学者提出了大量的修正 SQP 算法. 其中 Powell 提出了一个修正方法：通过引进辅助变量 ξ, 在求解 (1.2) 之前先解如下线性规划

$$\begin{aligned}&\max\ \xi\\&\text{s.t.}\ \ \xi g_j(x^k)+\nabla g_j(x^k)^Td\leq 0,\quad j\in V,\\&\qquad g_j(x^k)+\nabla g_j(x^k)^Td\leq 0,\quad j\in S,\end{aligned}\tag{1.4}$$

其中

$$V=\{j\in I:g_j(x)>0\},S=\{j\in I:g_j(x)\leq 0\}.$$

Fletcher[15] 将计算搜索方向的二次规划子问题 (1.2) 转化为一个无约束优化问题，通过求解这个约束优化问题的极小点得到搜索方向. Tone[16] 也给出了两种修正的二次子规划，并证明它们的解不为 0 时是效益函数 -l_1 精确罚函数的下降方向. Schittkowski[17] 则将 (1.2) 化为一个具有线性约束的最小二乘问题求解.

作为对 Tone 的方法的改进， Spellucci[18] 给出了一种克服二次规划子问题不相容的新方法. Han 和 Burke[20] 也给出了一个修正的二次子规划，该子问题总是相容的. 受 Han 和 Burke 方法的启发， Zhou[21] 给出了一种克服二次规划子问题不相容的方法：该方法先解一个具有有界约束的线性规划，后解一个修正二次规划得到修正方向. 高，贺和赖 [61] 通过引进一个新的罚函数，结合积极集估计技术，给出了一个具有可解子问题的 SQP 算法. Facchinei[26] 利用 Lucidi[25] 给出的精确罚函数作为效益函数，给出了一个克服二次规划子问题不相容的方法：如果子问题 (1.2) 相容且其解可接受，则将其解作为搜索方向；否则，利用效益函数梯度的一个合理近似得到搜索方向. 最近， Mo, Zhang 和 Wei[24] 给出了一个新的二次子规划，该子问题也总是相容的.

正如 Maratos[27] 指出，传统的不可行点 SQP 算法可能存在 Maratos 效应. 即当迭代点充分靠近最优解时，不一定能保证步长恒为 1 或趋于 1, 从而影响该算法的超线性收敛性. 为克服这个缺陷，学者们给出了几种克服 Maratos 效应的方法. 第一种是用光滑精确罚函数代替 l_1 精确罚函数作为效益函数，具体见文献 [28]-[32]. 第二种方法是二阶修正步方法，具体见文献 [33]-[38]. 第三种方法是 Watchdog 技术和非单调线搜索技巧，具体见文献 [39]-[47].

在传统的不可行点 SQP 算法的收敛性分析中，在解处线性无关约束规格成立对分析不可行点 SQP 算法的二次收敛性是必要的. 为减弱这个条件， Wright[48] 通过对 (1.2) 修改，在 Mangasarian-Fromovitz 约束规格下，给出了一个二次收敛的 SQP 算法，该算法称作稳定 SQP 算法. 后来， Hager[50] 证明了 Wright 的算法在无严格互补条件下也是二次收敛的，只不过要求在解处强二阶充分条件成立. Li[51] 通过求解线性方程组也给出了一个稳定 SQP 算法，并在较弱的条件下证明了该算法的超线性收敛性.

由于传统的不可行点 SQP 算法大多涉及罚参数的调整，为克服这个缺陷，Fletcher, Leyffer, Toint, *Wächter* 和 Biegler 等人给出了一类新的序列二次规划算法： SQP-filter 算法，而且在一定条件下，给出了该算法的全局收敛性和超线性收敛性，具体见文献 [52]-[56].

此外，有学者结合信赖域法，内点法和大规模优化的特点，给出了几类新的 SQP 算法：大规模 SQP 算法可参见文献 [69]-[72], 信赖域 SQP 算法可参见文献 [73]-[76], 内点型 SQP 算法可参见文献 [77]-[79].

1.1.2 可行点序列二次规划算法

由于大多数求解约束优化问题的 SQP 算法从任意的初始点出发，使用精确罚函数作为效益函数，这使得迭代点可能不是问题 (1.1) 的可行点. 而对许多实际问题而言，算法产生可行迭代点是非常重要的，例如对某些工程实际问题以及某

些实际问题，其目标函数在可行域外往往没有定义，此时要求迭代点必须可行. 为此， Panier 和 Tits [88] 结合可行方向法和 SQP 算法，提出了一类可行点序列二次规划 (FSQP) 算法，在该算法中，由于任意的迭代点 x^k 均属于 (1.1) 的可行域，故二次规划 (1.2) 总是相容的，即问题 (1.1) 的可行域总是非空的，而且目标函数可以直接作为线搜索中的效益函数. 但是该算法每个迭代步需求解三个不同的二次规划子问题，有时还要求出一个一阶可行下降方向，而且该算法的收敛速度仅是二步超线性收敛. 为克服上述 SQP 算法的不足，许多学者对其进行了改进，Panier 和 Tits [101] 对上述算法进行了改进，给出了一个组合可行与下降的超线性收敛的 SQP 算法. 该算法的主搜索方向由其对应二次规划子问题的解和另一二次规划的解的凸组合得到. 为避免 Maratos 效应，需求解第三个二次规划子问题得到一个二阶校正方向. 高，吴 [62] 也对上述算法进行了修正，每步迭代只需计算一个二次子规划和一个逆矩阵，而且在较弱的条件下证明了算法的全局收敛和一步超线性收敛性. Jian[57][58] 结合广义投影梯度也对上述算法进行了修正，给出一个超线性收敛和二次收敛的可行点 SQP 算法. 为减少上述算法每步迭代的计算量，最近， Zhu [102] 给出了一个简化可行点 SQP 算法，其主搜索方向由其对应二次规划子问题的解和一线性方程组的解的凸组合得到. 为避免 Maratos 效应，一个二阶校正方向通过解另一线性方程组得到.

另外，文献 [83, 84, 85] 中提出了另一类 FSQP 算法，在算法的当前迭代点 x^k 处，通过求解如下二次规划子问题得到主搜索方向：

$$\begin{array}{ll} \min\limits_{(z,d)} & z+\frac{1}{2}d^TH_kd \\ \text{s.t.} & \nabla f(x^k)^Td\le z, \\ & g_i(x^k)+\nabla g_i(x^k)^Td\le \sigma_k z,\quad i\in I, \end{array} \tag{1.5}$$

其中 H_k 是一对称正定矩阵， σ_k 是一正参数. 从 (1.5) 中可以看出，如果 $\sigma_k>0$ 且上述二次规划的解 $d_{\sigma_k}\neq 0$, 则 d_{σ_k} 是问题 (1.1) 在 x^k 处的可行下降方向. 文献 [83] 中， Birge,Qi,Wei 讨论了问题 (1.1) 的 FJ 点的求解. 他们给参数 σ 一个正的初始值，当单位步长不被接受时，适当增加参数 σ 的值，正如作者们指出，这种参数调整方法破坏了算法的超线性收敛性；文献 [84] 中， Lawrence 和 Tits 也给出了一个类似算法：需通过求解一个等式约束二次规划来修正参数 σ, 使得 $\sigma=O(\|d_\sigma\|^2)$, 为克服 Maratos 效应，需求解另一个等式二次规划来修正搜索方向 d_σ. Kostreva 和 Chen [85] 也通过求解二次规划 (1.5) 提出了一个 FSQP 算法，其中要求 $\sigma=o(\|d_\sigma\|)$, 但未具体给出参数 σ 的修正方式.

1.1.3 QP-free 算法

由上述讨论可知，SQP 类算法每次迭代需求解至少一个含不等式约束的二次规划，而相对而言，解线性方程组比求解含不等式约束的二次规划要简单得多.

因此，在研究 SQP 类算法的同时，有学者平行地提出了另一类算法：QP-free 算法 (或称序列线性方程组算法)([103]-[109]). 该类算法的产生基于以下事实：给定一个合适的指标集 $\bar{I} \subseteq I$, 考虑如下关于 (d, λ) 的线性方程组

$$\begin{aligned} &\nabla f(x) + Hd + N_{\bar{I}}\lambda = 0, \\ &g_{\bar{I}}(x) + N_{\bar{I}}^T d = 0, \end{aligned} \tag{1.6}$$

其中，

$$g_{\bar{I}}(x) = (g_j(x), j \in \bar{I}),\ N_{\bar{I}} = (\nabla g_j(x), j \in \bar{I}).$$

在可行点 x 处，如果 $d = 0, \lambda \geq 0$, 则

$$\begin{aligned} &\nabla f(x) + N_{\bar{I}}\lambda = 0, \\ &g_j(x) = 0,\ \lambda_j \geq 0,\ j \in \bar{I}. \end{aligned} \tag{1.7}$$

若令 $\lambda_j = 0,\ j \in I \setminus \bar{I}$, 则 (1.7) 表明 x 为问题 (1.1) 的 KKT 点.

此类算法的思想最早见于文献 [63][64], 而由 Panier,Tits 和 Herskovits[103] 于 1988 年正式提出，该算法每步迭代需求解两个不同的线性方程组和一个线性最小二乘问题，而且为保证系数矩阵序列的一致非奇异性和近似乘子序列的有界性，必须假定在所有可行点处严格互补条件成立. Gao,He 和 Wu[104] 给出了一个序列线性方程组算法，该算法在没有假定聚点孤立的条件下是全局收敛的. 但是在收敛性分析中必须要求乘子序列有界. Qi,Qi[105] 基于互补函数和 KKT 条件，提出了一个求解问题 (P) 的可行点 QP-free 算法，他们在无严格互补条件下证明了迭代矩阵的一致非奇异性和近似乘子序列的有界性. Yang,Li 和 Qi[106] 通过引进一个工作集的概念，提出了一个新的求解问题 (P) 的可行点 QP-free 算法，这个新算法仅考虑工作集内的约束，这使得计算量大大减少. 在这个算法的每一个迭代步，为得到搜索方向，仅要求解四个系数相同的线性方程组. 在合适的条件下，这个算法具有全局收敛性和局部一步超线性收敛速度，甚至具有二次收敛速度. 数值实验表明这个算法特别适合求解大规模不等式约束优化问题. 最近， Zhu[109] 给出了一个新的内点型可行点 QP-free 算法，在此算法的每一个迭代步，为得到搜索方向，只需解三个相同系数矩阵的线性方程组，但是系数矩阵序列需在严格互补条件下保持一致非奇异性. 而且为得到全局收敛性，稳定点数目有限这个条件是必要的. 以上均为可行点 QP-free 算法，有关不可行点 QP-free 算法的研究参见文献 [19][59][60][107]. 最近，也有学者研究了求解目标函数是 SC^1 函数 (即一阶导数半光滑) 的约束优化问题的可行点 QP-free 算法，可参见文献 [66][67].

1.2 本文主要工作及创新点

- 第 2 章在文献 [83, 84, 85] 的基础上，结合积极集估计技术，提出一个积极集可行点 SQP 算法，该算法的主要优点在于：算法的主搜索方向由一个低维的凸二次规划 (2.3) 确定，而且不需求解任何子问题来修正 (2.3) 中参数 σ_k，只需取 $\sigma_k = \|d^{k-1}\|^\nu$，其中 d^{k-1} 为前一个迭代点处的主搜索方向. 为克服 Maratos 效应，通过求解一个最小二乘问题得到高阶修正方向. 在适当的条件下，我们证明算法具有全局收敛性和超线性收敛性.

- 第 3 章提出一个求解极大极小问题的修正 SQP 算法. 该算法的主搜索方向通过求解一个二次规划得到. 为克服 Maratos 效应，不同于文献 [95][98]，我们通过求解一个线性方程组得到高阶修正方向，而 [95] 和 [98] 中需通过求解一个二次规划子问题得到高阶修正方向. 在较弱的条件下，我们证明该算法是全局收敛和一步超线性收敛的.

- 第 4 章提出一个求解非线性约束优化问题的可行点 SQP 算法. 在该算法的每一个迭代步，分别通过求解二次规划子问题和线性方程组得到一个下降方向和一个可行方向，在此基础上，我们构造一个可行下降方向. 为避免 Maratos 效应，我们通过解一线性方程组得到高阶修正方向. 在适当的条件下，该算法被证明是全局收敛和超线性收敛的. 与已有算法相比，本章提出算法的优点是：算法中线性方程组不涉及乘子估计，所求解的两个线性方程组的系数矩阵相同且比文献 [102] 中系数矩阵的结构简单，因而可减少计算量.

- 第 5 章以文献 [106] 中算法为基础，提出一个求解非线性不等式约束优化问题的非内点型可行点 QP-free 算法，这个算法不要求迭代点必须是可行域的内点. 在算法的每一个迭代步，为得到搜索方向，只需求解四个系数相同的线性方程组. 在适当的条件下，我们建立该算法的全局收敛性和超线性收敛定理.

- 第 6 章以文献 [105] 中算法为基础，提出一个改进的内点型可行点 QP-free 算法，在该算法的每一个迭代步，为得到搜索方向，只需解三个系数矩阵相同的线性方程组，而且在无严格互补条件下得到系数矩阵序列的一致非奇异性和近似乘子序列的有界性，且其全局收敛性分析不受稳定点数目有限的限制.

1.3 本文所用的记号

x:	实向量
$f(x)$:	目标函数
n:	问题的维数，即 x 的分量数目
$\Re$:	全体实数组成的集合
$\Re^n$:	全体 n 维实向量组成的集合
I_n:	n 阶单位矩阵
A^{-1}:	非奇异方阵 A 的逆
A^T:	矩阵 A 的转置
$det(A)$:	矩阵 A 的行列式
$\|F\|$:	集合 F 包含的元素的个数
$\nabla f(x)$:	$f(x)$ 的梯度
$\nabla^2_{xx} f(x)$:	$f(x)$ 的海色矩阵
$\\|\cdot\\|$:	向量的欧氏范数

第 2 章 求解约束优化问题的一个积极集可行点 SQP 算法

2.1 引言

本章考虑求解如下不等式约束优化问题：

$$
(P)\quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_j(x) \le 0, \quad j \in I, \end{array} \tag{2.1}
$$

其中 m 是正整数，函数 f, $g_j(j \in I): R^n \to R$ 连续可微.

最近，文献 [83, 84, 85] 中提出了一类求解问题 (2.1) 的 FSQP 算法，在算法的每一个迭代点 x^k 处，通过求解如下二次规划子问题得到主搜索方向：

$$
\begin{array}{ll} \min\limits_{(z,d)} & z + \frac{1}{2} d^T H_k d \\ \text{s.t.} & \nabla f(x^k)^T d \le z, \\ & g_i(x^k) + \nabla g_i(x^k)^T d \le \sigma_k z, \quad i \in I, \end{array} \tag{2.2}
$$

其中 H_k 是一对称正定矩阵，σ_k 是一正参数. 从 (2.2) 中可以看出，如果 $\sigma_k > 0$ 且上述二次规划的解 $d_{\sigma_k} \ne 0$, 则 d_{σ_k} 是问题 (2.1) 在 x^k 处的可行下降方向. 文献 [83] 中，Birge,Qi,Wei 讨论了问题 (2.1) 的 FJ 点的求解. 他们给参数 σ_k 一个正的初始值，当单位步长不被接受时，适当增加参数 σ_k 的值，但这种参数修正方法破坏了算法的超线性收敛性；文献 [84] 中，Lawrence 和 Tits 也给出了一个类似算法：通过求解一个等式约束二次规划来修正参数 σ_k, 使得 $\sigma_k = O(\|d\sigma_k\|^2)$, 另外，为克服 Maratos 效应，需求解另一个等式二次规划来修正搜索方向 d_{σ_k}. Kostreva 和 Chen [85] 也通过求解二次规划 (2.2) 提出了一个 FSQP 算法，其中要求 $\sigma_k = o(\|d\sigma_k\|)$, 但未具体给出参数 σ_k 的修正方式.

本章中，我们在文献 [83, 84, 85] 的基础上，结合近似积极集技术，提出了一个积极集可行点 SQP 算法. 该算法的主搜索方向由求解一个低维的二次规划 (2.3) 得到，而且不需求解任何子问题来修正 (2.3) 中参数 σ_k, 只需取 $\sigma_k = \|d^{k-1}\|^\nu$. 为克服 Maratos 效应，通过求解一个线性方程组得到一个高阶修正方向. 在适当的条件下，我们证明该算法具有全局收敛性和超线性收敛性.

2.2 算法描述

首先我们定义问题 (P) 的可行域 X 为

$$
X = \{x \in R^n : g_i(x) \le 0, i \in I\},
$$

而且对每一个可行点 $x \in X$, 定义积极集为

$$I(x) = \{i \in I : g_i(x) = 0\}.$$

本章中我们总假定可行集 X 非空且如下假设成立:
H1 对每一个可行点 $x \in X$, 向量组 $\{\nabla g_i(x), i \in I(x)\}$ 是线性无关的.

对 $x \in X$, 我们使用如下近似积极集，可参见文献 [86][106].

$$A(x;\varepsilon) = \{i : g_i(x) + \varepsilon\rho(x, \lambda(x)) \geq 0\},$$

其中 ε 是一非负参数， $\rho(x, \lambda(x)) = \sqrt{\| \Phi(x,\lambda) \|}$,

$$\Phi(x, \lambda(x)) = \begin{bmatrix} \nabla_x L(x, \lambda(x)) \\ \min\{-f(x), \lambda(x)\} \end{bmatrix}, f(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \cdots \\ g_m(x) \end{bmatrix},$$

$$\lambda(x) = -(\nabla g(x)^T \nabla g(x) + diag(g_i(x))^2)^{-1} \nabla g(x)^T \nabla f(x),\ \ \nabla g(x) = (\nabla g_i(x), i \in I).$$

易知 (x^*, λ^*) 是问题 (P) 的 KKT 点当且仅当 $\Phi(x^*, \lambda^*) = 0$ 或 $\rho(x^*, \lambda^*) = 0$. Facchinei 等人在文献 [86] 中证明了如果二阶充分条件和 Mangassarian-Fromovotz 约束规格成立，那么对任意 $\varepsilon > 0$, 当 x 充分接近 x^*, $A(x;\varepsilon) = I(x^*)$.

下面我们给出求解问题 (P) 的一个积极集可行点 SQP 算法步骤.

有关参数 r, $r_j > 0 (j \in I)$, $\tau \in (2,3)$, $\nu \in (0, +\infty)$, $\beta \in (0,1)$, $\gamma \in (0,1)$, $\alpha \in (0, \frac{1}{2})$.

有关数据 给定初始点 $x^1 \in X$, 一个对称正定矩阵 H_1, $\sigma_1 > 0$ 和 $\varepsilon^0 \geq 0$. 令 $k = 1$.

步 1 令 $\varepsilon = \varepsilon^{k-1}$.

步 2 令 $A^k(\varepsilon) = A(x^k, \varepsilon)$. 如果 $\nabla g_{A^k(\varepsilon)}(x^k)$ 非满秩，则令 $\varepsilon := \sigma\varepsilon$, 进入步 2.(这里 $\nabla g_{A^k(\varepsilon)}(x^k) = (\nabla g_i(x^k),\ i \in A^k(\varepsilon)))$

步 3 令 $\varepsilon^k = \varepsilon$, $A^k = A^k(\varepsilon)$.

步 4 (计算主搜索方向)

对当前迭代点 x^k, 求解

$$(NQP) \quad \begin{array}{ll} \min & rz + \frac{1}{2} d^T H_k d \\ \text{s.t.} & \nabla f(x^k)^T d \leq rz, \\ & g_i(x^k) + \nabla g_i(x^k)^T d \leq r_i \sigma_k z, \quad i \in A^k, \end{array} \tag{2.3}$$

得 (z_k, d^k), 令 $(u_0^k, u_{A^k}^k)$ 是其相应的 Lagrange 乘子. 如果 $d^k = 0$, 那么 x^k 是问题 (P) 的 KKT 点，停；否则进入步 5.

步 5 (计算高阶修正方向)

通过求解下面的最小二乘问题得高阶修正方向 $\widetilde{d}^k$:

$$\begin{array}{lll}(LS) & \min & \frac{1}{2}\|d\|_{H_k}^2 \\ & \text{s.t.} & g_i(x^k+d^k)+\nabla g_i(x^k)^T d = -(1-\rho_k)\|d^k\|^\tau+\rho_k r_i \sigma_k^\gamma z_k \|d^k\|, \quad i\in A^k.\end{array} \tag{2.4}$$

其中

$$\rho_k=\begin{cases}0, & \text{if} \quad \|d^k\|^2 \geq -(\max\limits_{i\in A^k} r_i)\sigma_k z_k; \\ 1, & \text{if} \quad \|d^k\|^2 < -(\max\limits_{i\in A^k} r_i)\sigma_k z_k.\end{cases}$$

如果 $\|\widetilde{d}^k\|>\|d^k\|$, 令 $\widetilde{d}^k=0$.

步 6 (线搜索)

求序列 $\{1,\beta,\beta^2,...\}$ 中满足如下不等式组的第一个数作为 λ_k

$$f(x^k+\lambda d^k+\lambda^2\widetilde{d}^k)\leq f(x^k)+\alpha\lambda\nabla f(x^k)^T d^k, \tag{2.5}$$

$$g_i(x^k+\lambda d^k+\lambda^2\widetilde{d}^k)\leq 0, \quad \forall i\in I. \tag{2.6}$$

步 7 令 $x^{k+1}=x^k+\lambda_k d^k+\lambda_k^2\widetilde{d}^k$, $\sigma_{k+1}=\|d^k\|^\nu$.

步 8 计算一个新的对称正定矩阵 H_{k+1}, 令 $k:=k+1$, 返回步 1.

为讨论方便, 对任意的 k, 我们令 $u_i^k=0$, $i\in I\setminus A^k$.

下面我们证明该算法是适定的.

引理 2.2.1 *对 $x^k\in X$, 如果条件 H1 成立, 则该算法在步 2 中的循环经有限步终止.*

该引理的证明类似于文献 [87] 中引理 1.1 和引理 2.8 的证明.

引理 2.2.2 *如果 H_k 是一对称正定矩阵, 且参数 r, $r_j(j\in A^k)$ 都是正数以及 $\sigma\geq 0$, 则 (NQP) 总存在唯一最优解.*

该引理的证明见文献 [84] 中引理 1 的证明.

引理 2.2.3 *若 H_k 是一对称正定矩阵, 则 (LS) 总存在唯一最优解.*

利用 H_k 对称正定性和 $g_{A^k}(x^k)$ 的满秩性易得该引理的证明.

引理 2.2.4 *设引理 2.2.2 中的条件成立. 如果 (z_k,d^k) 是 (2.2) 的最优解, 则*

(i) $rz_k+\frac{1}{2}(d^k)^T H_k d^k\leq 0$ 和 $z_k\leq 0$;

(ii) $z_k=0 \Longleftrightarrow d^k=0 \Longleftrightarrow x^k$ 是问题 (P) 的 KKT 点;

(iii) $z_k<0 \Longrightarrow d^k$ 是问题 (P) 在 x^k 处的可行下降方向.

该引理的证明见文献 [85] 中引理 3.4 的证明.

引理 2.2.5 *该算法步 6 中的线搜索是适定的.*

证明: 首先, 由 (2.5) 可知:

$$\begin{aligned}
a_k &\triangleq f(x^k+\lambda d^k+\lambda^2\widetilde{d}^k)-f(x^k)-\alpha\lambda\nabla f(x^k)^Td^k\\
&= \nabla f(x^k)^T(\lambda d^k+\lambda^2\widetilde{d}^k)-\alpha\lambda\nabla f(x^k)^Td^k+o(\lambda)\\
&= (1-\alpha)\lambda\nabla f(x^k)^Td^k+o(\lambda).
\end{aligned}$$

又 f 连续可微, $\alpha\in(0,\frac{1}{2})$, 故由引理 2.2.4 可知: 存在 $\overline{\lambda}>0$, 使得任意 $\lambda\in[0,\overline{\lambda}]$, $a_k\le 0$.

其次, 由 (2.6) 和引理 2.2.4 可知: 当 $i\in A^k$ 时,

$$\begin{aligned}
b_i^k &\triangleq g_i(x^k+\lambda d^k+\lambda^2\widetilde{d}^k)\\
&= g_i(x^k)+\lambda\nabla g_i(x^k)^Td^k+o(\lambda)\\
&\le (1-\lambda)g_i(x^k)+\lambda z+o(\lambda).
\end{aligned}$$

故存在 $\overline{\lambda}_i>0$, 使得任意 $\lambda\in[0,\overline{\lambda}_i]$, $b_i^k\le 0$.
当 $i\notin A^k$ 时, $g_i(x^k)<-\varepsilon\rho(x^k,\lambda(x^k))<0$. 由 $F(\lambda)=g_i(x^k+\lambda d^k+\lambda^2\widetilde{d}^k)$ 的连续性知: 存在 $\overline{\lambda}_i'>0$, 使得任意 $\lambda\in[0,\overline{\lambda}_i']$, $b_i^k\le 0$.

令 $\widehat{\lambda}=\min\{\overline{\lambda},\overline{\lambda}_i,\overline{\lambda}_i',\ i\in I\}$, 因此结论成立.

2.3 全局收敛性分析

本节我们将分析 2.2 节中算法的全局收敛性. 首先做如下假设.

H2. 该算法所产生的序列 $\{x^k\}$ 是有界的.

H3. 对任意 k 和所有 $d\in R^n$, 存在两个正数 a, $b>0$ 满足

$$a\|d\|^2\le d^TH_kd\le b\|d\|^2.$$

我们不妨设 x^* 是序列 $\{x^k\}$ 的一个聚点. 注意到 A^k 和 J_k 是有限集合 I 的子集, 故存在一子集 K 使得

$$\lim_{k\in K}x^k=x^*,\quad A^k\equiv A,\quad J_k\equiv J,\ \forall k\in K, \tag{2.7}$$

其中

$$J_k=\{i\in A^k:g_i(x^k)+\nabla g_i(x^k)^Td^k=r_i\sigma_kz_k\}.$$

引理 2.3.1 *如果假设 H2,H3 成立, 则序列* $\{d^k:k\in K\},\{z_k:k\in K\}$ *和* $\{\widetilde{d}^k:k\in K\}$ *都是有界的.*

证明: 首先，由 $\nabla f(x^k) \to \nabla f(x^*), k \in K$ 知：存在常数 $c_0 > 0$ 使得 $\| \nabla f(x^k) \| \leq c_0$，$\forall k \in K$. 而且，结合引理 2.2.4, (2.3) 和假设 H3 可得

$$\begin{aligned} 0 \geq r z_k + \tfrac{1}{2}(d^k)^T H_k d^k &\geq \nabla f(x^k)^T d^k + \tfrac{a}{2}\|d^k\|^2 \\ &\geq -\|\nabla f(x^k)\|\|d^k\| + \tfrac{a}{2}\|d^k\|^2 \\ &\geq -c_0\|d^k\| + \tfrac{a}{2}\|d^k\|^2, \qquad k \in K. \end{aligned}$$

这表明 $\{d^k : k \in K\}$ 是有界的.

其次，$\{z_k : k \in K\}$ 的有界性可从 $\{d^k : k \in K\}$ 的有界性和下面的不等式推得.

$$0 \geq z_k \geq \tfrac{1}{r}\nabla f(x^k)^T d^k \geq -\tfrac{1}{r}\|\nabla f(x^k)\| \cdot \|d^k\| \geq -\tfrac{c_0}{r}\|d^k\|, \quad k \in K. \tag{2.8}$$

最后，$\{\widetilde{d}^k : k \in K\}$ 的有界性可从 $\{d^k : k \in K\}$ 的有界性得到.

下面我们给出 (NQP) 的 KKT 条件如下：

$$H_k d^k + u_0^k \nabla f(x^k) + \sum_{j\in A} u_j^k \nabla g_j(x^k) = 0, \quad u_A^k = (u_j^k, j \in A), \tag{2.9}$$

$$r = r u_0^k + \sum_{j\in A} u_j^k r_j \sigma_k, \quad u_j^k \geq 0, j \in A, u_0^k \geq 0, \tag{2.10}$$

$$0 \leq u_0^k \perp (r z_k - \nabla f(x^k)^T d^k) \geq 0, \tag{2.11}$$

$$0 \leq u_j^k \perp (r_j \sigma_k z_k - g_j(x^k) - \nabla g_j(x^k)^T d^k) \geq 0, \quad j \in A, \tag{2.12}$$

其中记号 $x \perp y$ 表示 $x^T y = 0$.

引理 2.3.2 *(i) 乘子序列 $\{u_0^k\}_{k=0}^{\infty}$ 是有界的.*

(ii) 设条件 H1,H2,H3 成立，且令乘子向量 $u^k = (u_A^k, 0_{I\setminus A^k(\epsilon)}) = (u_J^k, 0_{I\setminus J})$. 如果 $\lim\limits_{k\in K} x^k = x^*$ 以及 $\lim\limits_{k\in K} d^k = 0$, 则 $\{u^k : k \in K\}$ 是有界的.

证明: (i) 根据 (2.3) 的 KKT 条件可得

$$r = r u_0^k + \sum_{j\in A} u_j^k r_j \sigma_k \geq r u_0^k,$$

即，$0 \leq u_0^k \leq 1$.

(ii) 假设 (ii) 的结论不成立，则存在一个子集 $K' \subseteq K$ 使得 $\| u^k \| = \| u_J^k \| \to \infty$，$k \in K'$. 因此，用 $\|u_J^k\|$ 去除 (2.9) 两边可得

$$\frac{1}{\|u_J^k\|} H_k d^k + \frac{u_0^k}{\|u_J^k\|}\nabla f(x^k) + \sum_{j\in J} \frac{u_j^k}{\|u_J^k\|}\nabla g_j(x^k) = 0. \tag{2.13}$$

注意到 $\{\frac{u_J^k}{\|u_J^k\|} : k \in K'\}$ 是一模长为 1 的序列，故我们不妨设

$$\frac{u_j^k}{\|u_J^k\|} \to \overline{u}_j, \quad k \in K', \quad j \in J, \quad 0 \leq (\overline{u}_j,\ j \in J) \neq 0. \tag{2.14}$$

因此，对 (2.13) 两边取极限 ($k \in K'$, $k \to \infty$), 考虑到假设 H3 和给定的条件可得：

$$\sum_{j \in J} \overline{u}_j \nabla g_j(x^*) = 0. \tag{2.15}$$

另一方面，根据给定的条件可得： $J \subseteq I(x^*)$, 因此根据 (2.14), (2.15) 和 H1 可得一矛盾，故 $\{u^k : k \in K\}$ 是有界的.

考虑到序列 $\{u_0^k : k \in K\}$, $\{u^k : k \in K\}$ 和 $\{\sigma_k\}$ 的有界性，我们不妨设

$$u^k = (u_j^k, j \in I) \to u^* = (u_j^*, j \in I), \quad u_0^k \to u_0^*, \quad \sigma_k \to \sigma_*, \quad k \in K. \tag{2.16}$$

引理 2.3.3 *设 $\{x^k\}$ 是该算法所产生的序列，如果 $\lim\limits_{k \in K} x^k = x^*$ 且 $\lim\limits_{k \in K} d^k = 0$, 则 x^* 是问题 (P) 的 KKT 点.*

证明：由 $\lim\limits_{k \in K} x^k = x^*$ 和 $\lim\limits_{k \in K} d^k = 0$ 可知： $\lim\limits_{k \in K} z^k = 0$. 因此，对 (2.9)—(2.12) 两边取极限 ($k \in K$, $k \to \infty$) 可得：

$$\begin{aligned}
&u_0^* \nabla f(x^*) + \sum_{j \in J} u_j^* \nabla g_j(x^*) = 0, \\
&u_j^* g_j(x^*) = 0, \quad u_j^* \geq 0, \quad g_j(x^*) \leq 0, \quad j \in A, \\
&r = r u_0^* + \sigma_* \sum_{j \in J} r_j u_j^*, \quad u_0^* \geq 0.
\end{aligned} \tag{2.17}$$

由 (2.17) 的第三式易知： $(u_0^*, u_J^*) \neq 0$, 而且结合假设 H1 可得： $u_0^* > 0$. 这就表明 $(x^*, \dfrac{u^*}{u_0^*})$ 是问题 (P) 的 KKT 点.

基于引理 2.3.1, 引理 2.3.2 和引理 2.3.3, 我们给出如下全局收敛性定理.

定理 2.3.1 *如果假设 H1,H2,H3 成立，那么该算法或有限步终止问题 (P) 的 KKT 点或产生一无限序列 $\{x^k\}$, 其每一个聚点 x^* 都是问题 (P) 的 KKT 点. 而且 $\{\frac{u^k}{u_0^k} : k \in K\}$ 收敛到对应于 x^* 的 KKT 乘子且 $\lim\limits_{k \in K} u_0^k > 0$.*

证明：该定理的第一个结论是显然的. 因此我们不妨设 $\{x^k\}$ 为该算法所产生的无限序列且 (2.16) 成立. 下面分 $\sigma_* = 0$ 和 $\sigma_* > 0$ 两种情况给与证明.

(1). 当 $\sigma_* = 0$ 时. 由步 7 知，存在一无限子集 $K_1 \subseteq K$ 使得 $\lim\limits_{k \in K_1} d^{k-1} = 0$. 再次由步 7 知：

$$\| x^k - x^{k-1} \| \leq t_k \| d^{k-1} \| + t_k^2 \| \widetilde{d}^{k-1} \| \leq 2t^k \| d^{k-1} \| \longrightarrow 0, \quad k \in K_1.$$

因此，根据 $\lim\limits_{k \in K_1} x^k = x^*$ 可得 $\lim\limits_{k \in K_1} x^{k-1} = x^*$. 进一步由引理 2.3.3 知 x^* 是问题 (P) 的 KKT 点.

(2). 当 $\sigma_* > 0$ 时. 显然，只需证明 $\lim\limits_{k \in K} d^k = 0$ 即可. 为此，我们假设 $\lim\limits_{k \in K} d^k \neq 0$, 则存在一无限子集 $K' \subseteq K$ 和一常数 $\Delta > 0$ 使得对任意 $k \in K'$, 有 $\| d^k \| \geq \Delta$. 下面的证明分两步进行.

a. 首先证明存在 $\overline{\lambda}>0$ 使得对任意 $k\in K'$, 有 $\lambda_k\geq\overline{\lambda}$.

$$\begin{aligned}
f(x^k+\lambda d^k+\lambda^2\widetilde{d}^k)-f(x^k) &- \alpha\lambda\nabla f(x^k)^Td^k\\
&= \nabla f(x^k)^T(\lambda d^k+\lambda^2\widetilde{d}^k)-\alpha\lambda\nabla f(x^k)^Td^k+o(\lambda)\\
&\leq \lambda(1-\alpha)\nabla f(x_k)^Td^k+o(\lambda)\\
&\leq \lambda(1-\alpha)rz_k+o(\lambda)\\
&\leq -\tfrac{1}{2}\lambda(1-\alpha)(d^k)^TH_kd^k+o(\lambda)\\
&\leq -\tfrac{1}{2}a\lambda(1-\alpha)\|d^k\|^2+o(\lambda)\\
&\leq -\tfrac{1}{2}a\lambda(1-\alpha)\Delta^2+o(\lambda).
\end{aligned}$$

上面的最后一个不等式表明对 $k\in K'$ 充分大和 $\lambda>0$ 充分小，有 (2.5) 成立.

下面分析 (2.6): 如果 $j\notin A$, 即， $g_j(x^k)<-\varepsilon\rho(x^k,\lambda(x^k))<0$. 结合 $g_j(x)$ 的连续性， $\{d^k:k\in K\}$ 和 $\{\widetilde{d}^k:k\in K\}$ 的有界性，则对 $k\in K'$ 充分大和 $\lambda>0$ 充分小，有 $g_j(x^k+\lambda d^k+\lambda^2\widetilde{d}^k)\leq 0$ 成立.

如果 $j\in A$, 则由泰勒展式和 (2.3) 可得

$$\begin{aligned}
g_j(x^k+\lambda d^k+\lambda^2\widetilde{d}^k) &= g_j(x^k)+\lambda\nabla g_j(x^k)^Td^k+o(\lambda)\\
&\leq g_j(x^k)+\lambda(r_j\sigma_kz_k-g_j(x^k))+o(\lambda)\\
&= (1-\lambda)g_j(x^k)+\lambda r_j\sigma_kz_k+o(\lambda)\\
&\leq \lambda r_j\sigma_kz_k+o(\lambda).
\end{aligned}$$

从而由 (2.3) 和假设 H3 可得

$$\begin{aligned}
g_j(x^k+\lambda d^k+\lambda^2\widetilde{d}^k) &\leq \lambda r_j\sigma^*(-\tfrac{1}{2r}(d^k)^TH_kd^k)+o(\lambda)\\
&\leq -\lambda r_j\sigma^*\tfrac{1}{2r}a\parallel d^k\parallel^2+o(\lambda)\\
&\leq -\lambda r_j\sigma^*\tfrac{1}{2r}a\Delta^2+o(\lambda).
\end{aligned}$$

因此上面的不等式表明对 $k\in K'$ 充分大和 $\lambda>0$ 充分小，有 (2.6) 成立.

综合上面的分析可知：存在 $\overline{\lambda}>0$, 使得对 $k\in K'$ 有 $\lambda_k\geq\overline{\lambda}$.

b. 其次利用 $\lambda_k\geq\overline{\lambda}>0$ 得出一个矛盾.

由 (2.5), (2.3) 和假设 H3 可知

$$\begin{aligned}
f(x^{k+1}) &\leq f(x^k)+\alpha\lambda_k\nabla f(x^k)^Td^k\leq f(x^k)+\alpha\lambda_krz_k\\
&\leq f(x^k)-\tfrac{1}{2}\alpha\lambda_k(d^k)^TH_kd^k\leq f(x^k)-\tfrac{1}{2}\alpha\lambda_ka\parallel d^k\parallel^2,\qquad\forall k.
\end{aligned}$$

因此序列 $\{f(x^k)\}$ 是单调下降的，考虑到 $\lim\limits_{k\in K'}f(x^k)=f(x^*)$, 故 $\lim\limits_{k\to\infty}f(x^k)=f(x^*)$.
又

$$f(x^{k+1})\leq f(x^k)-\frac{1}{2}a\alpha\overline{\lambda}\Delta^2,\quad\forall k\in K'.$$

对上面不等式两边取极限 ($k\in K'$ 且 $k\to\infty$) 可得 $-\frac{1}{2}a\alpha\overline{\lambda}\Delta^2\geq 0$, 这是一个矛盾 !因此， $d^*=0$. 故由引理 2.3.3 知： x^* 是问题 (P) 的 KKT 点，而且 $\{\frac{u^k}{u_0^k}:k\in K\}$ 收敛到对应于 x^* 的 KKT 乘子且 $\lim\limits_{k\in K}u_0^k>0$.

2.4 收敛速度分析

本节我们将分析 2.2 节中算法的收敛速度. 为此, 我们做如下假设.

H4 (i) 函数 $f(x)$, $g_j(x)(j \in I)$ 二次连续可微.

(ii) 在序列 $\{x^k\}$ 的聚点 x^* 处二阶充分条件成立, 即,

$$d^T \nabla^2_{xx} L(x^*, u^*) d > 0, \quad \forall d \in \Omega \stackrel{def}{=} \{d \in R^n : d \neq 0, \nabla g_{I(x^*)}(x^*)^T d = 0\}, \tag{2.18}$$

其中

$$L(x, u) = f(x) + \sum_{j \in I} u_j g_j(x). \tag{2.19}$$

(iii) 在 x^* 处严格互补条件成立, 即 $\lambda^* - g(x^*) > 0$.

引理 2.4.1 *(i) 如果假设 H1,H2 成立, 则* $\lim\limits_{k\to\infty} d^k = 0$, $\lim\limits_{k\to\infty} \widetilde{d}^k = 0$, $\lim\limits_{k\to\infty} z_k = 0$, $\lim\limits_{k\to\infty} \sigma_k = 0$ *和* $\lim\limits_{k\to\infty} \|x^{k+1} - x^k\| = 0$.
(ii) 如果假设 H1,H2,H3 成立, 则 $\lim\limits_{k\to\infty} x^k = x^*$.

证明: (i) 首先用反证法证明 $d^k \to 0$, $z_k \to 0$. 假设结论不成立, 则由 $\{d^k\}, \{z_k\}$ 的有界性和引理 2.3.1 知: 存在一子集 K 使得

$$d^k \to d^*, \; z_k \to z_* < 0 (k \in K).$$

根据 $\{d^{k-1}, u_0^k, u_i^k\}$ 的有界性得: 存在子集 $K_1 \subseteq K$ 使得

$$u_0^k \to u_0^*, \; u_i^k \to u_i^*, i \in I, \; \sigma_k \to \sigma_* (k \in K_1).$$

而由 $d^k \to d^* \neq 0$ 和定理 2.3.1 的证明知: $\sigma_* = 0$. 对 (NQP) 的 KKT 条件 (2.9)-(2.12) 两边取极限 $(k \in K_1,\; k \to \infty)$ 得:

$$H_* d^* + \nabla f(x^*) + \sum_{i \in A} u_i^* \nabla g_i(x^*) = 0;$$

$$u_0^* = 1, \; \nabla f(x^*)^T d^* = z_* < 0;$$

$$\nabla g_i(x^*)^T d^* \le 0, i \in I(x^*).$$

结合上两式和 Farkas 引理知: 存在 $\lambda^* \in R^{|I(x^*)|}$ 使得

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*) = 0, \; \lambda_i^* \ge 0, \; i \in I(x^*).$$

这表明 x^* 是问题 (P) 的 KKT 点, 由引理 2.2.4 知: $d^* = 0$. 矛盾 故 $d^k \to 0$, $z_k \to 0$.

由上述结论和步 5 和步 7 分别可知：$\lim\limits_{k\to\infty}\widetilde{d}^k=0$ 和 $\lim\limits_{k\to\infty}\sigma^k=0$.

而又从上面的结论可得：

$$\lim_{k\to\infty}\|x^{k+1}-x^k\|=\lim_{k\to\infty}\|\lambda_k d^k+\lambda_k^2\widetilde{d}^k\|\le\lim_{k\to\infty}(\|d^k\|+\|\widetilde{d}^k\|)=0.$$

(ii) 类似于文献 [88] 中的性质 4.1 的证明可知：x^* 是问题 (P) 的一个孤立 KKT 点，因此由定理 2.3.1 知：x^* 是 $\{x^k\}$ 的一个孤立聚点，这和 $\lim\limits_{k\to\infty}\|x^{k+1}-x^k\|=0$ 一起表明 $\lim\limits_{k\to\infty}x^k=x^*$.

类似于文献 [89] 中引理 3.1 的证明可得如下引理.

引理 2.4.2 *在上述假设下，当 k 充分大时，矩阵*

$$M_k\overset{def}{=}\begin{bmatrix}H_k & \nabla g_{A^k}(x^k)\\ \nabla g_{A^k}(x^k)^T & 0\end{bmatrix}.$$

是非奇异的，而且存在常数 $C>0$ 使得 $\|M_k^{-1}\|\le C$.

引理 2.4.3 *如果假设 H1,H2,H3,H4(ii) 成立，则*

$$J_k=I(x^*)=A^k. \tag{2.20}$$

$$|z_k|=O(\|d^k\|),\quad \|\widetilde{d}^k\|=O(\max\{\|d^k\|^2,\ -(\max_{j\in A^k}r_j)\sigma_k z_k\})=o(\|d^k\|), \tag{2.21}$$

证明：首先从 $\lim\limits_{k\to\infty}(x^k,d^k,z_k,\sigma_k)=(0,0,0,0)$ 可得：$J_k\subseteq I(x^*)$. 反过来，对任意 $i\in I(x^*)$, 由 H4(ii) 知：$\lambda_i^*>0$, 结合定理 2.3.1 知：$u_i^k>0$, 即，$i\in J_k$. 类似于文献 [86] 中的定理 2.3 和定理 3.7 可得 $I(x^*)=A^k$.

由 (2.3) 的第一个不等式可得

$$-\|\nabla f(x^k)\|\|d^k\|\le rz_k,\quad |z_k|\le\frac{1}{r}\|\nabla f(x^k)\|\|d^k\|.$$

因此不难证明 $|z_k|=O(\|d^k\|)$.

其次，我们证明 (2.21) 中的第二式. 注意到 (LS) 等价于如下线性方程组：

$$\begin{bmatrix}H_k & \nabla g_{A^k}(x^k)\\ \nabla g_{A^k}(x^k)^T & 0\end{bmatrix}\begin{bmatrix}\widetilde{d}^k\\ \widetilde{\lambda}^k\end{bmatrix}=\begin{bmatrix}0\\ -(1-\rho)\|d^k\|^\tau e_{A^k}+\rho\sigma_k^\gamma z_k\|d^k\|r_{A^k}+\widetilde{g}_{A^k}\end{bmatrix},$$

其中 $\widetilde{g}_{A^k}=-g_{A^k}(x^k+d^k)$, $e_{A^k}=(1,...,1)^T\in R^{|A^k|}$, $r_{A^k}=(r_j,\ j\in A^k)^T$.

由泰勒展式和 (2.20) 可得：$\widetilde{g}_{A^k}=O(\|d^k\|^2)+\sigma_k z_k r_{A^k}$. 因此结合引理 2.4.1, 引理 2.4.2 和 $\tau\in(2,3)$ 有：$\|\widetilde{d}^k\|=O(\max\{\|d^k\|^2,\ -(\max\limits_{j\in A^k}r_j)\sigma_k z_k\})=o(\|d^k\|)$.

引理 2.4.4 *如果假设 H1,H2,H3 成立，则 $\{\frac{u^k}{u_0^k}\}$ 收敛到问题 (P) 相关于 x^* 的 KKT 乘子且 $\lim\limits_{k\to\infty} u_0^k = 1$.*

证明: 利用 $\lim\limits_{k\to\infty}(x^k, d^k, \widetilde{d}^k, z_k) = (0,0,0,0)$, 类似于引理 2.3.2(ii) 的证明可知：$\{u^k\}$ 是有界的. 类似于定理 2.3.1 的证明可得：对任意无限子集 K, 序列 $\{\frac{u^k}{u_0^k} : k \in K\}$ 存在一个聚点 $\widetilde{u}^*$ 且 $(x^*, \widetilde{u}^*)$ 是问题 (P) 的 KKT 点. 又对应于 x^* 的 Lagrange 乘子是唯一的，因此，序列 $\{\frac{u^k}{u_0^k}\}$ 收敛到问题 (P) 对应于 x^* 的 Lagrange 乘子. 最后，由 (2.10), $\{u^k\}$ 的有界性以及引理 2.4.1(i) 可知

$$\lim_{k\to\infty} u_0^k = \lim_{k\to\infty}(1 - \frac{1}{r}\sum_{j\in J_k} u_j^k r_j \sigma_k) = 1.$$

为确保 k 充分大时，$\lambda_k \equiv 1$. 我们再如下假设.

H5 假设 $\|P_k(\nabla_{xx}^2 L(x^k, \frac{u_{J_k}^k}{u_0^k}) - H_k)d^k\| = o(\|d^k\|)$, 其中

$$R_k = (\nabla g_j(x^k), j \in J_k), \quad P_k = I_n - R_k(R_k^T R_k)^{-1} R_k^T.$$

$$L(x, \frac{u_{J_k}^k}{u_0^k}) = f(x) + \sum_{j\in J_k} \frac{u_j^k}{u_0^k} g_j(x).$$

引理 2.4.5 *如果假设 H1,H2,H3,H4,H5 成立，则当 k 充分大时，步长 $\lambda_k \equiv 1$.*

证明: 首先我们证明当 k 充分大且 $\lambda = 1$ 时 (2.6) 成立. 对 $j \notin I(x^*)$, 即, $g_j(x^*) < 0$. 由 $(x^k, d^k, \widetilde{d}^k) \to (x^*, 0, 0)(k \to \infty)$ 和 $g_i(x)$ 的连续性可知：$g_j(x^k + d^k + \widetilde{d}^k) \le 0$.

对 $j \in I(x^*) = A^k$, 由泰勒展式，(2.3), (2.4) 和 (2.21) 可得

$$\begin{aligned} g_j(x^k + d^k + \widetilde{d}^k) &= g_j(x^k + d^k) + \nabla g_j(x^k + d^k)^T \widetilde{d}^k + O(\|\widetilde{d}^k\|^2) \\ &= g_j(x^k + d^k) + \nabla g_j(x^k)^T \widetilde{d}^k + O(\|d^k\|\|\widetilde{d}^k\|) + O(\|\widetilde{d}^k\|^2) \\ &= -(1-\rho_k)\|d^k\|^\tau + \rho_k r_j \sigma_k^\gamma z_k \|d^k\| + O(\max\{\|d^k\|^3, -(\max_{j\in A^k} r_j)\sigma_k z_k \|d^k\|\}) \end{aligned} \tag{2.22}$$

因此从 ρ_k 的定义和 (2.22) 可知：

$$g_j(x^k + d^k + \widetilde{d}^k) \le 0.$$

这表明 (2.6) 当 k 充分大且 $\lambda = 1$ 时成立.

下面证明 (2.5) 当 k 充分大且 $\lambda = 1$ 时成立.

由泰勒展式和 (2.21) 可得

$$\begin{aligned} \omega_k &\stackrel{def}{=} f(x^k + d^k + \widetilde{d}^k) - f(x^k) - \alpha \nabla f(x^k)^T d^k \\ &= \nabla f(x^k)^T(d^k + \widetilde{d}^k) + \tfrac{1}{2}(d^k)^T \nabla_{xx}^2 f(x^k) d^k - \alpha \nabla f(x^k)^T d^k + o(\|d^k\|^2). \end{aligned} \tag{2.23}$$

另外，由 (2.3) 的 KKT 条件和 (2.21) 知

$$u_0^k\nabla f(x^k)^T(d^k+\widetilde{d}^k)=-(d^k)^TH_kd^k-\sum_{j\in J_k}u_j^k\nabla g_j(x^k)^T(d^k+\widetilde{d}^k)+o(\|d^k\|^2). \tag{2.24}$$

$$\begin{aligned} u_0^k\nabla f(x^k)^Td^k &= -(d^k)^TH_kd^k-\sum_{j\in J_k}u_j^k\nabla g_j(x^k)^Td^k \\ &= -(d^k)^TH_kd^k+\sum_{j\in J_k}u_j^kg_j(x^k)-\sum_{j\in J_k}u_j^kr_j\sigma_kz_k. \end{aligned}$$

再次由 (2.22) 的第三式和泰勒展式得

$$g_j(x^k)+\nabla g_j(x^k)^T(d^k+\widetilde{d}^k)+\frac{1}{2}(d^k)^T\nabla_{xx}^2g_j(x^k)d^k=o(\|d^k\|^2),\quad j\in J_k.$$

因此，

$$-\sum_{j\in J_k}u_j^k\nabla g_j(x^k)^T(d^k+\widetilde{d}^k)=\sum_{j\in J_k}u_j^kg_j(x^k)+\tfrac{1}{2}(d^k)^T(\sum_{j\in J_k}u_j^k\nabla_{xx}^2g_j(x^k))d^k+o(\|d^k\|^2). \tag{2.25}$$

把 (2.25) 代入 (2.24) 得

$$u_0^k\nabla f(x^k)^T(d^k+\widetilde{d}^k)=-(d^k)^TH_kd^k+\sum_{j\in J_k}u_j^kf_j(x^k)+\tfrac{1}{2}(d^k)^T(\sum_{j\in J_k}u_j^k\nabla_{xx}^2g_j(x^k))d^k+o(\|d^k\|^2). \tag{2.26}$$

把 (2.26) 和 (2.24) 的第三式代入 (2.23) 得

$$\begin{aligned} \omega_k &= \tfrac{1}{u_0^k}(\alpha-1)(d^k)^TH_kd^k+\tfrac{1}{2}(d^k)^T\nabla_{xx}^2L(x^k,\tfrac{u^k}{u_0^k})d^k \\ &+ (1-\alpha)\sum_{j\in J_k}\tfrac{u_j^k}{u_0^k}g_j(x^k)+\alpha\sum_{j\in J_k}\tfrac{u_j^k}{u_0^k}r_j\sigma_kz_k+o(\|d^k\|^2) \\ &\le (\tfrac{1}{u_0^k}(\alpha-1)+\tfrac{1}{2})a\|d^k\|^2+\tfrac{1}{2}(d^k)^T(\nabla_{xx}^2L(x^k,\tfrac{u^k}{u_0^k})-H_k)d^k \\ &+ (1-\alpha)\sum_{j\in J_k}\tfrac{u_j^k}{u_0^k}g_j(x^k)+\alpha\sum_{j\in J_k}\tfrac{u_j^k}{u_0^k}r_j\sigma_kz_k+o(\|d^k\|^2). \end{aligned}$$

根据 P_k 定义可得

$$\begin{aligned} d^k &= P_kd^k+R_k(R_k^TR_k)^{-1}R_k^Td^k \\ &= P_kd^k+R_k(R_k^TR_k)^{-1}(-g_{I(x^*)}(x^k)+\sigma_kz_ke) \\ &= P_kd^k+O(\|g_{I(x^*)}(x^k)\|)+O(\|\sigma_kz_ke\|). \end{aligned}$$

从而有

$$\begin{aligned} \omega_k &\le (\tfrac{1}{u_0^k}(\alpha-1)+\tfrac{1}{2})a\|d^k\|^2+\tfrac{1}{2}((d^k)^TP_k+R_k(R_k^TR_k)^{-1}R_k^Td^k)(\nabla_{xx}^2L(x^k,\tfrac{u^k}{u_0^k})-H_k)d^k \\ &+ (1-\alpha)\sum_{j\in J_k}\tfrac{u_j^k}{u_0^k}g_j(x^k)+\alpha\sum_{j\in J_k}\tfrac{u_j^k}{u_0^k}r_j\sigma_kz_k+o(\|d^k\|^2) \\ &= (\tfrac{1}{u_0^k}(\alpha-1)+\tfrac{1}{2})a\|d^k\|^2+o(\|g_{I(x^*)}(x^k)\|)+o(\|\sigma_kz_ke\|) \\ &+ (1-\alpha)\sum_{j\in J_k}\tfrac{u_j^k}{u_0^k}g_j(x^k)+\alpha\sum_{j\in J_k}\tfrac{u_j^k}{u_0^k}r_j\sigma_kz_k+o(\|d^k\|^2). \end{aligned}$$

因此，由 $\alpha\in(0,\frac{1}{2})$，$u_0^k\to 1$，可知 (2.5) 当 $\lambda=1$ 时成立.

由假设 H4(ii) 易得如下引理：

引理 2.4.6 在上述假设下，当 k 充分大时，矩阵

$$D_k \overset{def}{=} \begin{bmatrix} P_k \nabla^2_{xx} L(x^*, u^*) \\ R_k^T \end{bmatrix}$$

是列满秩的.

定理 2.4.1 在上述假设下，该算法是超线性收敛的. 即，序列 $\{x^k\}$ 满足 $\|x^{k+1} - x^*\| = o(\|x^k - x^*\|)$.

证明: 因为 $P_k R_k = 0$, 故由泰勒展式， (2.9) 和引理 2.4.4 知:

$$\begin{aligned} \frac{1}{u_0^k} P_k H_k d^k &= -P_k \nabla f(x^k) = -P_k(\nabla f(x^k) + R_k u^*_{I(x^*)}) \\ &= -P_k \nabla^2_{xx} L(x^*, u^*_{I(x^*)})(x^k - x^*) + O(\|x^k - x^*\|^2), \end{aligned}$$

即

$$P_k(\frac{1}{u_0^k} H_k - \nabla^2_{xx} L(x^*, u^*_{I(x^*)}))d^k = -P_k \nabla^2_{xx} L(x^*, u^*_{I(x^*)})(x^k + d^k - x^*) + O(\|x^k - x^*\|^2).$$

又由泰勒展式可得:

$$g_{I(x^*)}(x^k) = g_{I(x^*)}(x^k) - g_{I(x^*)}(x^*) = R_k^T(x^k - x^*) + O(\|x^k - x^*\|^2).$$

故由 (2.7) 和引理 2.4.3 知

$$R_k^T(x^k + d^k - x^*) = o(\|d^k\|) + O(\|x^k - x^*\|^2).$$

从而由引理 2.4.6 知

$$x^k + d^k - x^* = (D_k^T D_k)^{-1} D_k^T \begin{bmatrix} -P_k(\frac{1}{u_0^k} H_k - \nabla^2_{xx} L(x^*, u^*_{I(x^*)}))d^k \\ 0 \end{bmatrix} + O(\|x^k - x^*\|^2).$$

由上式，引理 2.4.4 和假设 H5 知:

$$\|x^k + d^k - x^*\| = o(\|d^k\|) + O(\|x^k - x^*\|^2).$$

根据引理 2.4.3 和引理 2.4.5 得 $\|\widetilde{d}^k\| = o(\|d^k\|)$. 从而有

$$\begin{aligned} \|x^{k+1} - x^*\| &= \|x^k + d^k + \widetilde{d}^k - x^*\| \le o(\|d^k\|) + O(\|x^k - x^*\|^2) \\ &= o(\|d^k + \widetilde{d}^k\|) + O(\|x^k - x^*\|^2) \\ &= \frac{o(\|d^k + \widetilde{d}^k\|)}{\|d^k + \widetilde{d}^k\|}(\|(x^{k+1} - x^*) - (x^k - x^*)\|) + O(\|x^k - x^*\|^2) \\ &\le \frac{|o(\|d^k + \widetilde{d}^k\|)|}{\|d^k + \widetilde{d}^k\|}(\|x^{k+1} - x^*\| + \|x^k - x^*\|) + O(\|x^k - x^*\|^2) \\ &= o(\|x^{k+1} - x^*\|) + o(\|x^k - x^*\|). \end{aligned}$$

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

2.5 数值试验

本节，我们对 2.2 节中的算法和文献 [84] 中的算法进行数值比较．2.2 节中算法用 Algo21 表示，文献 [84] 中的算法用 Algo22 表示．两种算法都用 MATLAB 6.5 编程，都在 Windows XP 的个人计算机实现．其中 (2.3) 和 (2.4) 都使用优化工具箱求解.

整个实验中，我们使用如下 BFGS 公式：

$$H_{k+1} = H_k - \frac{H_k s^k (s^k)^T H_k}{(s^k)^T H_k s^k} + \frac{\widetilde{y}^k (\widetilde{y}^k)^T}{(s^k)^T \widetilde{y}^k}, \quad (k \geq 0),$$

其中

$$s^k = x^{k+1} - x^k,\ \widetilde{y}^k = y^k + a_k(\gamma_k s^k + R_k R_k^T s^k),\ \gamma_k = \min\{\|d^k\|^2, \xi\},\ \ \xi \in (0,1),$$

$$y^k = \nabla_x L(x^{k+1}, \lambda^k) - \nabla_x L(x^k, \lambda^k),$$

$$a_k = \begin{cases} 0, & \text{if} \quad (s^k)^T y^k \geq \delta \|s^k\|^2, \delta \in (0,1); \\ 1, & \text{if} \quad 0 \leq (s^k)^T y^k < \delta \|s^k\|^2; \end{cases}$$

整个实验中，令

$$H_1 = I_n,\ \ \tau = 2.7,\ \ r = 0.25,\ \ r_j = 0.002 (j \in I),\ \ \beta = 0.5,$$

$$\gamma = 0.5,\ \ \alpha = 0.3,\ \ \sigma_1 = 0.001,\ \ \nu = 2,\ \ \sigma = 0.5,\ \ \varepsilon^0 = 2.$$

表 2.1 的测试问题来自文献 [110] 和文献 [111]．所选测试问题的初始点和文献 [110], 文献 [111] 中相同．表 2.1 的列表示如下意义：**prob** 表示来自文献 [110] 和文献 [111] 的测试问题；**Ni** 表示求解的迭代次数；**Nf, Ng** 分别表示约束函数值和约束函数梯度值的迭代次数．**objective, dnorm** 和 **eps** 表示目标函数值，模 d^k 和精度 ϵ.

表 2.1 数值结果

Algo	Prob	Ni	Nf	Ng	objective	dnorm	eps
Algo21	HS12	11	49	9	-30.0000	1.7668e-006	1e-05
Algo22	-	12	51	10	-30.0000	3.4720e-005	1e-05
Algo21	HS29	14	67	11	-22.6274	5.6545e-006	1e-05
Algo22	-	16	84	14	-22.6274	1.3650e-005	1e-05
Algo21	HS31	12	455	28	-6.0000	7.1223e-006	1e-05
Algo22	-	10	385	23	-6.0000	1.6319e-005	1e-05
Algo21	HS34	13	176	33	-0.8340	1.5991e-006	1e-05
Algo22	-	11	155	24	-0.8340	1.5562e-005	1e-05
Algo21	HS35	8	125	16	0.1111	1.1247e-006	1e-05
Algo22	-	8	119	14	0.1111	1.6929e-005	1e-05
Algo21	HS43	22	132	45	-44.0000	4.2772e-006	1e-05
Algo22	-	14	126	34	-44.0000	1.7061e-005	1e-05
Algo21	HS100	27	562	51	680.6301	9.4936e-006	1e-05
Algo22	-	30	598	57	680.6301	2.6100e-005	1e-05
Algo21	HS113	25	752	95	24.3062	2.2422e-006	1e-05
Algo22	-	18	541	20	24.3062	2.8480e-006	1e-05
Algo21	S264	21	312	42	-44.1134	8.7955e-006	1e-05
Algo22	-	23	333	54	-44.1134	1.4441e-005	1e-05

第 3 章　求解极大极小问题的一个修正 SQP 算法

3.1　引言

本章考虑求解如下极大极小问题：

$$(P)\quad \min_{x\in R^n} F(x). \tag{3.1}$$

其中 $F(x)=\max\{f_j(x), j\in I\}$ 且 $f_j(x)$ 连续可微．问题 (3.1) 有较强的实际背景，它经常出现在许多工程设计问题中 (具体见文献 [90, 91])．

由于当 $f_i(x), i\in I$ 都是可微函数时，目标函数 $F(x)$ 一般也是不可微的．因此把传统的光滑优化方法直接用于求解极大极小问题可能很难找到其最优解．为克服这个困难，许多学者把问题 (3.1) 转化为如下等价问题：

$$\begin{array}{lll} & \min\limits_{(x,z)} & z \\ (P) & \text{s.t.} & f_j(x)\le z,\ \ j\in I, \end{array} \tag{3.2}$$

显然，问题 (3.2) 的 KKT 条件可表述如下：

$$\begin{aligned} &\begin{bmatrix}0\\1\end{bmatrix}+\sum_{j\in I}\lambda_j\begin{bmatrix}\nabla f_j(x)\\-1\end{bmatrix}=0,\\ &\lambda_j\ge 0, f_j(x)-z\le 0, \lambda_j(f_j(x)-z)=0, j\in I. \end{aligned} \tag{3.3}$$

上述关系等价于

$$\sum_{j\in I}\lambda_j\nabla f_j(x)=0,\ \sum_{j\in I}\lambda_j=1,\ \lambda_j(f_j(x)-F(x))=0,\ \lambda_j\ge 0, j\in I, \tag{3.4}$$

其中点 $x\in R^n$ 称作问题 (P) 的稳定点而 λ 称作乘子向量．

由于序列二次规划方法 (SQP) 是求解非线性优化问题的一类比较受欢迎的方法．因此许多学者尝试用 SQP 方法求解极大极小问题 (具体见文献 [92, 93, 94, 95, 97, 99])．其中，Zhou 和 Tits[95] 提出了一个算法：该算法的搜索方向通过求解两个二次规划得到，为避免 Maratos 效应，采用非单调线搜索技术．但是，该算法仅得到两步超线性收敛速度．为克服这个缺陷，Xue[97] 给出了一个改进算法，但为得到一步超线性收敛速度，需假设该算法是强收敛的，而且还假设了当 k 充分大时其步长恒为 1．最近，Zhu[98] 给出了一个求解极大极小优化问题的 SQP 算法，该算法的主搜索方向通过求解一个二次规划得到．为克服 Maratos 效应，其修正方向通过求解另一个二次规划得到．但和标准的非线性规划问题相比，为得到一步超线性收敛性，[98] 中的假设 H6 和 H7 是较强的．

本章中，我们提出一个求解极大极小问题 (3.1) 修正的 SQP 算法. 该算法的主搜索方向通过求解一个二次规划得到. 为克服 Maratos 效应，不同于文献 [95][98]，本章所给算法的高阶修正方向通过求解一个线性方程组得到，无需求解二次规划子问题. 在较弱的条件下，我们证明该算法具有全局收敛性和一步超线性收敛性.

3.2 算法描述

为表述方便，对 $x \in R^n$, 本章中我们使用如下记号：

$$f(x) = (f_j(x), j \in I)^T, I(x) = \{j \in I : f_j(x) = F(x)\}, \tag{3.5}$$

$$g_j(x) = \nabla f_j(x), j \in I, g(x) = (\nabla f_j(x), j \in I). \tag{3.6}$$

而且我们做如下假设.

H1 函数 f_j $(j \in I)$ 连续可微.

H2 R^{n+1} 中的向量组 $\left(\begin{bmatrix} g_j(x) \\ -1 \end{bmatrix}, j \in I(x)\right)$ 是线性无关的.

对给定点 x^k, 基于 H2, 我们使用文献 [86] 中的技术产生 ε- 近似积极集 $I_k \supseteq I(x^k)$ 使得矩阵 $A_k \triangleq \left(\begin{bmatrix} g_j(x^k) \\ -1 \end{bmatrix}, j \in I_k\right)$ 是列满秩的.

首先，我们给出如下记号：

$$y = (x^T, z)^T,\ L(y, \lambda) = z + \sum_{j \in I} \lambda_j (f_j(x) - z),\ \nabla_y L(y, \lambda) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \sum_{j \in I} \lambda_j \begin{bmatrix} \nabla f_j(x) \\ -1 \end{bmatrix},$$

$$G(x) = diag(f_j(x) - F(x)),\ \nabla c_j(x) = \begin{bmatrix} \nabla f_j(x) \\ -1 \end{bmatrix},\ \nabla c(x) = (\nabla c_j(x), j \in I),$$

$$M(x) = \nabla c(x)^T \nabla c(x) + G^2(x),\ \lambda(x) = -M^{-1}(x) \nabla c(x)^T \nabla f_0(x),\ \nabla f_0(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$\Phi(x, \lambda) = \begin{bmatrix} \nabla_y L(y, \lambda) \\ min\{F(x) - f_j(x), \lambda\} \end{bmatrix}.\ \rho(x, \lambda) = \sqrt{\|\Phi(x, \lambda)\|}.$$

下面我们定义近似积极集如下：

$$I(x, \varepsilon) = \{i : f_i(x) - F(x) + \varepsilon \rho(x, \lambda(x)) \geq 0\},$$

其中 ε 是一非负参数. 显然 (x^*, λ^*) 是问题 (P) 的稳定点当且仅当 $\Phi(x^*, \lambda^*) = 0$ 或 $\rho(x^*, \lambda^*) = 0$. 而且由文献 [86] 知在二阶充分条件和 Mangassarian-Fromovotz 约束规格下，对任意 $\varepsilon > 0$, 当 x 充分接近 x^* 时， $I(x, \varepsilon) = I(x^*)$.

该积极集估计技术通过如下步骤实现.

算法 A

步 (i) 给定初始参数 $\varepsilon = \varepsilon_{k-1} > 0$.

步 (ii) 生成 ε- 近似积极集 $I(x^k, \varepsilon)$ 和矩阵 N_k, 其中

$$N_k = \left(\begin{bmatrix} g_j(x^k) \\ -1 \end{bmatrix}, j \in I(x^k, \varepsilon) \right). \tag{3.7}$$

步 (iii) 如果 $\det\left(N_k^T N_k\right) \geq \varepsilon$, 则令 $I_k = I(x^k, \varepsilon)$, $A_k := N_k$ 和 $\varepsilon_k = \varepsilon$, 停; 否则令 $\varepsilon := \frac{1}{2}\varepsilon$ 重复步 (ii).

引理 3.2.1 *如果假设 (H1) 和 (H2) 成立, 则上述算法中的循环经有限步终止.*

该引理的证明类似于文献 [87] 中引理 1.1 和引理 2.8 的证明.

对当前迭代点 x^k 和一对称正定矩阵 $H_k = H(x^k)$, 我们求解如下二次规划子问题得到主搜索方向 d^k.

$$\begin{array}{lll} & \min & z + \frac{1}{2} d^T H_k d \\ (QP) & \text{s.t.} & f_j(x^k) + g_j(x^k)^T d - F(x^k) \leq z, \quad j \in I_k. \end{array} \tag{3.8}$$

下面的引理刻画了上述 QP 子问题的某些重要性质.

引理 3.2.2 *如果矩阵 H_k 是一对称正定矩阵, 则*

(i) 二次规划子问题 (3.8) 有最优解;

(ii) (z_k, d^k) 是 (3.8) 的最优解当且仅当它是 (3.8) 的 KKT 点.

证明: (i) 由于 $F(x^k) = \max\{f_j(x^k), j \in I\}$, 故 $(\bar{z}, \bar{d}) = (0, 0) \in R^{n+1}$ 是 (3.8) 的一可行解, 即 (3.8) 的可行集非空. 故二次规划子问题 (3.8) 有最优解.

(ii) 如果 (z_k, d^k) 是 (3.8) 的 KKT 点, 注意到 (3.8) 是一凸二次规划, 则它是 (3.8) 的 KKT 点. 反过来, 如果 (z_k, d^k) 是 (3.8) 的最优解, 注意到 Abadie 约束规格 ((3.8) 的约束都是线性的) 成立, 则 (z_k, d^k) 是 (3.8) 的 KKT 点.

引理 3.2.3 *如果假设 (H1) 和 (H2) 成立, 且 (z_k, d^k) 是 (3.8) 的最优解, 则*

(i) $z_k + \frac{1}{2}(d^k)^T H_k d^k \leq 0, z_k \leq 0$;

(ii) $d^k = 0 \Leftrightarrow z_k = 0, d^k = 0 \Leftrightarrow x^k$ *是问题 (P) 的 KKT 点;*

(iii) 如果 $d^k \neq 0$, 则 $z_k < 0$, 而且, d^k 是 $F(x)$ 在 x^k 处的下降方向.

证明: (i) 注意到 (0,0) 是 (3.8) 的可行解且 H_k 正定可得

$$z_k + \frac{1}{2}(d^k)^T H_k d^k \leq 0, \quad z_k \leq -\frac{1}{2}(d^k)^T H_k d^k \leq 0.$$

(ii) 如果 $d^k=0$, 那么由 (3.8) 知:

$$F(x^k)-f_j(x^k)+z_k\ge 0,\quad j\in I_k.$$

由 $I(x^k)\subseteq I_k$ 知 $z_k\ge 0$. 又 $z_k\le 0$, 因此 $z_k=0$.
反过来, 如果 $z_k=0$, 则 $\frac{1}{2}(d^k)^TH_kd^k=\frac{1}{2}(d^k)^TH_kd^k+z_k\le 0$, 结合 H_k 的正定性可得 $d^k=0$.
由引理 3.2.2(ii) 知: (3.8) 的最优解 (z_k,d^k) 是 (3.8) 的 KKT 点, 则存在乘子向量 $\lambda^k=(\lambda_j^k,j\in I_k,0_{I\setminus I_k})$ 满足

$$\begin{aligned}
&\begin{bmatrix}1\\ H_kd^k\end{bmatrix}+\sum_{j=1}^m\lambda_j^k\begin{bmatrix}-1\\ g_j(x^k)\end{bmatrix}=0,\\
&f_j(x^k)+g_j(x^k)^Td^k-F(x^k)-z_k\le 0, j\in I_k,\\
&\left(f_j(x^k)+g_j(x^k)^Td^k-F(x^k)-z_k\right)\lambda_j^k=0, j\in I_k,\\
&\lambda_j^k\ge 0, j\in I_k;\quad \lambda_j^k=0,\ j\in I\setminus I_k.
\end{aligned}\tag{3.9}$$

如果 $d^k=0$, 则 $z_k=0$, 从而由上述式子可得

$$\sum_{j=1}^m\lambda_j^kg_j(x^k)=0,$$

$$\begin{aligned}
&\sum_{j=1}^m\lambda_j^k=1, f_j(x^k)-F(x^k)\le 0, j\in I,\\
&\left(f_j(x^k)-F(x^k)\right)\lambda_j^k=0, \lambda_j^k\ge 0, j\in I.
\end{aligned}$$

因此 x^k 是 (P) 的稳定点.
反过来, 如果 x^k 是 (P) 的稳定点, 则 $z_k=0$ 和 $d^k=0$ 满足 (3.8) 的 KKT 条件. 因此由引理 3.2.2 知 (0,0) 是 (3.8) 的最优解. 因此 $d^k=0$.
(iii) 由 $z_k+\frac{1}{2}(d^k)^TH_kd^k\le 0, d^k\ne 0$ 和矩阵 H_k 的正定性知: $z_k<0$. 而且, 由 (3.8) 的约束知:

$$g_j(x^k)^Td^k\le z_k+F(x^k)-f_j(x^k)=z_k<0,\quad j\in I(x^k).$$

另一方面, 易知 $F(x)$ 在点 x 处沿着方向 d 的方向导数 $F'(x;d)$ 能被表示成

$$F'(x;d)=\lim_{\lambda\to 0^+}\frac{F(x+\lambda d)-F(x)}{\lambda}=\max\{g_j(x)^Td,\ \ j\in I(x)\}.\tag{3.10}$$

因此

$$F'(x^k;d^k)\le z_k<0,\tag{3.11}$$

即, d^k 是 $F(x)$ 在点 x^k 处的下降方向.

下面我们给出本章算法的算法步骤.

算法 B

有关参数: $\varepsilon_{-1}>0,\ \ \tau\in(2,3),\ \ \alpha\in(0,0.5),\ \ \beta\in(0,1)$.

步 0 初始数据: $x^0\in R^n$, 对称正定矩阵 $H_0\in R^{n\times n}$. 令 $k:=0$.

步 1 产生 ε- 近似积极集 I_k: 令 $\varepsilon=\varepsilon_{k-1}$, 由算法 A 产生近似积极集 I_k.

步 2 计算主搜索方向 d^k:

求解 (QP)(3.8) 得其解 (z_k,d^k), 其乘子向量为 $\lambda^k_{I_k}=(\lambda^k_j,j\in I_k)$. 如果 $d^k=0$, 则 x^k 是 (P) 的稳定点, 停; 否则, 进入步 3.

步 3 计算修正方向 $\tilde{d}^k$:

求解如下线性方程组得 $\tilde{d}^k$

$$\begin{bmatrix}\widetilde{H}_k & A_k\\ A_k^T & 0\end{bmatrix}\begin{bmatrix}\widetilde{t}\\ \widetilde{\gamma}\end{bmatrix}=\begin{bmatrix}0\\ -\|d^k\|^\tau e-\widetilde{f}_k\end{bmatrix},\tag{3.12}$$

其中 $\widetilde{H}_k=\begin{bmatrix}H_k & 0\\ 0 & 1\end{bmatrix},\widetilde{t}=\begin{bmatrix}\widetilde{d}\\ \widetilde{z}\end{bmatrix},e=(1,1,\cdots,1)^T\in R^{|I_k|}$ 和 $\widetilde{f}_k=(\widetilde{f}^k_j,j\in I_k)$, $\widetilde{f}^k_j=f_j(x^k+d^k)-F(x^k)-z_k,\ \ j\in I_k$. 如果 $\|\widetilde{d^k}\|\geq\|d^k\|$, 令 $\widetilde{d^k}=0$.

步 4 线搜索:

求序列 $\{1,\beta,\beta^2,\cdots\}$ 中满足如下不等式的第一个数作为 t_k

$$F(x^k+td^k+t^2\tilde{d}^k)\leq\max_{l=0,1,2}F(x^{k-l})-\alpha t(d^k)^TH_kd^k.\tag{3.13}$$

步 5 产生一个新的对称正定矩阵 H_{k+1}, 令 $x^{k+1}=x^k+t_kd^k+t_k^2\tilde{d}^k$ 和 $k:=k+1$, 返回到步 1.

为说明算法 B 是适定的, 我们给出如下引理.

引理 3.2.4 *如果 $d^k\neq 0$, 则步 4 中的线搜索是适定的. 即, 存在 $\bar{t}_k>0$ 使得 (3.13) 成立.*

证明 对于二次规划 (3.8), 由于 $(d,z)=(0,0)$ 为其可行点, 则

$$z_k+\frac{1}{2}(d^k)^TH_kd^k\leq 0.$$

假设对 $\lambda=\beta^j,j=1,2,\ldots,$ 线搜索 (3.13) 不能成立, 则从 (3.11), (3.10), $\alpha\in(0,0.5)$, $\beta\in(0,1)$ 和引理 3.2.3 (i) 知:

$$\begin{aligned}z_k &\geq F'(x^k;d^k)=\lim_{j\to\infty}\frac{F(x^k+\beta^jd^k)-F(x^k)}{\beta^j}=\lim_{j\to\infty}\frac{F(x^k+\beta^jd^k+\beta^{2j}\tilde{d}^k)-F(x^k)}{\beta^j}\\ &\geq\lim_{j\to\infty}\frac{F(x^k+\beta^jd^k+\beta^{2j}\tilde{d}^k)-\max\limits_{l=0,1,2}F(x^{k-l})}{\beta^j}\\ &\geq-\lim_{j\to\infty}\alpha(d^k)^TH_kd^k>-\tfrac{1}{2}(d^k)^TH_kd^k\geq z_k.\end{aligned}$$

这是一个矛盾！故步 4 中的线搜索是适定的.

3.3 全局收敛性分析

本节，我们将分析算法 B 的全局收敛性. 如果在步 2 中求得的 $d^k=0$, 则算法 B 在点 x^k 终止，且从引理 3.2.3(ii) 知 x^k 是问题 (P) 的稳定点. 如果 $d^k\neq 0$, 则从引理 3.2.3(iii) 知 d^k 是 $F(x)$ 在 x^k 处的下降方向.

我们不妨假设算法 B 产生一无限序列 $\{x^k\}$, 接下来的目标是证明 $\{x^k\}$ 的聚点 x^* 是问题 (P) 的稳定点. 为此，我们做如下的假设.

H3 矩阵序列 $\{H_k\}$ 是一致正定的. 即，存在两个正数 a 和 b 使得

$$a\|d\|^2\le d^TH_kd\le b\|d\|^2,\ \forall d\in R^n,\ \forall k. \tag{3.14}$$

H4 对任意 $x^0\in R^n$, 集合 $\Omega=\{x\in R^n: f(x)\le f(x^0)\}$ 是一紧集.

在本章余下部分，我们设 x^* 是 $\{x^k\}$ 的聚点. 因此，注意到 I_k 是有限集 I 的子集和引理 3.2.1, 不妨假设存在一子集 K 使得

$$x^k\to x^*,\quad I_k\equiv I',\ \forall k\in K. \tag{3.15}$$

下面的引理见文献 [95].

引理 3.3.1 *序列 $\{x^k\}$ 是有界的，且 $\{t_kd^k\}\to 0$ 和 $\{x^{k+1}-x^k\}\to 0$.*

引理 3.3.2 *如果假设 H1,H2,H3 成立，则*
(i) 序列 $\{z_k,k\in K\},\{d^k,k\in K\}$ 和 $\{\tilde{d}^k,k\in K\}$ 都是有界的；
(ii) $\lim\limits_{k\in K}d^k=\lim\limits_{k\in K}\tilde{d}^k=0,\lim\limits_{k\in K}z_k=0.$

证明(i) 由于 (0,0) 是 (3.8) 的可行解，注意到假设 H3 和 (3.8) 的约束可得

$$\begin{aligned}0\ &\ge z_k+\tfrac{1}{2}(d^k)^TH_kd^k\ge f_j(x^k)+g_j(x^k)^Td^k-F(x^k)+\tfrac{1}{2}(d^k)^TH_kd^k\\ &=g_j(x^k)^Td^k+\tfrac{1}{2}(d^k)^TH_kd^k\ge-\|g_j(x^k)\|\cdot\|d^k\|+\tfrac{1}{2}a\|d^k\|^2,\quad \forall j\in I(x^k),\forall k.\end{aligned}$$

上述不等式表明 $\{z_k,k\in K\}$ 和 $\{d^k,k\in K\}$ 是有界的. 由算法 B 步 3 中 $\tilde{d}^k$ 的定义可知 $\{\tilde{d}^k,k\in K\}$ 是有界的.
(ii) 类似于文献 [95] 中的定理 3.1, 可以证明 $\lim\limits_{k\in K}d^k=0$, 这表明 (ii) 的结论是成立的.

引理 3.3.3 *乘子序列 $\{\lambda^k=(\lambda^k_{I_k},0_{I\setminus I_k})\}$ 是有界的.*

证明 由 (3.8) 的 KKT 条件知：$\sum\limits_{j=1}^{m}\lambda^k_j=1$ 和 $\lambda^k_j\ge 0,j\in I$. 即 $\{\lambda^k\}$ 是有界的.

基于引理 3.3.1, 引理 3.3.2 和引理 3.3.3, 我们给出算法 B 的如下全局收敛性定理.

定理 3.3.1 *如果假设 H1,H2,H3,H4 成立，则算法 B 或在有限步终止于问题 (P) 的稳定点，或产生一无限序列 $\{x^k\}$，其每一聚点 x^* 都是问题 (P) 的稳定点.*

该定理的证明见文献 [95] 中的定理 3.1.

3.4 收敛速度分析

本节，我们将分析算法 B 的收敛速度. 首先，我们证明如下命题.

命题 3.4.1 *如果假设 H1 和 H2 成立，则对应于问题 (P) 的稳定点 $\tilde{x}$ 的乘子是唯一的.*

证明: 假设 $\tilde{\lambda},\tilde{\mu}$ 是关于问题 (P) 的稳定点 $\tilde{x}$ 的乘子，则从 (3.2) 知

$$\sum_{j\in I(\tilde{x})}\tilde{\lambda}_j\begin{bmatrix}g_j(\tilde{x})\\-1\end{bmatrix}=\begin{bmatrix}0\\-1\end{bmatrix},\ \sum_{j\in I(\tilde{x})}\tilde{\mu}_j\begin{bmatrix}g_j(\tilde{x})\\-1\end{bmatrix}=\begin{bmatrix}0\\-1\end{bmatrix},\ \tilde{\lambda}_{I\setminus I(\tilde{x})}=\tilde{\mu}_{I\setminus I(\tilde{x})}=0.$$

上面两个等式表明：

$$\sum_{j\in I(\tilde{x})}(\tilde{\mu}_j-\tilde{\lambda}_j)\begin{bmatrix}g_j(\tilde{x})\\-1\end{bmatrix}=0.$$

由假设 H2 可知：$\tilde{\lambda}=\tilde{\mu}$.

定理 3.4.1 *如果假设 H1,H2,H3,H4 成立，则* $\lim\limits_{k\to\infty}d^k=\lim\limits_{k\to\infty}\tilde{d}^k=0,\lim\limits_{k\to\infty}z_k=0.$

该引理的证明见文献 [95] 中的定理 3.1.

为证明算法 B 的超线性收敛性，我们进一步做如下假设.

H5(i) 函数 $f_j(x)$ $(j\in I)$ 都是二次连续可微的；

(ii) 在问题 (P) 的稳定点 (x^*,μ^*) 处，二阶充分条件成立.

$$d^T\nabla^2_{xx}L(x^*,\mu^*)d>0,\ \forall d\in R^n:\ d\neq 0,\ (g_j(x^*)-g_{t_0}(x^*))^Td=0,\ j\in I(x^*)\setminus\{t_0\},\ t_0\in I(x^*).$$

其中

$$\nabla^2_{xx}L(x^*,\mu^*)=\sum_{j\in I}\mu_j^*\nabla^2 f_j(x^*)=\sum_{j\in I(x^*)}\mu_j^*\nabla^2 f_j(x^*).$$

(iii) 在 (x^*,μ^*) 处严格互补条件成立. 即，$\mu_j^*>0,\ \forall j\in I(x^*)$.

下面我们证明 x^* 是问题 (P) 的孤立稳定点.

引理 3.4.1 *如果假设 (H2) 和 (H5) 成立，则 x^* 是问题 (P) 的孤立稳定点.*

该引理的证明类似于文献 [88] 中的性质 4.1.

定理 3.4.2 *如果假设 H2,H3,H4,H5 成立，则* $\lim_{k\to\infty} x^k = x^*$.

证明：由引理 3.4.1 知：x^* 是问题 (P) 的孤立稳定点. 而且，从定理 3.3.1 知：x^* 是 $\{x^k\}$ 的孤立聚点. 这和引理 3.3.1 一起表明 $\lim_{k\to\infty} x^k = x^*$. (参见文献 [88] 中的性质 4.1).

下面的引理表明即使在解处严格互补条件不成立，当 k 充分大时，积极集能被精确达到.

引理 3.4.2 *设 x^* 是问题 (P) 的稳定点，如果假设 (H5)(i)(ii) 成立，则存在 x^* 的一个邻域，使得对此邻域的任何点 x, 有*

$$I(x, \widehat{\varepsilon}) = I(x^*).$$

该引理见文献 [86].

引理 3.4.3 *如果假设 H1,H2 和 H5(iii) 成立，则当 k 充分大时，*

$$J_k = I(x^*),$$

其中 $J_k = \{j \mid f_j(x^k) + g_j(x^k)^T d^k - F(x^k) = z_k\}$.

证明：对任意 $j \in J_k$, 有

$$f_j(x^k) + g_j(x^k)^T d^k - F(x^k) = z_k.$$

考虑到定理 3.4.1 和定理 3.4.2, 通过对上面等式取极限得：$J_k \subseteq I(x^*)$. 反过来，对任意 $j \in I(x^*)$, 由 (H5)(iii) 知：当 k 充分大时，$\lambda_j^k > 0$. 注意到 (3.8) 的 KKT 条件可得：$I(x^*) \subseteq J_k$.

引理 3.4.4 *在上述假设下，当 k 充分大时，矩阵*

$$M_k \stackrel{def}{=} \begin{bmatrix} \widetilde{H}_k & A_k \\ A_k^T & 0 \end{bmatrix}.$$

是非奇异的，而且，存在常数 $C > 0$ 使得 $\|M_k^{-1}\| \leq C$.

类似于文献 [105] 中的引理 3.1 的证明可得上述引理.

引理 3.4.5 *如果假设 (H2),(H3) 和 (H5)(iii) 成立，则* $\|\widetilde{d}^k\| = O(\|d^k\|^2)$.

证明：由泰勒展式，$\widetilde{f}_j^k$ 的定义，定理 3.4.1, 引理 3.4.2 和引理 3.4.3 可得

$$\begin{aligned} \widetilde{f}_j^k &= f_j(x^k + d^k) - F(x^k) - z_k = f_j(x^k) + g_j(x^k)^T d^k + O(\|d^k\|^2) - F(x^k) - z_k \\ &= O(\|d^k\|^2), \ j \in I_k. \end{aligned}$$

因此，从引理 3.4.4, $\tau \in (2,3)$ 和 (3.3) 可知：$\|\widetilde{d}^k\| = O(\|d^k\|^2)$.

引理 3.4.6 *如果假设 H2,H3,H4,H5 成立，那么 (3.8) 对应于 (z_k, d^k) 的 KKT 乘子 $\lambda_{I_k}^k$ 满足*

$$\lim_{k\to\infty} \lambda^k = \mu^*, \lambda^k = (\lambda_{I_k}^k, 0_{I\setminus I_k}).$$

证明: 假设

$$\lim_{k\to\infty} \lambda^k \neq \mu^*,$$

则存在子集 K 和常数 $\bar{a} > 0$ 使得

$$\|\lambda^k - \mu^*\| \geq \bar{a}, k \in K.$$

注意到 $\lim\limits_{k\to\infty} x^k = x^*$ 和 $\{\lambda^k\}$ 的有界性，存在另一个子集 $K' \subseteq K$ 使得

$$x^k \to x^*, \|\lambda^k - \mu^*\| \geq \bar{a}, \lambda^k \to \lambda^*, k \in K' \subseteq K. \tag{3.16}$$

注意到定理 3.4.1 并对 (3.8) 的 KKT 条件两边取极限 ($k \in K', k \to \infty$) 可得

$$\begin{aligned}
&\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \sum_{j=1}^{m} \lambda_j^* \begin{bmatrix} -1 \\ g_j(x^*) \end{bmatrix} = 0, \\
&f_j(x^*) - F(x^*) \leq 0, (f_j(x^*) - F(x^*))\lambda_j^* = 0, \lambda_j^* \geq 0, j \in I.
\end{aligned}$$

从上可知 (x^*, λ^*) 是 (P) 的稳定点，因此 $\lambda^* = \mu^*$. 这和 (3.16) 矛盾！

为得到算法 B 的超线性收敛性，首先必须保证当 k 充分大时，搜索步长恒为 1. 为此，我们再做如下假设.

H6 假设 $\|P_k(\nabla_{xx}^2 L(x^k, \lambda^k) - H_k)d^k\| = o(\|d^k\|)$. 其中

$$P_k = I_n - R_k(R_k^T R_k)^{-1} R_k^T, \ R_k = R(x^k) = \left(g_j(x^k) - g_{t_0}(x^k), j \in I_*\setminus\{t_0\}\right).$$

下面的引理见文献 [95].

引理 3.4.7 *d^k 满足如下关系：*

$$d^k = P_k d^k + d_k',$$

其中 $d_k' = R_k(R_k^T R_k)^{-1}\overline{f}_j(x^k), \overline{f}_j(x^k) = [f_i(x^k) - f_j(x^k) : i \in I(x^*)\setminus\{j\}, j \in I(x^*)]$. 而且存在常数 $c_2 > 0$ 使得当 k 充分大时，

$$\|\overline{f}_j(x^k)\| \geq c_2\|d_k'\|, j = 1, 2, \cdots, m.$$

进一步，存在常数 $c_3 > 0$ 使得当 k 充分大时，对满足 $f_{j_k}(x^k) = F(x^k)$ 的任意 $j_k \in I(x^*)$，有

$$\overline{f}_j(x^k)^T \overline{\lambda}_k \leq -c_3\|d_k'\|,$$

其中 $\overline{\lambda}_k = [\lambda_{k,i} : i \in I(x^*) \setminus \{j_k\}]^T$.

定理 3.4.3 *如果假设 H2,H3,H4,H5,H6 成立，则当 k 充分大时，搜索步长恒为 1. 即，* $t_k \equiv 1$.

证明: 我们只需证明 (3.13) 当 $t=1$ 时成立即可. 首先由泰勒展式，定理 3.4.1, 3.4.2 和引理 3.4.5, 可得

$$\begin{aligned} f_i(x^k+d^k+\tilde{d}^k) &= f_i(x^k+d^k)+g_i(x^k+d^k)^T\tilde{d}^k+O(\|\tilde{d}^k\|^2) \\ &= f_i(x^k+d^k)+g_i(x^k)^T\tilde{d}^k+O(\|d^k\|^3), i\in I(x^*). \end{aligned} \tag{3.17}$$

从 (3.12) 可知

$$A_k^T\begin{bmatrix}\tilde{d}^k\\ \widetilde{z}_k\end{bmatrix} = -\|d^k\|^\tau e-\tilde{f}^k,\ \ g_i(x^k)^T\tilde{d}^k=\widetilde{z}_k-\|d^k\|^\tau-\tilde{f}_i^{\,k}, i\in I(x^*). \tag{3.18}$$

因此根据 (3.17), (3.18) 和 $\tilde{f}_i^k$ 的定义得

$$\begin{aligned} f_i(x^k+d^k+\tilde{d}^k) &= f_i(x^k+d^k)+\widetilde{z}_k-\|d^k\|^\tau-\tilde{f}_i^{\,k}+O(\|d^k\|^3) \\ &= F(x^k)+z_k+\widetilde{z}_k-\|d^k\|^\tau+O(\|d^k\|^3), \quad i\in I(x^*). \end{aligned} \tag{3.19}$$

故

$$f_j(x^k+d^k+\tilde{d}^k)=F(x^k)+z_k+\widetilde{z}_k-\|d^k\|^\tau+O(\|d^k\|^3), j\in I(x^*).$$

由上两式知

$$f_i(x^k+d^k+\tilde{d}^k)=f_j(x^k+d^k+\tilde{d}^k)+O(\|d^k\|^3), \forall i,j\in I(x^*). \tag{3.20}$$

注意到当 k 充分大时，有 $I(x^k+d^k+\tilde{d}^k)\subseteq I(x^*)$. 故对 $\forall j_k\in I(x^k+d^k+\tilde{d}^k)$, 有

$$F(x^k+d^k+\tilde{d}^k)=f_{j_k}(x^k+d^k+\tilde{d}^k)=f_j(x^k+d^k+\tilde{d}^k)+O(\|d^k\|^3), \forall j\in I(x^*).$$

另一方面，从 (3.9), 引理 3.3.2, 泰勒展式和引理 3.4.5 知

$$\sum_{j\in I(x^*)}\lambda_j^k=1, \lambda_j^kF(x^k+d^k+\tilde{d}^k)=\lambda_j^kf_j(x^k+d^k+\tilde{d}^k)+O(\|d^k\|^3),\ \ j\in I(x^*),$$

$$\begin{aligned} F(x^k+d^k+\tilde{d}^k) &= \sum_{j\in I(x^*)}\lambda_j^kF(x^k+d^k+\tilde{d}^k)=\sum_{j\in I(x^*)}\lambda_j^kf_j(x^k+d^k+\tilde{d}^k)+O(\|d^k\|^3) \\ &= \sum_{j\in I(x^*)}\lambda_j^k\left(f_j(x^k)+g_j(x^k)^T(d^k+\tilde{d}^k)+\tfrac{1}{2}(d^k)^T\nabla_{xx}^2f_j(x^k)d^k\right)+o(\|d^k\|^2). \end{aligned} \tag{3.21}$$

又根据 (3.9) 和引理 3.4.5 有

$$\sum_{j\in I(x^*)}\lambda_j^kg_j(x^k)^T(d^k+\tilde{d}^k)=-(d^k)^TH_kd^k+o(\|d^k\|^2), \tag{3.22}$$

和

$$\sum_{i\in I(x^*)}\lambda_i^k f_i(x^k)=\sum_{i\in I(x^*)}\lambda_j^k F(x^k)+\sum_{j\in I(x^*)}\lambda_i^k\{f_i(x^k)-f_j(x^k)\}=f_j(x^k)+\overline{f}_j(x^k)^T\overline{\lambda}_k. \tag{3.23}$$

因此，从 (3.21)-(3.23),(H3),(H5) 和引理 3.4.7, 可得

$$\begin{aligned}
F(x^k+d^k+\tilde{d}^k)\ &\le \max_{l=0,1,2}F(x^{k-l})-(d^k)^TH_kd^k-c_3\|d_k'\|\\
&+\tfrac{1}{2}(d^k)^T(\sum_{j\in I(x^*)}\lambda_j^k\nabla_{xx}^2f_j(x^k))d^k+o(\|d^k\|^2)\\
&=\max_{l=0,1,2}F(x^{k-l})-c_3\|d_k'\|-\tfrac{1}{2}(d^k)^TH_kd^k\\
&+\tfrac{1}{2}(d^k)^T\left(\sum_{j\in I(x^*)}\lambda_j^k\nabla_{xx}^2f_j(x^k)-H_k\right)d^k+o(\|d_k'\|)+o(\|d^k\|^2)\\
&=\max_{l=0,1,2}F(x^{k-l})-c_3\|d_k'\|-\tfrac{1}{2}(d^k)^TH_kd^k\\
&+\tfrac{1}{2}(d^k)^TP_k\left(\sum_{j\in I(x^*)}\lambda_j^k\nabla_{xx}^2f_j(x^k)-H_k\right)d^k+o(\|d_k'\|)+o(\|d^k\|^2)\\
&=\max_{l=0,1,2}F(x^{k-l})-c_3\|d_k'\|-\alpha(d^k)^TH_kd^k\\
&+(\alpha-\tfrac{1}{2})(d^k)^TH_kd^k+o(\|d_k'\|)+o(\|d^k\|^2)\\
&\le\max_{l=0,1,2}F(x^{k-l})-\alpha(d^k)^TH_kd^k+(\alpha-\tfrac{1}{2})a\|(d^k)\|^2\\
&-c_3\|d_k'\|+o(\|d_k'\|)+o(\|d^k\|^2).
\end{aligned}$$

注意到 $\alpha\in(0,\frac{1}{2})$, 易知当 k 充分大时

$$F(x^k+d^k+\tilde{d}^k)\le\max_{l=0,1,2}F(x^{k-l})-\alpha(d^k)^TH_kd^k,$$

即， (3.13) 当 $t=1$ 和 k 充分大时成立.

下面的引理对分析该算法的超线性收敛性是必要的.

引理 3.4.8 如果假设 *H2,H3,H4,H5* 成立，则当 k 充分大时，矩阵

$$G_k=\begin{bmatrix}P_k\nabla_{xx}^2L(x^*,\mu^*)\\ R_k^T\end{bmatrix} \tag{3.24}$$

是列满秩的.

该引理的类似于引理 3.4.4.

定理 3.4.4 如果假设 *H2,H3,H4,H5,H6* 成立，则算法 B 是超线性收敛的，即，序列 $\{x^k\}$ 满足

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

证明：由 (3.9) 可得

$$H_kd^k + \sum_{j\in J}\lambda_j^k g_j(x^k) = 0,\ \ \sum_{j\in J}\lambda_j^k = 1,\ \ \sum_{j\in J}\lambda_j^k g_{t_0}(x^k) = g_{t_0}(x^k), \tag{3.25}$$

$$g_j(x^k)^T d^k = F(x^k) + z_k - f_j(x^k), j\in J. \tag{3.26}$$

又由 (3.25) 和 (3.26) 可知

$$H_kd^k + \sum_{j\in J}\lambda_j^k(g_j(x^k) - g_{t_0}(x^k)) = H_kd^k + R_k\lambda_J^k = -g_{t_0}(x^k), \tag{3.27}$$

$$(g_j(x^k)-g_i(x^k))^Td^k = f_i(x^k)-f_j(x^k),\ \ \forall i,j\in J;\ \ R_k^Td^k = (f_{t_0}(x^k)-f_j(x^k), j\in J). \tag{3.28}$$

定义 $h(x)$ 如下

$$h(x) = \sum_{j\in J}\mu_j^*(g_j(x) - g_{t_0}(x)) = R(x)\mu_J^*.$$

则从泰勒展式和 $\sum\limits_{j\in J}\mu_j^* = 1$ 知

$$\begin{aligned}
h(x^k) &= R_k\mu_J^* = h(x^*) + \nabla h(x^*)^T(x^k - x^*) + o(\|x^k - x^*\|)\\
&= \sum_{j\in J}\mu_j^*(g_j(x^*) - g_{t_0}(x^*)) + \sum_{j\in J}\mu_j^*(\nabla_{xx}^2 f_j(x^*) - \nabla_{xx}^2 f_{t_0}(x^*))(x^k - x^*) + o(\|x^k - x^*\|)\\
&= \sum_{j\in J}\mu_j^*(g_j(x^*) - g_{t_0}(x^*)) + (\sum_{j\in J}\mu_j^*\nabla_{xx}^2 f_j(x^*) - \nabla_{xx}^2 f_{t_0}(x^*))(x^k - x^*) + o(\|x^k - x^*\|)\\
&= \sum_{j\in J}\mu_j^*(g_j(x^*) - g_{t_0}(x^*)) + \nabla_{xx}^2 L(x^*,\mu^*)(x^k - x^*)\\
&\quad -\nabla_{xx}^2 f_{t_0}(x^*)(x^k - x^*) + o(\|x^k - x^*\|).
\end{aligned}$$

而由 P_k 的定义有 $P_kR_k = 0$. 故

$$\begin{aligned}
0 &= P_kR_k\mu_J^* = P_kh(x^k)\\
&= P_k\sum_{j\in J}\mu_j^*(g_j(x^*) - g_{t_0}(x^*)) + P_k\nabla_{xx}^2 L(x^*,\mu^*)(x^k - x^*) - P_k\nabla_{xx}^2 f_{t_0}(x^*)(x^k - x^*)\\
&\quad +o(\|x^k - x^*\|).
\end{aligned}$$

这和 $\sum\limits_{j\in J}\mu_j^* g_j(x^*) = 0$ 和 $\sum\limits_{j\in J}\mu_j^* = 1$ 一起表明

$$\begin{aligned}
&P_k\nabla_{xx}^2 L(x^*,\mu^*)(x^k - x^*)\\
&= P_k\nabla_{xx}^2 f_{t_0}(x^*)(x^k - x^*) - P_k\sum_{j\in J}\mu_j^*(g_j(x^*) - g_{t_0}(x^*)) + o(\|x^k - x^*\|)\\
&= P_k\nabla_{xx}^2 f_{t_0}(x^*)(x^k - x^*) - P_k(\sum_{j\in J}\mu_j^* g_j(x^*) - g_{t_0}(x^*)) + o(\|x^k - x^*\|)\\
&= P_k\nabla_{xx}^2 f_{t_0}(x^*)(x^k - x^*) + P_kg_{t_0}(x^*) + o(\|x^k - x^*\|).
\end{aligned} \tag{3.29}$$

而从定理 3.4.3, (3.29), 引理 3.4.4, $\sum_{j\in J}\mu_j^* g_j(x^*)=0$ 和 (3.28) 有

$$\begin{aligned}
&P_k\nabla^2_{xx}L(x^*,\mu^*)(x^{k+1}-x^*)\\
&=P_k\nabla^2_{xx}L(x^*,\mu^*)(x^k-x^*)+P_k\nabla^2_{xx}L(x^*,\mu^*)(d^k+\tilde{d}^k)\\
&=P_k\nabla^2_{xx}f_{t_0}(x^*)(x^k-x^*)+P_kg_{t_0}(x^*)+P_k\nabla^2_{xx}L(x^*,\mu^*)d^k+o(\|x^k-x^*\|)+o(\|d^k\|)\\
&=P_k\nabla^2_{xx}f_{t_0}(x^*)(x^k-x^*)+P_kg_{t_0}(x^*)+P_k(\nabla^2_{xx}L(x^*,\mu^*)-H_k)d^k+P_kH_kd_k+o(\|x^k-x^*\|)\\
&\quad+o(\|d^k\|)\\
&=P_k\nabla^2_{xx}f_{t_0}(x^*)(x^k-x^*)+P_kg_{t_0}(x^*)+P_kH_kd^k+o(\|x^k-x^*\|)+o(\|d^k\|).
\end{aligned}$$

由 (3.27) 知 $P_kH_kd^k=-P_kg_{t_0}(x^k)$, 结合上面最后一个等式和泰勒展式可得

$$\begin{aligned}
&P_k\nabla^2_{xx}L(x^*,\mu^*)(x^{k+1}-x^*)\\
&=P_k\left(\nabla^2_{xx}f_{t_0}(x^*)(x^k-x^*)+g_{t_0}(x^*)-g_{t_0}(x^k)\right)+o(\|x^k-x^*\|)+o(\|d^k\|)\\
&=o(\|x^k-x^*\|)+o(\|d^k\|),
\end{aligned}$$

即,

$$P_k\nabla^2_{xx}L(x^*,\mu^*)(x^{k+1}-x^*)=o(\|x^k-x^*\|)+o(\|d^k\|). \tag{3.30}$$

另外, 由 $J_k=I(x^*)$ 和泰勒展式可得

$$0=f_j(x^*)-f_{t_0}(x^*)=f_j(x^k)-f_{t_0}(x^k)+(g_j(x^k)-g_{t_0}(x^k))^T(x^*-x^k)+o(\|x^k-x^*\|),j\in J_k,$$

$$0=(f_j(x^k)-f_{t_0}(x^k),j\in J)+R_k^T(x^*-x^k)+o(\|x^k-x^*\|). \tag{3.31}$$

因此由 (3.31) 和 (3.28) 也有

$$R_k^T(x^k-x^*)=(f_j(x^k)-f_{t_0}(x^k),j\in J)+o(\|x^k-x^*\|),$$

$$\begin{aligned}
R_k^T(x^{k+1}-x^*)&=R_k^T(x^k-x^*)+R_k^T(d^k+\tilde{d}^k)\\
&=(f_j(x^k)-f_{t_0}(x^k),j\in J)+R_k^Td^k+o(\|x^k-x^*\|)+o(\|d^k\|)\\
&=o(\|x^k-x^*\|)+o(\|d^k\|).
\end{aligned}$$

即,

$$R_k^T(x^{k+1}-x^*)=o(\|x^k-x^*\|)+o(\|d^k\|). \tag{3.32}$$

故从 (3.30) 和 (3.32) 有

$$\begin{bmatrix}P_k\nabla^2_{xx}L(x^*,\mu^*)\\ R_k^T\end{bmatrix}\begin{bmatrix}x^{k+1}-x^*\\ 0\end{bmatrix}=o(\|d^k\|)+o(\|x^k-x^*\|).$$

由引理 3.4.5 和引理 3.4.8 知

$$\begin{aligned}
\|x^{k+1}-x^*\|&=o(\|d^k\|)+o(\|x^k-x^*\|)\\
&=o(\|d^k+\tilde{d}^k\|)+o(\|x^k-x^*\|)\\
&=o(\|(x^{k+1}-x^*)-(x^k-x^*)\|)+o(\|x^k-x^*\|)\\
&\le o(\|x^{k+1}-x^*\|)+o(\|x^k-x^*\|).
\end{aligned}$$

因此,

$$\frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|}\left(1-\frac{o(\|x^{k+1}-x^*\|)}{\|x^{k+1}-x^*\|}\right)\le\frac{o(\|x^k-x^*\|)}{\|x^k-x^*\|}.$$

即,

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

3.5 数值试验

本节，我们对 3.2 节中的算法 (为表述方便，我们用 Algo31 表示) 和文献 [95] 中的算法 (用 Algo32 表示) 进行了数值比较. 两种算法都用 MATLAB 6.5 编程，都在 Windows XP 的个人计算机实现. 其中 (3.8) 和 (3.12) 都使用优化工具箱求解.

整个实验中，我们使用如下 BFGS 公式:

$$H_{k+1}=H_k-\frac{H_ks^k(s^k)^TH_k}{(s^k)^TH_ks^k}+\frac{\widetilde{y}^k(\widetilde{y}^k)^T}{(s^k)^T\widetilde{y}^k},\quad (k\ge 0),$$

其中

$$s^k=x^{k+1}-x^k,\ \widetilde{y}^k=y^k+a_k(\gamma_ks^k+B_kB_k^Ts^k),\ \gamma_k=\min\{\|d^k\|^2,\xi\},\quad \xi\in(0,1),$$

$$y^k=\nabla_xL(x^{k+1},\lambda^k)-\nabla_xL(x^k,\lambda^k),\ B_k=(\nabla f_j(x^k),j\in J_k),$$

$$\nabla_xL(x,\lambda)=\sum_{j\in I}\lambda_j\nabla f_j(x),$$

$$a_k=\begin{cases}0, & \text{if}\quad (s^k)^Ty^k\ge\delta\|s^k\|^2,\delta\in(0,1);\\ 1, & \text{if}\quad 0\le(s^k)^Ty^k<\delta\|s^k\|^2;\\ 1+\frac{\gamma_k\|s^k\|^2-(s^k)^Ty^k}{\gamma_k\|s^k\|^2+(s^k)^TB_k(B_k)^Ts^k}, & \text{otherwise}.\end{cases}$$

整个实验中，令

$$H_0=I_n,\quad \tau=2.5,\quad \beta=0.6,\quad \alpha=0.45,\quad \varepsilon_{-1}=2.1.$$

表 3.1 的测试问题来自文献 [97] 和文献 [100]. 所选测试问题的初始点和文献 [97] 和文献 [100] 中相同. 表 3.1 的列表示如下意义: **prob** 表示来自文献 [97] 和文献 [100] 的测试问题; **Ni** 表示求解的迭代数; **objective**, **dnorm** 和 **eps** 表示目标函数值，模 d^k 和精度 ϵ.

表 3.1 数值结果

Algo	Prob	Ni	objective	dnorm	eps
Algo31	1	10	1.9522	4.2478e-006	1e-05
Algo32	-	11	1.9522	9.6073e-006	1e-05
Algo31	2	30	2.0000	1.9868e-014	1e-05
Algo32	-	28	2.0000	5.2721e-006	1e-05
Algo31	3	11	-44.0000	8.2939e-006	1e-05
Algo32	-	11	-43.9900	1.2939e-005	1e-05
Algo31	4	22	0.6164	8.7974e-006	1e-05
Algo32	-	21	0.6164	1.0136e-006	1e-05
Algo31	5	14	3.5997	1.8319e-006	1e-05
Algo32	-	15	3.5997	8.6761e-006	1e-05
Algo31	8	23	680.6301	1.5379e-006	1e-05
Algo32	-	24	680.6380	2.7637e-006	1e-05
Algo31	9	35	24.3012	2.5676e-006	1e-05
Algo32	-	39	24.3062	1.2946e-006	1e-05
Algo31	Vardi-3	10	-48.0158	4.5843e-008	1e-05
Algo32	-	12	-48.0158	6.3061e-008	1e-05

第 4 章　求解约束优化问题一个组合可行与下降的可行点 SQP 算法

4.1　引言

本章继续考虑求解如下不等式约束优化问题:

$$
(P)\quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_j(x) \le 0,\ \ j \in I, \end{array} \tag{4.1}
$$

其中 m 是正整数, 函数 f, $g_j(j\in I): R^n \to R$ 连续可微, 我们记 $X=\{x: g_i(x)\le 0, i\in I\}$.

Panier 和 Tits 在文献 [88] 中提出一个可行点 SQP 算法. 该算法的主搜索方向通过求解如下二次规划得到.

$$
(QP)\quad \begin{array}{ll} \min & \nabla f(x^k)^T d + \frac{1}{2} d^T H_k d \\ \text{s.t.} & g_j(x^k) + \nabla g_j(x^k)^T d \le 0, \quad j \in I, \end{array} \tag{4.2}
$$

由于迭代点 x^k 满足问题 (P) 的所有约束, 故其对应二次规划子问题总是有解的. 为避免 Maratos 效应, 一个二阶校正方向通过解一线性最小二乘问题得到. 但是, 为得到该算法的收敛性, 该算法需求出一个一阶可行下降方向. 后来, Panier 和 Tits [101] 对上述算法进行了改进, 提出一个组合可行与下降的超线性收敛的 SQP 算法. 该算法的主搜索方向由二次规划 (4.2) 的解和另一二次规划的解的凸组合得到. 为避免 Maratos 效应, 需求解第三个二次规划子问题得到一个二阶校正方向. 为减少每步迭代的计算量, Zhu [102] 提出一个简化可行点 SQP 算法, 其主搜索方向由二次规划 (4.2) 的解和一线性方程组的解的凸组合得到. 为避免 Maratos 效应, 一个二阶校正方向通过解另一线性方程组得到. 而且在合适的条件下, 该算法被证明是全局收敛和超线性收敛的. 但是, Zhu 的算法中的线性方程组涉及乘子估计. 正如文献 [103] 中所指出的, 当近似积极约束所对应的乘子估计非常小时, 这些线性方程组可能会变为病态的, 从而导致算法不稳定.

本章中, 我们提出一个求解问题 (P) 的可行点 SQP 算法. 在该算法的每一个迭代步, 通过分别求解二次规划 (4.3) 和一线性方程组得到一个下降方向和一个可行方向, 在此基础上做一适当组合得到可行下降方向. 为避免 Maratos 效应, 高阶校正方向通过解一线性方程组得到. 而且在适当的条件下, 该算法被证明是全局收敛和超线性收敛的. 本章所给算法的优点是: 算法中线性方程组不涉及乘子估计, 所求解的两个线性方程组的系数矩阵相同且比文献 [102] 中系数矩阵的结构简单, 因而可减少计算量.

4.2 算法描述

对 $x \in X$, 定义积极集为

$$I(x) = \{i \in I : g_i(x) = 0\}.$$

本章中，我们假设可行集 X 非空和下面的假设成立.

H1 对 $x \in X$, 向量组 $\{\nabla g_j(x), j \in I(x)\}$ 线性无关.

对 $x \in X$, 我们也使用如下积极集估计技术，有关记号见第 2 章 2.2 节.

$$A(x;\varepsilon) = \{i : g_i(x) + \varepsilon\rho(x,\lambda(x)) \geq 0\}.$$

下面给出该算法的算法步骤.

有关参数 $\tau \in (2,3)$, $\beta \in (0,1)$, $\alpha \in (0,\frac{1}{2})$, $\delta > 2$, $\rho > 0$, $\gamma > 0$.

初始数据 给定初始可行点 $x^1 \in X$, 一个对称正定矩阵 H_1 和 $\varepsilon^0 > 0$. 令 $k = 1$.

步 1 令 $\varepsilon = \varepsilon^{k-1}$.

步 2 令 $A^k(\varepsilon) = A(x^k, \varepsilon)$. 如果 $\nabla g_{A^k(\varepsilon)}(x^k)$ 非满秩，则令 $\varepsilon := \sigma\varepsilon$, 进入步 2.(其中 $\nabla g_{A^k(\varepsilon)}(x^k) = (\nabla g_j(x^k),\ j \in A^k(\varepsilon)))$

步 3 令 $\varepsilon^k = \varepsilon$, $A^k = A^k(\varepsilon)$.

步 4(计算主搜索方向)

1.1 计算下降方向 d_0^k:

$$\begin{array}{lll} & \min & \nabla f(x^k)^T d + \frac{1}{2} d^T H_k d \\ (QP) & \text{s.t.} & g_j(x^k) + \nabla g_j(x^k)^T d \leq 0, \quad j \in A^k, \end{array} \tag{4.3}$$

通过求解上述二次规划得 d_0^k, 令 $u_{A^k}^k$ 是对应 Lagrange 乘子. 如果 $d^k = 0$, 则 x^k 是问题 (P) 的 KKT 点，停；否则进入 1.2.

1.2 计算可行方向 d_1^k:

求解如下线性方程组：

$$\begin{bmatrix} H_k & \nabla g_{A^k}(x^k) \\ \nabla g_{A^k}(x^k)^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \lambda \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) \\ -\|d_0^k\|^\delta e_{A^k} \end{bmatrix}, \tag{4.4}$$

其中 $e_{A^k} = (1, ..., 1)^T \in R^{|A^k|}$. 令 $(d_1^k, \pi_{A^k}^k)$ 是其解.

1.3 计算可行下降方向 d^k:

建立 d_0^k 和 d_1^k 的一个组合：

$$d^k = d_0^k + \rho_k d_1^k, \tag{4.5}$$

$$\rho_k = \begin{cases} \rho, & \text{if } \nabla f(x^k)^T d_1^k \leq 0; \\ \frac{-\nabla f(x^k)^T d_0^k}{\nabla f(x^k)^T d_1^k + \gamma}, & otherwise. \end{cases} \tag{4.6}$$

步 5 通过求解如下线性方程组得高阶修正方向 $\widetilde{d}^k$:

$$\begin{bmatrix} H_k & \nabla g_{A^k}(x^k) \\ \nabla g_{A^k}(x^k)^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ -\|d^k\|^\tau e_{A^k} - g_{A^k}(x^k + d^k) \end{bmatrix}, \tag{4.7}$$

如果 $\|\widetilde{d}^k\| > \|d^k\|$, 令 $\widetilde{d}^k = 0$.

步 6(线搜索)

求序列 $\{1, \beta, \beta^2, ...\}$ 中满足如下不等式组的第一个数作为 λ_k

$$f(x^k + \lambda d^k + \lambda^2 \widetilde{d}^k) \le f(x^k) + \alpha\lambda \nabla f(x^k)^T d^k, \tag{4.8}$$

$$g_j(x^k + \lambda d^k + \lambda^2 \widetilde{d}^k) \le 0, \quad \forall j \in I. \tag{4.9}$$

步 7 令 $x^{k+1} = x^k + \lambda_k d^k + \lambda_k^2 \widetilde{d}^k$. 生成一个新的对称正定矩阵 H_{k+1}, 令 $k := k+1$, 返回到步 1.

注 4.2.1 *不同于文献 [102] 中的算法, 上述算法中线性方程组 (4.4),(4.7) 的系数矩阵不涉及乘子估计.*

4.3 全局收敛性分析

为讨论方便, 对任意的 k, 我们令 $u_i^k = \pi_i^k = 0, \ \forall i \in I \setminus A^k$.

本节, 我们首先证明该算法是适定的.

引理 4.3.1 *如果假设 H1 成立, 则算法步 2 中的循环经有限步终止.*

该引理的证明类似于文献 [87] 中引理 1.1 和引理 2.8 的证明.

引理 4.3.2 *如果 H_k 是对称正定的, 则在假设 H1 下, 线性方程组 (4.4) 和 (4.7) 总存在唯一最优解.*

证明: 由 H_k 的对称正定性和假设 H1 知

$$\begin{bmatrix} H_k & \nabla g_{A^k}(x^k) \\ \nabla g_{A^k}(x^k)^T & 0 \end{bmatrix}$$

是非奇异的. 因此结论成立.

引理 4.3.3 *(1) 如果 $d_0^k = 0$, 则 x^k 问题 (P) 的 KKT 点.*

(2) 如果 $d_0^k \neq 0$, 则

$$\nabla f(x^k)^T d_0^k \le -\frac{1}{2}(d_0^k)^T H_k d_0^k, \quad \nabla f(x^k)^T d^k \le \theta \nabla f(x^k)^T d_0^k (\theta \in (0,1)),$$

$$\nabla g_j(x^k)^T d^k < 0, \quad j \in I(x^k).$$

证明: (1) 由问题 (P) 和 (4.3) 的 KKT 点的定义可知, x^k 问题 (P) 的 KKT 点.
(2) 如果 $d_0^k \neq 0$, 由 (4.3) 可得:

$$\nabla f(x^k)^T d_0^k \leq -\frac{1}{2}(d_0^k)^T H_k d_0^k < 0,\ \nabla g_j(x^k)^T d_0^k \leq 0,\ j \in I(x^k).$$

而从 (4.4) 可知:

$$\nabla g_j(x^k)^T d_1^k = -\|d_0^k\|^\delta.$$

又 $\rho_k > 0$, 故

$$\nabla g_j(x^k)^T d^k = \nabla g_j(x^k)^T d_0^k + \rho_k \nabla g_j(x^k)^T d_1^k \leq \rho_k \nabla g_j(x^k)^T d_1^k < 0, j \in I(x^k).$$

由 d^k 的定义可得

$$\nabla f(x^k)^T d^k = \nabla f(x^k)^T d_0^k + \rho_k \nabla f(x^k)^T d_1^k.$$

而从 ρ_k 的定义知: $\nabla f(x^k)^T d^k \leq \theta \nabla f(x^k)^T d_0^k$.

引理 4.3.4 *步 6 中的线搜索是适定的.*

由引理 4.3.3, 类似于引理 2.2.5 的证明, 可得上述引理.

为分析 4.2 节中算法的全局收敛性, 我们做如下假设.

H2 该算法产生的序列 $\{x^k\}$ 是有界的.

H3 对任意的 k 和所有的 $d \in R^n$, 存在常数 a, $b > 0$ 使得

$$a\|d\|^2 \leq d^T H_k d \leq b\|d\|^2.$$

不妨设 x^* 是 $\{x^k\}$ 的一个聚点. 注意到 A^k 和 J_k 是有限集 I 的子集, 故存在子集 K 使得

$$\lim_{k \in K} x^k = x^*, \quad A^k \equiv A, \quad J_k \equiv J, \ \forall k \in K, \tag{4.10}$$

其中

$$J_k = \{j \in A^k : f_j(x^k) + \nabla f_j(x^k)^T d^k = 0\}.$$

引理 4.3.5 *如果假设 H2,H3 成立, 则序列* $\{d_0^k : k \in K\}, \{d_1^k : k \in K\}$, $\{\widetilde{d}^k : k \in K\}, \{u^k : k \in K\}$ *和* $\{\pi^k : k \in K\}$ *都是有界的.*

证明: 由 x^k 的可行性和 (4.3) 可知:

$$-\|\nabla f(x^k)\|\|d_0^k\| + a\|d_0^k\|^2 \leq \nabla f(x^k)^T d_0^k + \frac{1}{2} {d_0^k}^T H_k d_0^k < 0.$$

又 $\|\nabla f(x^k)\|, k \in K$ 有界, 故不难证明 $\{d_0^k : k \in K\}$ 有界.

由引理 4.3.2 和 $\{d_0^k : k \in K\}$ 有界性易证 $\{\widetilde{d}^k : k \in K\}$,$\{u^k : k \in K\}$ 和 $\{\pi^k : k \in K\}$ 的有界性.

基于引理 4.3.5, 我们给出如下全局收敛性定理. 首先，我们不妨设存在一个子集 K, 使得

$$x^k \to x^*, H_k \to H_*, d_0^k \to d_0^*, u^k \to u^*, d^k \to d^*, \lambda^k \to \lambda^*,\ k \in K.$$

定理 4.3.1 *如果假设 H1,H2,H3 成立，则该算法或在有限步终止于问题 (P) 的 KKT 点，或产生一无限序列 $\{x^k\}$, 其每一聚点 x^* 都是问题 (P) 的 KKT 点.*

证明: 该定理的第一个结论由引理 4.3.3 易得. 下证第二个结论. 由 $\{f(x^k)\}$ 的单调下降性和 f 的连续性并结合 $x^k \to x^*(k \in K)$ 可知:

$$f(x^k) \to f(x^*),\ k \to \infty. \tag{4.11}$$

由 (4.3) 易知: d_0^* 是下面二次规划的唯一解, u_A^* 是相应的 Lagrange 乘子.

$$\begin{array}{lll} & \min & \nabla f(x^*)^T d + \frac{1}{2} d^T H_* d \\ (QP') & \text{s.t.} & g_j(x^*) + \nabla g_j(x^*)^T d \le 0, \quad j \in A, \end{array}$$

如果 $d_0^* = 0$, 则由 (QP') 的 KKT 点的定义知: x^* 都是问题 (P) 的 KKT 点. 假设 $d_0^* \neq 0$, 则 $\nabla f(x^*)^T d_0^* < 0$.

由 (4.4) 易知: (d_1^*, π_A^*) 是下面线性方程组的唯一解:

$$\begin{bmatrix} H_* & \nabla g_A(x^*) \\ \nabla g_A(x^*)^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \lambda \end{bmatrix} = \begin{bmatrix} -\nabla f(x^*) \\ -\|d_0^*\|^\delta e_A \end{bmatrix}$$

类似于 d^k 的定义，我们可定义 d_*, 模仿引理 4.3.3 的证明可得:

$$\nabla f(x^*)^T d_0^* < 0,\ \ \nabla f(x^*)^T d_* < 0,\ \ \nabla g_j(x^*)^T d_* < 0,\ \ j \in I(x^*).$$

由引理 4.3.4 的证明，易知步 6 中的线搜索所得步长 t_k 满足下面的关系:

$$t_k \ge t_* > 0,\ k \in K.$$

因此由 (4.8) 和上面的关系式，可得:

$$0 = \lim_{k \in K} (f(x^{k+1}) - f(x^k)) \le \lim_{k \in K} \alpha t_k \nabla f(x^k)^T d^k \le \alpha t_* \theta \nabla f(x^*)^T d_* < 0.$$

上述矛盾表明: $d_0^* = 0$. 故结论成立.

4.4 收敛速度分析

本节, 我们将分析 4.2 节中算法的收敛速度. 为此, 我们进一步做如下假设.

H4(i) 函数 $f(x)$, $g_j(x)(j \in I)$ 二阶连续可微.

(ii) 在 KKT 点对 (x^*, u^*) 处二阶充分条件成立, 即,

$$d^T \nabla_{xx}^2 L(x^*, u^*) d > 0, \quad \forall d \in \Omega \stackrel{def}{=} \{d \in R^n : d \neq 0, \nabla g_{I(x^*)}(x^*)^T d = 0\}, \tag{4.12}$$

其中

$$L(x,u) = f(x) + \sum_{j \in I} u_j g_j(x). \tag{4.13}$$

(iii) 在 (x^*, u^*) 处严格互补条件成立, 即, $u^* - g(x^*) > 0$.

引理 4.4.1 *(i) 如果假设 H1,H2 成立, 则* $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$ *和* $\lim_{k\to\infty} x^k = x^*$.
(ii) 如果假设 H1,H2,H3 成立, 则 $\lim_{k\to\infty} d_0^k = 0, \lim_{k\to\infty} d_1^k = 0, \lim_{k\to\infty} d^k = 0, \lim_{k\to\infty} \widetilde{d}^k = 0.$

证明: (i) 由 (4.8),(4.11) 和引理 4.3.3 有

$$\begin{aligned} 0 = \lim_{k\to\infty} (f(x^{k+1}) - f(x^k)) & \leq \lim_{k\to\infty} \alpha \lambda_k \nabla f(x^k)^T d^k \\ & \leq \alpha\theta\lambda_k \nabla f(x^k)^T d_0^k \\ & \leq -\lim_{k\to\infty} \tfrac{1}{2}\alpha a \theta \lambda_k \|d_0^k\|^2 \leq 0. \end{aligned}$$

故

$$\lim_{k\to\infty} \lambda_k \nabla f(x^k)^T d^k = 0, \lim_{k\to\infty} \lambda_k \nabla f(x^k)^T d_0^k = 0, \lim_{k\to\infty} \lambda_k \|d_0^k\| = 0.$$

另外, 由 (4.4) 知

$$\begin{aligned} \nabla f(x^k)^T d_1^k & = -(d_1^k)^T H_k d_1^k - (d_1^k)^T \nabla g_{A^k}(x^k) \pi_{A^k}^k \\ & = -(d_1^k)^T H_k d_1^k + \sum_{i \in I} \|d_0^k\|^\delta \pi_i^k \\ & \leq -a\|d_1^k\|^2 + |A^k| \|d_0^k\|^\delta c, \end{aligned}$$

其中 $|\pi_i^k| \leq c$, c 为一正常数 (由 π_i^k 的有界性可得).

又

$$\begin{aligned} 0 & = \lim_{k\to\infty} \lambda_k \nabla f(x^k)^T d^k - \lim_{k\to\infty} \lambda_k \nabla f(x^k)^T d_0^k \\ & = \lim_{k\to\infty} \rho_k \lambda_k \nabla f(x^k)^T d_1^k \\ & \leq -a\rho_k \lambda_k \|d_1^k\|^2 + \rho_k \lambda_k |A^k| \|d_0^k\|^\delta c \leq \rho_k \lambda_k |A^k| \|d_0^k\|^\delta c. \end{aligned}$$

因此

$$\lambda_k \|d^k\| \leq \lambda_k \|d_0^k\| + \rho_k \lambda_k \|d_1^k\| \to 0, k \to \infty.$$

从而

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = \lim_{k\to\infty} \|\lambda_k d^k + \lambda_k^2 \widetilde{d}^k\| \leq \lim_{k\to\infty} 2\lambda_k \|d^k\| = 0.$$

由假设 H4(ii) 知 x^* 是问题 (P) 的孤立 KKT 点，因此由定理 4.3.1 知：x^* 是 $\{x^k\}$ 的孤立聚点，这和

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$$

一起表明 $\lim_{k\to\infty} x^k = x^*$.

(ii) 考虑到 $x^k \to x^*$, $k \to \infty$, 类似于定理 4.3.1 的证明知：

$$\lim_{k\to\infty} d_0^k = 0,\ \lim_{k\to\infty} d_1^k = 0,\ \lim_{k\to\infty} d^k = 0,\ \lim_{k\to\infty} \widetilde{d}^k = 0.$$

引理 4.4.2 *在上述假设下，当 k 充分大时，矩阵*

$$M_k = \begin{bmatrix} H_k & \nabla g_{A^k}(x^k) \\ \nabla g_{A^k}(x^k)^T & 0 \end{bmatrix}.$$

是非奇异的. 而且，存在常数 $C > 0$ 使得 $\|M_k^{-1}\| \le C$.

类似于文献 [105] 中引理 3.1 的证明，可得上述引理.

引理 4.4.3 *如果假设 H1,H2,H3 成立，则*

$$\|d_1^k\| = o(\|d_0^k\|^2),\quad \|d^k\| = \|d_0^k\| + o(\|d_0^k\|^2), \tag{4.14}$$

$$J_k = I(x^*) = A^k,\quad \|\widetilde{d}^k\| = O(\|d^k\|^2), \tag{4.15}$$

$$\lim_{k\to\infty} u^k = u^*. \tag{4.16}$$

证明：首先，由 (4.4), (4.5) 和 (4.6) 知：

$$\|d_1^k\| = o(\|d_0^k\|^2),\quad \|d^k\| = \|d_0^k\| + o(\|d_0^k\|^2).$$

其次，我们将证明 (4.15). 由 $\lim_{k\to\infty}(x^k, d^k) = (0,0)$ 知：$J_k \subseteq I(x^*)$. 对任意 $i \in I(x^*)$, 由 H4(ii) 知：$u_i^* > 0$. 从而由定理 4.3.1 知：$u_i^k > 0$. 即 $i \in J_k$. 由文献 [86] 中的定理 2.3 和定理 3.7 结合假设 H1,H4(ii) 知 $I(x^*) = A^k$. 注意到 (4.7) 等价于如下线性方程组：

$$\begin{bmatrix} H_k & \nabla g_{A^k}(x^k) \\ \nabla g_{A^k}(x^k)^T & 0 \end{bmatrix} \begin{bmatrix} \widetilde{d}^k \\ \widetilde{\lambda}^k \end{bmatrix} = \begin{bmatrix} 0 \\ -\|d^k\|^\tau e_{A^k} - g_{A^k}(x^k + d^k) \end{bmatrix}.$$

因此，不难证明 $\|\widetilde{d}^k\| = O(\|d^k\|^2)$.

类似于文献 [102] 中的证明，有 $\lim_{k\to\infty} u^k = u^*$.

为确保当 k 充分大时，$\lambda_k \equiv 1$, 再做如下假设.

H5 假设 $\|P_k(\nabla_{xx}^2 L(x^k, u_k) - H_k)d_0^k\| = o(\|d_0^k\|)$, 其中

$$L(x, u_k) = f(x) + \sum_{j\in J_k} u_j^k g_j(x),$$

$$P_k = I_n - R_k(R_k^T R_k)^{-1} R_k^T,\ R_k = R(x^k) = \left(\nabla g_j(x^k), j \in J_k\right).$$

引理 4.4.4 *如果假设 H1,H2,H3,H4,H5 成立，则当 k 充分大时其步长恒为 1. 即，$\lambda_k \equiv 1$.*

证明：该定理的证明只需证明当 $\lambda = 1$ 时 (4.8) 和 (4.9) 成立即可.

首先证明当 $\lambda = 1$ 时 (4.9) 成立. 当 $j \notin I(x^*)$ 时，即，$g_j(x^*) < 0$, 注意到 $(x^k, d^k, \widetilde{d}^k) \to (x^*, 0, 0)(k \to \infty)$, 故 $g_j(x^k + d^k + \widetilde{d}^k) \le 0$ 成立.

当 $j \in I(x^*) = A^k$ 时，由泰勒展式和 (4.15) 有

$$\begin{aligned} g_j(x^k + d^k + \widetilde{d}^k) &= g_j(x^k + d^k) + \nabla g_j(x^k + d^k)^T \widetilde{d}^k + O(\|\widetilde{d}^k\|^2) \\ &= g_j(x^k + d^k) + \nabla g_j(x^k)^T \widetilde{d}^k + O(\|d^k\|\|\widetilde{d}^k\|) + O(\|\widetilde{d}^k\|^2) \\ &= g_j(x^k + d^k) + \nabla g_j(x^k)^T \widetilde{d}^k + O(\|d^k\|^3). \end{aligned} \tag{4.17}$$

因此结合 (4.7) 和 (4.17) 知

$$g_j(x^k + d^k + \widetilde{d}^k) = -\|d^k\|^\tau + O(\|d^k\|^3) < 0.$$

这表明 (4.9) 当 $\lambda = 1$ 时成立.

下面证明 (4.8) 当 $\lambda = 1$ 时成立.

考虑到泰勒展式和 (4.15) 可得

$$\begin{aligned} \omega_k &\overset{def}{=} f(x^k + d^k + \widetilde{d}^k) - f(x^k) - \alpha \nabla f(x^k)^T d^k \\ &= \nabla f(x^k)^T (d^k + \widetilde{d}^k) + \tfrac{1}{2}(d^k)^T \nabla^2_{xx} f(x^k) d^k - \alpha \nabla f(x^k)^T d^k + o(\|d^k\|^2). \end{aligned} \tag{4.18}$$

另一方面，由 (4.3) 的 KKT 条件和 (4.14), 可得

$$\nabla f(x^k)^T (d^k + \widetilde{d}^k) = -(d_0^k)^T H_k d_0^k - \sum_{j \in J_k} u_j^k \nabla g_j(x^k)^T (d^k + \widetilde{d}^k) + o(\|d_0^k\|^2). \tag{4.19}$$

$$\begin{aligned} \nabla f(x^k)^T d^k &= -(d_0^k)^T H_k d_0^k - \sum_{j \in J_k} u_j^k \nabla g_j(x^k)^T d_0^k + o(\|d_0^k\|^2) \\ &= -(d_0^k)^T H_k d_0^k + \sum_{j \in J_k} u_j^k g_j(x^k) + o(\|d_0^k\|^2). \end{aligned}$$

又从 (4.17) 的第三式，(4.15) 和泰勒展式知

$$g_j(x^k) + \nabla g_j(x^k)^T (d^k + \widetilde{d}^k) + \frac{1}{2}(d^k)^T \nabla^2_{xx} g_j(x^k) d^k = o(\|d^k\|^2), \quad j \in J_k.$$

因此

$$-\sum_{j \in J_k} u_j^k \nabla g_j(x^k)^T (d^k + \widetilde{d}^k) = \sum_{j \in J_k} u_j^k g_j(x^k) + \tfrac{1}{2}(d_0^k)^T \Big(\sum_{j \in J_k} u_j^k \nabla^2_{xx} g_j(x^k)\Big) d_0^k + o(\|d_0^k\|^2). \tag{4.20}$$

把 (4.20) 代入 (4.19) 得

$$\nabla f(x^k)^T (d^k + \widetilde{d}^k) = -(d_0^k)^T H_k d_0^k + \sum_{j \in J_k} u_j^k g_j(x^k) + \tfrac{1}{2}(d_0^k)^T \Big(\sum_{j \in J_k} u_j^k \nabla^2_{xx} g_j(x^k)\Big) d_0^k + o(\|d_0^k\|^2). \tag{4.21}$$

把 (4.21) 和 (4.19) 的第三式代入 (4.18) 得

$$\begin{aligned}\omega_k &= (\alpha-1)(d_0^k)^T H_k d_0^k + \tfrac{1}{2}(d_0^k)^T \nabla_{xx}^2 L(x^k,u^k) d_0^k \\ &+ (1-\alpha)\sum_{j\in J_k} u_j^k g_j(x^k) + o(\|d_0^k\|^2) \\ &\le ((\alpha-1)+\tfrac{1}{2})a\|d_0^k\|^2 + \tfrac{1}{2}(d_0^k)^T(\nabla_{xx}^2 L(x^k,u^k) - H_k)d_0^k \\ &+ (1-\alpha)\sum_{j\in J_k} u_j^k g_j(x^k) + o(\|d_0^k\|^2).\end{aligned}$$

由 P_k 的定义有:

$$\begin{aligned}d^k &= P_k d^k + R_k(R_k^T R_k)^{-1} R_k^T d^k \\ &= P_k d^k + R_k(R_k^T R_k)^{-1}(-g_{I(x^*)}(x^k) + \rho_k\|d_0^k\|^\delta) \\ &= P_k d^k + O(\|-g_{I(x^*)}(x^k)\|) + o(\|d_0^k\|^2).\end{aligned}$$

根据假设 H5 和上述条件知

$$\omega_k \le ((\alpha-1)+\frac{1}{2})a\|d_0^k\|^2 + o(\|d_0^k\|^2).$$

因此, 由 $\alpha\in(0,\frac{1}{2})$ 可得结论成立.

定理 4.4.1 *在上述假设下, 4.2 节中算法是超线性收敛的. 即,* $\|x^{k+1}-x^*\| = o(\|x^k-x^*\|)$.

证明: 因为 $P_k R_k = 0$, 故由泰勒展式 d^k 的定义知:

$$\begin{aligned}P_k H_k d^k &= -P_k\nabla f(x^k) = -P_k(\nabla f(x^k) + R_k u_{I(x^*)}^*) \\ &= -P_k\nabla_{xx}^2 L(x^*, u_{I(x^*)}^*)(x^k-x^*) + O(\|x^k-x^*\|^2),\end{aligned}$$

即

$$P_k(H_k - \nabla_{xx}^2 L(x^*, u_{I(x^*)}^*))d^k = -P_k\nabla_{xx}^2 L(x^*, u_{I(x^*)}^*)(x^k+d^k-x^*) + O(\|x^k-x^*\|^2).$$

而从泰勒展式可得:

$$g_{I(x^*}(x^k) = g_{I(x^*}(x^k) - g_{I(x^*}(x^*) = R_k^T(x^k-x^*) + O(\|x^k-x^*\|^2).$$

由 (4.11) 知

$$R_k^T(x^k+d^k-x^*) = o(\|d^k\|) + O(\|x^k-x^*\|^2).$$

记

$$D_k = \begin{bmatrix} P_k\nabla_{xx}^2 L(x^*, u_{I(x^*)}^*) \\ R_k^T \end{bmatrix},$$

由 H4(ii) 知： D_k 是列满秩的. 从而有

$$(x^k+d^k-x^*) = (D_k^TD_k)^{-1}D_k^T\begin{bmatrix}-P_k(H_k-\nabla_{xx}^2L(x^*,u_{I(x^*)}^*))d^k\\ 0\end{bmatrix}+o(\|d^k\|)+O(\|x^k-x^*\|^2).$$

由上式和假设 H5 知：

$$\|x^k+d^k-x^*\| = o(\|d^k\|)+O(\|x^k-x^*\|^2).$$

根据引理 4.4.3 得 $\|\widetilde{d}^k\| = o(\|d^k\|)$. 从而有

$$\begin{aligned}\|x^{k+1}-x^*\| &= \|x^k+d^k+\widetilde{d}^k-x^*\| \le o(\|d^k\|)+O(\|x^k-x^*\|^2)\\ &= o(\|d^k+\widetilde{d}^k\|)+O(\|x^k-x^*\|^2)\\ &= \frac{o(\|d^k+\widetilde{d}^k\|)}{\|d^k+\widetilde{d}^k\|}(\|(x^{k+1}-x^*)-(x^k-x^*)\|)+O(\|x^k-x^*\|^2)\\ &\le \frac{|o(\|d^k+\widetilde{d}^k\|)|}{\|d^k+\widetilde{d}^k\|}(\|x^{k+1}-x^*\|+\|x^k-x^*\|)+O(\|x^k-x^*\|^2)\\ &= o(\|x^{k+1}-x^*\|)+o(\|x^k-x^*\|).\end{aligned}$$

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

4.5 数值实验

本节，我们对 4.2 节中算法和文献 [102] 中的算法进行了数值比较. 4.2 节中算法用 Algo41 表示，文献 [102] 中的算法用 Algo42 表示. 两种算法都用 MATLAB 6.5 编程，都在 Windows XP 的个人计算机实现. 其中 (4.3),(4.4) 和 (4.7) 都使用优化工具箱求解.

整个实验中，我们使用如下 BFGS 公式：

$$H_{k+1} = H_k-\frac{H_ks^k(s^k)^TH_k}{(s^k)^TH_ks^k}+\frac{\widetilde{y}^k(\widetilde{y}^k)^T}{(s^k)^T\widetilde{y}^k}, \quad (k\ge 0),$$

其中

$$s^k = x^{k+1}-x^k,\ \widetilde{y}^k = y^k+a_k(\gamma_ks^k+R_kR_k^Ts^k),\ \gamma_k = \min\{\|d^k\|^2,\xi\},\ \ \xi\in(0,1),$$

$$y^k = \nabla_xL(x^{k+1},\lambda^k)-\nabla_xL(x^k,\lambda^k),$$

$$a_k = \begin{cases}0, & \text{if}\ \ (s^k)^Ty^k\ge\delta\|s^k\|^2, \delta\in(0,1);\\ 1, & \text{if}\ \ 0\le(s^k)^Ty^k<\delta\|s^k\|^2;\end{cases}$$

整个实验中，令

$$H_1 = I_n,\ \ \tau = 2.8,\ \ \rho = 2,\ \ \beta = 0.6,\ \ \alpha = 0.2,\ \ \gamma = 1,\ \ \delta = 3,\ \ \varepsilon^0 = 2.$$

表 4.1 的测试问题来自文献 [110] 和文献 [111]. 所选测试问题的初始点和文献 [110], 文献 [111] 中相同. 表 4.1 的列表示如下意义: **prob** 表示来自文献 [110] 和文献 [111] 的测试问题; **Ni** 表示求解的迭代数; **Nf, Ng** 分别表示目约束函数值和约束函数梯度值的迭代次数. **objective, dnorm** 和 **eps** 表示目标函数值, 模 d^k 和精度 ϵ.

表 4.1 数值结果

Algo	Prob	Ni	Nf	Ng	objective	dnorm	eps
Algo41	HS001	6	62	11	1.0000	5.2373e-014	1e-06
Algo42	-	8	64	14	1.0000	4.2709e-012	1e-06
Algo41	HS12	11	51	8	-30.0000	2.5691e-008	1e-06
Algo42	-	9	37	6	-30.0000	7.5775e-008	1e-06
Algo41	HS29	13	56	9	-22.6274	1.0308e-008	1e-06
Algo42	-	16	87	11	-22.6274	1.2937e-008	1e-06
Algo41	HS31	9	434	14	-6.0000	5.4155e-008	1e-06
Algo42	-	10	455	16	-6.0000	4.5850e-008	1e-06
Algo41	HS34	18	984	53	-0.8340	7.2411e-007	1e-06
Algo42	-	35	1160	73	-0.8340	3.1356e-008	1e-06
Algo41	HS35	8	164	14	0.1111	1.0122e-008	1e-06
Algo42	-	9	166	17	0.1111	3.6653e-008	1e-06
Algo41	HS43	12	165	25	-44.0000	8.4210e-008	1e-06
Algo42	-	21	444	36	-44.0000	1.2956e-007	1e-06
Algo41	HS100	41	512	66	680.6301	4.6149e-007	1e-06
Algo42	-	31	147	56	680.6301	3.0494e-009	1e-06
Algo41	S225	7	105	11	2.0000	1.2382e-009	1e-06
Algo42	-	8	125	14	2.0000	3.4452e-009	1e-06
Algo41	S264	20	435	27	-44.1134	2.8061e-007	1e-06
Algo42	-	17	246	22	-44.1134	6.5058e-008	1e-06

第 5 章　求解约束优化问题的一个非内点型可行点 QP-free 算法

5.1　引言

本章我们继续考虑求解如下不等式约束优化问题

$$(P)\ \ \min_{x\in\Re^n} f(x) \qquad \text{s.t. } g_i(x)\le 0,\ i\in I. \tag{5.1}$$

其中 $f:\Re^n\to\Re$, $g_i:\Re^n\to\Re^m$ 都是连续可微函数. 我们记 $F=\{x\in\Re^n: g_i(x)\le 0, i\in I\}$.

为了克服传统 SQP 方法某些缺陷, 如 QP 子问题可能不相容, 求解 QP 子问题的计算量较大等. Panier, Tits 和 Herskovits [103] 提出一个求解问题 (P) 的可行点 QP-free 算法. 在这个算法的每一个迭代步, 仅要求解两个不同的线性方程组和一个线性最小二乘问题. 在适当的条件下, 这个算法具有全局收敛性和局部两步超线性收敛速度. Gao,He 和 Wu[104] 也提出一个求解问题 (P) 的可行点 QP-free 算法, 不同于 Panier, Tits 和 Herskovits 的算法, 在不要求稳定点数目有限的条件下, 由他们的算法所产生的点列的聚点都是问题 (P) 的 KKT 点. 但是, 在收敛性的证明中, 他们必须假定乘子序列有界. Qi,Qi[105] 基于互补函数和 KKT 条件, 提出了一个求解问题 (P) 的可行点 QP-free 算法, 他们在无严格互补条件下证明了迭代矩阵的一致非奇异性和近似乘子序列的有界性. Yang,Li 和 Qi[106] 通过引进一个工作集的概念, 提出了一个新的求解问题 (P) 的可行点 QP-free 算法, 这个新算法仅考虑工作集内的约束, 这使得计算量大大减少. 在这个算法的每一个迭代步, 为得到搜索方向, 仅要求解四个系数相同的线性方程组. 在适当的条件下, 这个算法具有全局收敛性和局部一步超线性收敛速度, 甚至具有二次收敛速度. 但是, 对于上述几种可行点 QP-free 算法, 迭代点必须是 F 的内点.

在本章中, 以文献 [106] 中算法为基础, 我们提出一个求解问题 (P) 的非内点型可行点 QP-free 算法. 这个算法不要求迭代点必须是 F 的内点, 在算法的每一个迭代步, 为得到搜索方向, 只需求解四个系数相同的线性方程组. 在适当的条件下, 我们证明该算法具有全局收敛性和超线性收敛性.

5.2　算法描述

5.2.1 Yang,Li,Qi 算法

本小节我们给出 Yang,Li,Qi 算法 [106]. 首先做如下两条假设:

H1 集合 F 是有界的.

H2 对任意 $x \in F$, 向量组 $\{\nabla g_i(x),\ i \in I(x)\}$ 是线性无关的，这里 $I(x) = \{i|g_i(x) = 0\}$.

下面的性质可见文献 [108].

性质 5.2.1 (i) 对每个 $x \in F$, 关于 λ 的二次函数 $\|\nabla_x L(x,\lambda)\|^2 + \|G(x)\lambda\|^2$ 存在唯一的极小解 $\lambda(x) = -M^{-1}\nabla g(x)^T \nabla f(x)$, 其中

$$L(x,\lambda) = f(x) + \lambda^T g(x), G(x) = diag(g_i(x)), M(x) = \nabla g(x)^T \nabla g(x) + G^2(x).$$

(ii) 乘子函数 $\lambda(x)$ 在 F 上是连续可微的.

(iii) 如果 $(x^*, \lambda^*) \in R^n \times R^m$ 是 (P) 的 KKT 点对，那么 $\lambda(x^*) = \lambda^*$.

对 $x \in F$, 一个近似积极集定义如下:

$$A(x;\varepsilon) = \{i|g_i(x) + \varepsilon\rho(x,\lambda(x)) \geq 0\},$$

其中 $\rho(x,\lambda(x)) = \sqrt{\|\Phi(x,\lambda)\|}$, $\Phi(x,\lambda) = \begin{bmatrix} \nabla_x L(x,\lambda) \\ \min\{-g(x),\lambda\} \end{bmatrix}$.

对一个对称正定矩阵 H 和 I 的一个子集 A, 定义

$$V(x,H,A) = \begin{bmatrix} H & \nabla g_A(x) \\ \nabla g_A(x)^T & 0 \end{bmatrix}.$$

Yang,Li,Qi 算法步骤:

参数及有关初始值: $\beta \in (0,1)$, $\mu \in (0,0.5)$, $\nu > 2$, $\tau \in (2,3)$, $\vartheta \in (0,1)$, $M^0 > 0$, $\sigma \in (0,1)$, $\sigma_1 > 1$, x^1 是 F 的一个严格可行点， H_1 是一个对称正定矩阵, $\varepsilon^0 > 0$, $k = 1$.

步 1. 令 $\varepsilon = \varepsilon^{k-1}$ 和 $M = M^{k-1}$.

步 2. 令 $A^k(\varepsilon) = A(x^k;\varepsilon)$ 和 $V_k(\varepsilon) = V(x^k, H_k; A^k(\varepsilon))$. 如果 $\nabla g_{A^k(\varepsilon)}(x^k)$ 非满秩或 $\|V_k(\varepsilon)^{-1}\| > M$, 那么令 $\varepsilon = \sigma\varepsilon$ 和 $M = \sigma_1 M$, 进入步 2.

步 3. 令 $\varepsilon^k = \varepsilon$, $M^k = M$, $A^k = A^k(\varepsilon^k)$, $V_k = V(x^k, H_k; A^k)$.

步 4. 搜索方向的计算:

(i) 求解如下线性方程组得 $(d^{k0}, z^{k0}_{A^k})$:

$$V_k \begin{bmatrix} d \\ z_{A^k} \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) \\ 0 \end{bmatrix}.$$

(ii) 求解如下线性方程组得 $(d^{k1}, z^{k1}_{A^k})$:

$$V_k \begin{bmatrix} d \\ z_{A^k} \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) \\ \varphi^k \end{bmatrix},$$

其中

$$\varphi_i^k = \begin{cases} z_i^{k0}, & \text{if } \ z_i^{k0} < 0; \\ -g_i(x^k), & \text{if } \ z_i^{k0} > 0; \\ 0, & \text{if } \ z_i^{k0} = 0. \end{cases}$$

如果 $d^{k1} = 0$, 停.

(iii) 求解如下线性方程组得 $(d^{k2}, z_{A^k}^{k2})$:

$$V_k \begin{bmatrix} d \\ z_{A^k} \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) \\ \varphi^k - \|d^{k1}\|^{\nu} e_{A^k} \end{bmatrix}.$$

(iv) 通过以下方式计算 $(d^k, z_{A^k}^k)$:

$$\begin{bmatrix} d^k \\ z_{A^k}^k \end{bmatrix} = (1-\phi^k) \begin{bmatrix} d^{k1} \\ z_{A^k}^{k1} \end{bmatrix} + \phi^k \begin{bmatrix} d^{k2} \\ z_{A^k}^{k2} \end{bmatrix},$$

其中 $\phi^k = (\vartheta - 1) \dfrac{\nabla f(x^k)^T d^{k1}}{1 + \|d^{k1}\|^{\nu} |\sum\limits_{i \in A^k} z_i^{k0}|}$.

步 5. 求解如下线性方程组得 $\widehat{d}^k$:

$$V_k \begin{bmatrix} d \\ z_{A^k} \end{bmatrix} = \begin{bmatrix} 0 \\ -\|d^k\|^{\tau} e_{A^k} - g_{A^k}(x^k + d^k) \end{bmatrix}.$$

如果 $\|\widehat{d}^k\| > \|d^k\|$, 令 $\widehat{d}^k = 0$.

步 6. 线搜索:

求序列 $\{1, \beta, \beta^2, \cdots,\}$ 中满足如下不等式组的第一个数作为 t_k:

$$f(x^k + td^k + t^2\widehat{d}^k) \le f(x^k) + \mu t \nabla f(x^k)^T d^k, \quad g_i(x^k + td^k + t^2\widehat{d}^k) < 0, \quad i \in I. \tag{5.2}$$

步 7. 令 $x^{k+1} = x^k + t_k d^k + t_k^2 \widehat{d}^k$, 产生一个新的对称正定矩阵 H_{k+1}, 令 $k = k+1$, 进入步 1.

注 5.2.1 从 $H2$ 和 H_k 的对称正定性易知, $V(x^k, H_k, A^k)$ 是非奇异的.

5.2.2 非内点型可行点 QP-free 算法

从 Yang,Li,Qi 算法可知, 该算法要求迭代点必须是 F 的内点, 在本章中, 我们仅通过修改

$$\varphi_i^k = \begin{cases} z_i^{k0}, & \text{if } \ z_i^{k0} < 0; \\ -g_i(x^k), & \text{if } \ z_i^{k0} \ge 0. \end{cases}$$

和线搜索第二个不等式组为

$$g_i(x^k+td^k+t^2\widehat{d}^k)\le 0,\quad i\in I.$$

得到一个非内点型可行 QP-free 算法 (其他步骤同 Yang,Li,Qi 算法).
类似于 Yang,Li,Qi 算法，不难得到如下关系：

$$\begin{aligned}
&d^{k0}=-Q_k\nabla f(x^k),\ z_{A^k}^{k0}=-B_k^T\nabla f(x^k);\\
&d^{k1}=d^{k0}+B_k\varphi^k,\ z_{A^k}^{k1}=z_{A^k}^{k0}-D_k^{-1}\varphi^k;\\
&d^{k2}=d^{k1}-\|d^{k1}\|^{\nu}B_ke_{A^k},\ z_{A^k}^{k2}=z_{A^k}^{k1}+\|d^{k1}\|^{\nu}D_k^{-1}e_{A^k};\\
&d^{k}=d^{k1}-\phi^k\|d^{k1}\|^{\nu}B_ke_{A^k},\ z_{A^k}^{k}=z_{A^k}^{k1}+\phi^k\|d^{k1}\|^{\nu}D_k^{-1}e_{A^k},
\end{aligned}\tag{5.3}$$

其中

$$N_{A^k}=\nabla g_{A^k}(x^k),\ D_k=N_{A^k}^TH_k^{-1}N_{A^k},$$

$$B_k=H_k^{-1}N_{A^k}D_k^{-1},\ Q_k=H_k^{-1}(I_n-N_{A^k}B_k^T).$$

为讨论方便，对任意的 k, 我们令 $z_i^{k0}=z_i^{k1}=z_i^{k2}=z_i^k=0,\ \ \forall i\notin A^k$.

引理 5.2.1 *(i) 如果 $d^{k1}=0$, 那么 x^k 是问题 (P) 的 KKT 点.*
(ii) 如果 $d^{k1}\neq 0$, 那么

$$\nabla f(x^k)^Td^{k0}=-(d^{k0})^TH_kd^{k0},\ \nabla f(x^k)^Td^{k1}=\nabla f(x^k)^Td^{k0}-(\varphi^k)^Tz_{A^k}^{k0}\le\nabla f(x^k)^Td^{k0},$$

$$\nabla f(x^k)^Td^k\le\vartheta\nabla f(x^k)^Td^{k1},\ \nabla g_i(x^k)^Td^k=\varphi^k-\phi^k\|d^{k1}\|^{\nu}<0,\ \forall i\in I(x^k).$$

证明：(i) 如果 $d^{k1}=0$, 那么由步 4(ii) 知：

$$\nabla f(x^k)+\nabla g_{A^k}(x^k)z_{A^k}^{k1}=0,$$

$$\varphi^k=0.$$

由 φ^k 的定义得： $z_{A^k}^{k0}\ge 0, g_{A^k}(x^k)=0$. 而从 (5.3) 得， $z_{A^k}^{k1}\ge 0$. 根据 (P) 的 KKT 点的定义知结论成立.
(ii) 该结论的证明类似于文献 [106] 中引理 2.2 的证明.

5.3 全局收敛性分析

本节我们将分析 5.2.2 节中算法的全局收敛性. 首先我们做如下假设：
H3 对所有的 k 和 $d\in R^n$, 存在正常数 C_1 和 C_2 满足

$$C_1\|d\|^2\le d^TH_kd\le C_2\|d\|^2.$$

下面两个引理见文献 [106].

引理 5.3.1 序列 $\{(d^{k0},z^{k0})\}$, $\{(d^{k1},z^{k1})\}$, $\{(d^{k2},z^{k2})\}$ 都是有界的.

引理 5.3.2 对所有的 k, 存在正常数 κ 满足 $\|d^k-d^{k1}\|\le\kappa\|d^{k1}\|^{\nu}$.

引理 5.3.3 设 x^* 是 5.2.2 节中算法所产生序列 $\{x^k\}$ 的一个聚点且 $\{x^k\}_{K_0}\to x^*$. 如果 $\{\nabla f(x^k)^T d^{k1}\}_{K_0}\to 0$, 那么 x^* 是问题 (P) 的 KKT 点且 $\{z^{k0}\}_{K_0}$ 存在一子列收敛到对应于 x^* 的唯一乘子向量 λ^*.

证明: 如果 $\{\nabla f(x^k)^T d^{k1}\}_{K_0}\to 0$, 由引理 5.2.1(ii) 和假设 H3 知:

$$\{d_0^k\}_{K_0}\to 0,\ \{\varphi^{kT}z_{A^k}^{k0}\}_{K_0}\to 0.$$

由引理 5.3.1 知: 存在子集 $K_1\subseteq K_0$, 使得 $\{z^{k0}\}_{K_1}\to z^*$(z^* 是 $\{z^{k0}\}_{K_0}$ 的一个聚点). 根据步 4(i) 和上述关系得:

$$\nabla f(x^*)+\nabla g(x^*)z^*=0,$$

$$z_i^*\ge 0,\ z_i^*g_i(x^*)=0,\ i\in I.$$

又 $g(x^*)\le 0$, 因此 x^* 是问题 (P) 的 KKT 点. 由乘子向量的唯一性知: $\{z^{k0}\}_{K_1}$ 收敛到对应于 x^* 的唯一乘子向量 λ^*.

类似于文献 [106] 中定理 3.1 的证明，我们可得如下全局收敛性定理.

定理 5.3.1 如果 (x^*,λ^*) 是该算法所产生序列 $\{(x^k,z^{k0})\}$ 的一个聚点, 那么 (x^*,λ^*) 是问题 (P) 的 KKT 点.

证明: 假设 (x^*,λ^*) 不是问题 (P) 的 KKT 点, 由定理的条件知: 存在一个子集 K 使得: $(x^k,z^{k0})_K\to(x^*,\lambda^*)$. 下证存在子集 $K_0\subseteq K$ 使得: $\{\nabla f(x^k)^T d^{k1}\}_{K_0}\to 0$. 否则, 存在 $\gamma>0$ 和 $\underline{d}>0$ 使得:

$$\nabla f(x^k)^T d^{k1}\le-\gamma,\ k\in K,\ \|d^{k1}\|\ge\underline{d}. \tag{5.4}$$

由引理 5.3.1, ϕ^k 的定义和上式知: 存在 $\widetilde{\phi}>0$, 使得 $\phi^k\ge\widetilde{\phi}$. 类似于文献 [103] 中的性质 3.9 和文献 [106] 中的定理 3.1 的证明有:

$$\begin{aligned}&f(x^k+td^k+t^2\widehat{d}^k)-f(x^k)-\mu t\nabla f(x^k)^T d^k\\ &\le t\{\sup_{\xi\in[0,1]}\|\nabla f(x^k+t\xi d^k+t^2\xi\widehat{d}^k)-\nabla f(x^k)\|\|d^k\|\\ &+2t\sup_{\xi\in[0,1]}\|\nabla f(x^k+t\xi d^k+t^2\xi\widehat{d}^k)\|\|\widehat{d}^k\|-(1-\mu)\vartheta C_1\underline{d}^2\},\end{aligned} \tag{5.5}$$

$$g_i(x^k+td^k+t^2\widehat{d}^k)\le g_i(x^k)+t\{u_i^k(t)+\nabla g_i(x^k)^T d^k\}, \tag{5.6}$$

其中

$$u_i^k(t)=\sup_{\xi\in[0,1]}\|\nabla g_i(x^k+t\xi d^k+t^2\xi\widehat{d}^k)-\nabla g_i(x^k)\|\|d^k\|+2t\sup_{\xi\in[0,1]}\|\nabla g_i(x^k+t\xi d^k+t^2\xi\widehat{d}^k)\|\|\widehat{d}^k\|.$$

由 (5.4), φ^k 的定义和上式知：对 $i\in A^k$, 有

$$\begin{aligned}
&g_i(x^k+td^k+t^2\widehat{d}^k)\\
&\le g_i(x^k)+t\{u_i^k t+\varphi_i^k-\phi^k\|d^{k1}\|^\nu\}\\
&\le g_i(x^k)+t\{u_i^k t+\varphi_i^k-\phi^k\|d^{k1}\|^\nu\}\\
&=\begin{cases} g_i(x^k)+tz_i^{k0}+t\{u_i^k t-\phi^k\|d^{k1}\|^\nu\}, & \text{if } z_i^{k0}<0;\\ (1-t)g_i(x^k)+t\{u_i^k t-\phi^k\|d^{k1}\|^\nu\}, & \text{if } z_i^{k0}\ge 0.\end{cases}
\end{aligned}$$

当 $i\notin A^k$ 时， $g_i(x^k)<-\bar{\varepsilon}\rho(x^k,\lambda(x^k))$. 因此由 (5.6) 得

$$g_i(x^k+td^k+t^2\widehat{d}^k)\le-\bar{\varepsilon}\rho(x^k,\lambda(x^k))+t\{u_i^k(t)+\nabla g_i(x^k)^Td^k\}.$$

而由 $\rho(x^*,\lambda(x^*))>0,\{x^k\}_K\to x^*,\|\widehat{d}^k\|\le\|d^k\|$ 和 $\{d^k\}$ 的有界性可得：存在 $t_i>0$, 使得对任意 $t\in[0,t_i]$ 和 $k\in K$ 充分大有

$$g_i(x^k+td^k+t^2\widehat{d}^k)\le 0.$$

由 (5.5) 知：存在 $t_f>0$, 使得对任意 $t\in[0,t_f]$ 和 $k\in K$ 充分大有

$$f(x^k+td^k+t^2\widehat{d}^k)-f(x^k)-\mu t\nabla f(x^k)^Td^k\le 0.$$

令 $\bar{t}=\min\{t_f,t_1,\cdots,t_m\}>0$, 则由引理 5.2.2 可得

$$f(x^k+td^k+t^2\widehat{d}^k)-f(x^k)\le-\mu\bar{t}\vartheta\gamma. \tag{5.7}$$

而由 $\{f(x^k)\}$ 的单调下降性和 f 的连续性并结合 $x^k\to x^*(k\in K)$ 可知：

$$f(x^k)\to f(x^*),\ k\to\infty.$$

显然这与 (5.7) 矛盾！因此由引理 5.3.3 知结论成立.

5.4 收敛速度分析

本节我们将分析 5.2.2 节中算法的收敛速度. 假定 (x^*,λ^*) 是该算法所产生序列 (x^k,z^{k0}) 的一个聚点，那么从全局收敛性定理知，(x^*,λ^*) 是问题 (P) 的 KKT 点. 下面我们做如下假设：

H4(i) 函数 $f(x),g_i(x)$ 都是二次连续可微的.

(ii) 在 (x^*,λ^*) 处严格互补条件成立，即 $\lambda^*-g(x^*)>0$.
H5 在 (x^*,λ^*) 处二阶充分条件条件成立，即在子空间 $\{u|\nabla g_i(x^*)^Tu=0,\forall i\in I(x^*)\}$ 上 $\nabla^2_{xx}L(x^*,\lambda^*)$ 是正定的.

引理 5.4.1 *若 $H1,H3,H4,H5$ 成立，则整个序列 $\{x^k\}$ 收敛到 x^*，即，$x^k\to x^*,k\to\infty$.*

证明: 从 (5.2) 和引理 5.2.1 可知:

$$f(x^{k+1})\le f(x^k)+\mu t_k\nabla f(x^k)^Td^k\le f(x^k)-\mu\vartheta t_k(d^{k1})^TH_kd^{k1}.$$

考虑到 $H1$ 和 $H3$, 可得到 $t_k\|d_{k1}\|\to 0$. 因此利用引理 5.3.2 可得，$t_k\|d_k\|\to 0$. 从而 $\|x^{k+1}-x^k\|\to 0$. 又 $H4,H5$ 成立，类似于文献 [88] 中的性质 4.1, 可得 $\lim\limits_{k\to\infty}x^k=x^*$.

下面的引理见文献 [86] 中的定理 2.3 和 2.7.

引理 5.4.2 *设 $H5$ 成立，如果 x^* 是问题 (P) 的 KKT 点，那么存在 x^* 的一个邻域，使得这个邻域中的任意 x, 有 $A(x,\bar{\varepsilon})=I(x^*)$.*

类似于文献 [106] 中的推论 4.1, 我们可得如下引理.

引理 5.4.3 *若 $H1,H2,H3,H5$ 成立，则对充分大的 k, 有 $A_k=I(x^*)$ 成立. 而且有如下关系成立:*
(i) $d^{k0}\to 0,\tilde{d}^{k1}\to 0,d^k\to 0(k\to\infty)$
(ii) $z^{k0}\to\lambda^*,z^{k1}\to\lambda^*,z^k\to\lambda^*(k\to\infty)$
(iii) 如果还有 $H4$ 成立，则对充分大的 k, 有 $\varphi_k=-g_{I(x^*)}(x^k)$.

为得到该算法的超线性收敛性，我们再做如下假设.
H6 矩阵序列 $\{H_k\}$ 满足

$$\|P_k(H_k-\nabla^2_{xx}L(x^*,\lambda^*))d^k\|=o(\|d^k\|),$$

其中 $P_k=I_n-N_k(N_k^TN_k)^{-1}N_k^T$, $N_k=\nabla g_{I(x^*)}(x^k)$.

引理 5.4.4 *(i) 当 k 充分大时，d^k 可分解为 $d^k=P_kd^k+\widetilde{d}^k$, 其中 $\widetilde{d}^k=O(\|g_{I(x^*)}(x^k)\|)+o(\|d^{k1}\|^2)$, 特别，当 $g_{I(x^*)}(x^k)=0$ 时，$\widetilde{d}^k=o(\|d^{k1}\|^2)$. (ii) 当 k 充分大时，$\|\widehat{d}^k\|=O(\|d^k\|^2)$.*

该引理的证明见文献 [106] 中引理 4.3, 引理 4.4.

引理 5.4.5 *若 $H1-H5$ 成立，则当 k 充分大时，步长 $t_k\equiv 1$.*

证明: 由线搜索可知只需证明: 当 k 充分大时,

$$f(x^k+d^k+\widehat{d}^k)\le f(x^k)+\mu\nabla f(x^k)^Td^k,$$

$$g_i(x^k+d^k+\widehat{d}^k)\le 0,\ i\in I.$$

当 $i\in I(x^*)$ 时, 由步 5 和引理 5.4.4 知:

$$\begin{aligned}g_i(x^k+d^k+\widehat{d}^k) &= g_i(x^k+d^k)+\nabla g_i(x^k+d^k)^T\widehat{d}^k+O(\|\widehat{d}^k\|^2)\\ &= g_i(x^k+d^k)+\nabla g_i(x^k)^T\widehat{d}^k+O(\|d^k\|\|\widehat{d}^k\|)+O(\|\widehat{d}^k\|^2)\\ &= -\|d^k\|^\tau+O(\|d^k\|^3).\end{aligned}\tag{5.8}$$

故当 k 充分大时, $g_i(x^k+d^k+\widehat{d}^k)<0,\ i\in I(x^*)$. 当 $i\notin I(x^*)$ 时, 由 $g_i(x)$ 的连续性和 $g_i(x^*)<0$ 易知结论成立.

下证 $f(x^k+d^k+\widehat{d}^k)\le f(x^k)+\alpha\nabla f(x^k)^Td^k$ 对充分大的 k 成立.
由泰勒展式和引理 5.4.4 有:

$$f(x^k+d^k+\widehat{d}^k)=f(x^k)+\nabla f(x^k)^T(d^k+\widehat{d}^k)+\frac{1}{2}(d^k)^T\nabla^2_{xx}f(x^k)d^k+o(\|d^k\|^2).$$

从步 4, 步 5 和 d^k 的定义得:

$$H_kd^k+\nabla f(x^k)+\sum_{i\in I(x^*)}z_i^k\nabla g_i(x^k)=0,$$

$$\nabla g_i(x^k)^Td^k=-g_i(x^k)-\phi^k\|d^{k1}\|^\nu,\ i\in I(x^*),\tag{5.9}$$

$$g_i(x^k+d^k)+\nabla g_i(x^k)^T\widehat{d}^k=o(\|d^k\|^2).\tag{5.10}$$

由上面第一式和引理 5.4.4 可得:

$$\nabla f(x^k)^Td^k=-d^{kT}H_kd^k-\sum_{i\in I(x^*)}z_i^k\nabla g_i(x^k)^Td^k,$$

$$\nabla f(x^k)^T\widehat{d}^k=-\sum_{i\in I_0}z_i^k\nabla g_i(x^k)^T\widehat{d}^k+o(\|d^k\|^2).$$

因此

$$\begin{aligned}f(x^k+d^k+\widehat{d}^k) &= f(x^k)+\tfrac{1}{2}\nabla f(x^k)^Td^k+\tfrac{1}{2}(d^k)^T(\nabla^2_{xx}f(x^k)-H_k)d^k\\ &- \tfrac{1}{2}\sum_{i\in I(x^*)}z_i^k\nabla g_i(x^k)^Td^k-\sum_{i\in I(x^*)}z_i^k\nabla g_i(x^k)^T\widehat{d}^k+o(\|d^k\|^2).\end{aligned}$$

由泰勒展式和 (5.10) 得:

$$g_i(x^k)+\nabla g_i(x^k)^T(d^k+\widehat{d}^k)+\frac{1}{2}(d^k)^T\nabla^2_{xx}g_i(x^k)d^k=o(\|d^k\|^2).$$

从 (5.9) 和上式知：

$$
\begin{aligned}
&-\frac{1}{2}\sum_{i\in I(x^*)} z_i^k\nabla g_i(x^k)^T d^k-\sum_{i\in I(x^*)} z_i^k\nabla g_i(x^k)^T\widehat{d}^k\\
&=\frac{1}{2}\sum_{i\in I(x^*)} z_i^k\nabla g_i(x^k)^T d^k-\sum_{i\in I(x^*)} z_i^k\nabla g_i(x^k)^T(d^k+\widehat{d}^k)\\
&=\frac{1}{2}\sum_{i\in I(x^*)} z_i^k g_i(x^k)-\frac{1}{2}\sum_{i\in I(x^*)}\phi^k z_i^k\|d^k\|^{\nu}\\
&+\frac{1}{2}\sum_{i\in I(x^*)} z_i^k d^{kT}\nabla_{xx}^2 g_i(x^k)d^k+o(\|d^k\|^2)\\
&=\frac{1}{2}\sum_{i\in I(x^*)} z_i^k g_i(x^k)+\frac{1}{2}\sum_{i\in I(x^*)} z_i^k d^{kT}\nabla_{xx}^2 g_i(x^k)d^k+o(\|d^k\|^2).
\end{aligned}
\tag{5.11}
$$

由假设 $H4$ 和引理 5.3.2 知：当 $g_{I(x^*)}(x^k)\neq 0$ 时，

$$
\frac{1}{2}\sum_{i\in I(x^*)} z_i^k g_i(x^k)+o(\|g_{I(x^*)}(x^k)\|<0.
$$

结合上两式和假设 $H6$, 由 (5.8) 可得：

$$
\begin{aligned}
f(x^k+d^k+\widehat{d}^k) &= f(x^k)+\tfrac{1}{2}\nabla f(x^k)^T d^k+\tfrac{1}{2}\sum_{i\in I(x^*)} z_i^k g_i(x^k)\\
&+\frac{1}{2}d^{kT}(\nabla_{xx}^2 f(x^k)+\nabla_{xx}^2 g_i(x^k)-H_k)d^k+o(\|d^k\|^2)\\
&= f(x^k)+\tfrac{1}{2}\nabla f(x^k)^T d^k+\tfrac{1}{2}\sum_{i\in I(x^*)} z_i^k g_i(x^k)+o(\|g_{I(x^*)}(x^k)\|)\\
&+\frac{1}{2}d^{kT}P_k(\nabla_{xx}^2 f(x^k)+\nabla_{xx}^2 g_i(x^k)-H_k)d^k+o(\|d^k\|^2)\\
&\le f(x^k)+\tfrac{1}{2}\nabla f(x^k)^T d^k\\
&+\frac{1}{2}\|d^k\|\|P_k(\nabla_{xx}^2 L(x^k,\lambda_k)-H_k)d^k\|+o(\|d^k\|^2)\\
&= f(x^k)+\tfrac{1}{2}\nabla f(x^k)^T d^k+o(\|d^k\|^2).
\end{aligned}
$$

由引理 5.3.2 和假设 $H3$ 知：

$$
\begin{aligned}
\nabla f(x^k)^T d^k &\le \theta\nabla f(x^k)^T d^{k1}\\
&\le -\theta(d^{k1})^T H_k d^{k1}\\
&\le -\theta C_1\|d^{k1}\|^2\\
&= -\theta C_1\|d^k\|^2+o(\|d^k\|^2).
\end{aligned}
$$

注意到 $\mu\in(0,1)$, (5.11) 和上式表明：当 k 充分大时， $f(x^k+d^k+\widehat{d}^k)\le f(x^k)+\mu\nabla f(x^k)^T d^k$.

定理 5.4.1 *若 $H1-H6$ 成立，则 5.2.2 节中算法所产生的序列 $\{x^k\}$ 超线性收敛，即， $\|x^{k+1}-x^*\|=o(\|x^k-x^*\|)$. 如果还有 $\nabla^2 f,\nabla^2 g_i(\forall i\in I)$ Lipschitz 连续且当 k 充分大时， $H_k=\nabla_{xx}^2 L(x^k,\lambda^k)$, 则收敛速度是二次的，即， $\|x^{k+1}-x^*\|=O(\|x^k-x^*\|^2)$.*

该定理的证明见文献 [106].

5.5 数值试验

本节，我们对 5.2.2 节中算法和文献 [106] 中的算法进行了数值比较. 两种算法都用 MATLAB 6.5 编程，都在 Windows XP 的个人计算机实现. 其中本章算法步 4 和步 5 中的线性方程组都使用优化工具箱求解. 5.2.2 节中算法用 Algo51 表示，文献 [102] 中的算法用 Algo52 表示.

整个实验中，我们使用如下 BFGS 公式：

$$H_{k+1} = H_k - \frac{H_k s^k (s^k)^T H_k}{(s^k)^T H_k s^k} + \frac{y^k (y^k)^T}{(s^k)^T y^k},$$

其中

$$y^k = \begin{cases} \widetilde{y}^k, & \text{if} \quad (s^k)^T \widetilde{y}^k \geq 0.2 s^{kT} H_k s^k; \\ \vartheta_k \widetilde{y}^k + (1-\vartheta_k) H_k s_k, & \text{otherwise}; \end{cases}$$

$$s^k = x^{k+1} - x^k, \ \widetilde{y}^k = \nabla_x L(x^{k+1}, \lambda^k) - \nabla_x L(x^k, \lambda^k),$$

$$\vartheta_k = \frac{0.8 s_k^T H_k s_k}{s_k^T H_k s_k - s_k^T \widetilde{y}^k},$$

整个实验中，令

$$H_1 = I_n, \quad \tau = 2.85. \quad \mu = 0.25, \quad \beta = 0.6, \quad \nu = 2.7, \quad \vartheta = 0.3,$$

$$\sigma = 0.4, \quad \sigma_1 = 3.5, \quad \varepsilon^0 = 2, \quad M^0 = 0.8.$$

表 5.1 的测试问题来自文献 [110] 和文献 [111]. 所选测试问题的初始点和文献 [110], 文献 [111] 中相同. 表 5.1 的列表示如下意义：**prob** 表示来自文献 [110] 和文献 [111] 的测试问题；**Ni** 表示求解的迭代数；**Nf, Ng** 分别表示约束函数值和约束函数梯度值的迭代次数. **objective, dnorm** 和 **eps** 表示目标函数值，模 d^k 和精度 ϵ.

表 5.1 数值结果

Algo	Prob	Ni	Nf	Ng	objective	dnorm	eps
Algo51	HS001	7	60	12	-1.0000	2.7509e-009	1e-06
Algo52	-	8	78	15	-1.0000	1.7988e-009	1e-06
Algo51	HS12	9	36	6	-30.0000	3.5097e-007	1e-06
Algo52	-	8	32	6	-30.0000	4.9214e-009	1e-06
Algo51	HS29	11	43	8	-22.6274	9.5834e-007	1e-06
Algo52	-	13	53	10	-22.6274	8.2007e-008	1e-06
Algo51	HS31	9	77	19	-6.0000	8.9923e-007	1e-06
Algo52	-	10	83	19	-6.0000	9.5086e-008	1e-06
Algo51	HS34	29	92	73	-0.8340	1.0992e-008	1e-06
Algo52	-	37	132	103	-0.8340	3.2684e-014	1e-06
Algo51	HS35	11	320	23	0.1111	2.3668e-010	1e-06
Algo52	-	10	312	21	0.1111	2.5179e-008	1e-06
Algo51	HS43	12	363	24	-44.0000	6.9779e-008	1e-06
Algo52	-	11	355	22	-44.0000	2.0101e-007	1e-06
Algo51	HS100	19	901	36	680.6301	5.4214e-007	1e-06
Algo52	-	19	632	29	680.6301	1.5996e-007	1e-06
Algo51	HS113	32	594	66	24.3062	5.4193e-007	1e-06
Algo52	-	33	611	68	24.3062	4.4202e-007	1e-06
Algo51	S225	7	110	9	2.0000	3.5510e-010	1e-06
Algo52	-	6	103	8	2.0000	1.8097e-010	1e-06
Algo51	S264	13	210	16	-44.1134	2.7452e-007	1e-06
Algo52	-	13	201	14	-44.1134	8.1741e-007	1e-06
Algo51	S388	79	1655	281	-5821.1000	3.8179e-007	1e-06
Algo52	-	76	1595	227	-5821.1000	3.0200e-007	1e-06

第 6 章　求解约束优化问题的一个内点型可行点 QP-free 算法

6.1　引言

本章我们继续考虑求解如下不等式约束优化问题

$$(P)\quad \min_{x\in\Re^n} f(x) \qquad \text{s.t. } g_i(x)\le 0, i\in I. \tag{6.1}$$

其中 $f: R^n\to R$, $g_i:\Re^n\to\Re$ 连续可微. 我们记

$$F=\{x\in R^n: g_i(x)\le 0, i\in I\},\ F^0=\{x\in\Re^n: g_i(x)<0, i\in I\}.$$

在文献 [105] 中， Qi,Qi 基于互补函数和 KKT 条件，提出了一个求解问题 (P) 的内点型可行点 QP-free 算法，他们在无严格互补条件下证明了系数矩阵的一致非奇异性和近似乘子序列的有界性. 但是，他们算法的每一个迭代步，为得到搜索方向，需解三个系数矩阵相同的线性方程组和一个最小二乘问题. 为减少计算量，最近， Zhu[109] 给出了一个新的内点型可行点 QP-free 算法，在此算法的每一个迭代步，为得到搜索方向，只需解三个相同系数矩阵的线性方程组，但是系数矩阵需在严格互补条件下保持一致非奇异性. 而且为得到全局收敛性，需要稳定点数目有限的限制.

本章中，我们以文献 [105] 中算法为基础，给出一个改进的内点型可行点 QP-free 算法，在该算法的每一个迭代步，为得到搜索方向，只需解三个系数矩阵相同的线性方程组，在无严格互补条件下得到系数矩阵的一致非奇异性和近似乘子序列的有界性，而且其全局收敛性不受稳定点数目有限的限制.

6.2　算法描述

首先，我们按照文献 [105] 给出有关记号.
令 $x^k\in F^0$, $(x^k,\mu^k)\in R^{n+m}$, 定义

$$\eta_i^k=\frac{g_i(x^k)}{\sqrt{g_i^2(x^k)+(\mu_i^k)^2}}+1,\ \theta_i^k=\left\{1-\frac{\mu_i^k}{\sqrt{g_i^2(x^k)+(\mu_i^k)^2}}\right\}^{\frac12},$$

$$\delta_i^k=\left\{\frac{1}{\sqrt{g_i^2(x^k)+(\mu_i^k)^2}(\mu_i^k+\sqrt{g_i^2(x^k)+(\mu_i^k)^2})}\right\}^{\frac12}.$$

易知，$(\eta_i^k)^2+(\theta_i^k)^4\geq 3-2\sqrt{2}$，$\theta_i^k=-\delta_i^k g_i(x^k)$.

若令 $A_k=\nabla g(x^k)=(\nabla g_i(x^k),\ i\in I)$，$\Gamma=diag(\eta_k)$，$\Theta_k=-\sqrt{2}diag(\theta_k)$，可定义矩阵

$$V_k=\begin{bmatrix} H_k & A_k \\ \Gamma_k A_k^T & \Theta_k \end{bmatrix},$$

其中 H_k 为一对称正定矩阵.

下面我们给出内点型可行点 QP-free 算法.

内点型可行点 QP-free 算法步骤:

参数及初始值: 初始点 $x^0\in F^0$，初始对称正定矩阵 $H_0\in R^{n\times n}$，$\theta,\beta\in(0,1)$，$\alpha\in(0,\frac{1}{2})$，$2<\tau<\delta<3,\gamma\in(0,1),\overline{\mu}>0,0<\mu_j^0<\overline{\mu},j=1,\cdots,m,\ k:=0$.

步 1. 牛顿步 d_0^k 的计算

解如下线性方程组得 (d_0^k,π_k)

$$V_k\begin{bmatrix} d \\ \lambda \end{bmatrix}=-\begin{bmatrix} \nabla f(x^k) \\ 0 \end{bmatrix}. \tag{6.2}$$

如果 $d_0^k=0$，停.

步 2. 搜索方向的计算

(i) 计算下降方向 d_1^k

解如下线性方程组得 $(d_1^k,\widetilde{\pi}_k)$

$$V_k\begin{bmatrix} d \\ \lambda \end{bmatrix}=-\begin{bmatrix} \nabla f(x^k) \\ \|d_0^k\|^\delta e \end{bmatrix}, \tag{6.3}$$

其中 $e=(1,1,\cdots,1)\in R^m$.

(ii) 计算主搜索方向 d^k

$$d^k=(1-\beta_k)d_0^k+\beta_k d_1^k,\quad \lambda^k=(1-\beta_k)\pi_k+\beta_k\widetilde{\pi}_k,$$

其中 $\beta_k=\max\{\beta\in(0,1)|(1-\beta)\nabla f(x^k)^T d_0^k+\beta\nabla f(x^k)^T d_1^k\leq\theta\nabla f(x^k)^T d_0^k\}$.

(iii) 计算修正方向 $\widetilde{d}^k$

解如下线性方程组得 $(\widetilde{d}^k,\widetilde{\lambda}_k)$

$$V_k\begin{bmatrix} d \\ \lambda \end{bmatrix}=-\begin{bmatrix} 0 \\ \Gamma_k(\Psi_k e+\widetilde{g}^k) \end{bmatrix}, \tag{6.4}$$

其中

$$I_k=\{i|g_i(x^k)\geq-\lambda_i^k\},\ \Psi_k=\begin{cases} \max\{\|d^k\|^\tau,\max\limits_{i\in I_k}|\frac{\eta_i^k}{\sqrt{2}\delta_i^k\lambda_i^k}-1|^\gamma\|d^k\|^2\}, & \text{if } i\in I_k; \\ 0, & \text{if } i\notin I_k. \end{cases}$$

$$\widetilde{g}^k=(\widetilde{g}_i^k,i\in I),\ \widetilde{g}_i^k=\begin{cases} g_i(x^k+d^k), & \text{if } i\in I_k;\\ 0, & \text{if } i\notin I_k.\end{cases}$$

如果 $\|\widetilde{d}^k\|\geq\|d^k\|$, 令 $\widetilde{d}^k=0$.

步 3. 按下式确定参数组 v_i^k, $i\in I$.

$$v_i^k=\begin{cases} 0, & \text{if } \lambda_i^k\geq 0;\\ 1, & \text{if } \lambda_i^k<0.\end{cases}$$

步 4. 线搜索

求序列 $\{\beta,\beta^2,\beta^3,\beta^4,\cdots\}$ 中满足如下不等式组的第一个数作为 t_k.

$$f(x^k+td^k+t^2\widetilde{d}^k)\leq f(x^k)+\alpha t\nabla f(x^k)^Td^k,\quad g_i(x^k+td^k+t^2\widetilde{d}^k)<v_i^kg_i(x^k),\quad i\in I. \tag{6.5}$$

步 5. 令 $\mu_j^{k+1}=\min\{\max\{\pi_j^k,\|d_0^k\|\},\overline{\mu}\},x^{k+1}=x^k+t_kd^k+t_k^2\widetilde{d}^k$, 生成一个新的对称正定矩阵 H_{k+1}, 令 $k:=k+1$, 进入步 1.

6.3 全局收敛性分析

本节我们将给出 6.2 节中算法的全局收敛性. 首先做如下假设.

H1 集合 F 是有界的.

H2 对任意 $x\in F$, 向量组 $\{\nabla g_i(x),i\in I(x)\}$ 线性无关, 其中 $I(x)=\{i|g_i(x)=0\}$.

H3 对任意 k, 任意 $d\in R^n$, 存在两个正常数 a,b, 满足

$$a\|d\|^2\leq d^TH_kd\leq b\|d\|^2.$$

下面的引理见文献 [105].

引理 6.3.1 *对任意的 k, 矩阵序列 $\{V_k^{-1}\}$ 是有界的.*

引理 6.3.2 *(i) 如果 $d_0^k=0$, 那么 x^k 是问题 (P) 的 KKT 点.*

(ii) 如果 $d_0^k\neq 0$, 那么

$$\nabla f(x^k)^Td_0^k<-d_0^{kT}H_kd_0^k,\quad \nabla f(x^k)^Td^k\leq\theta\nabla f(x^k)^Td_0^k<0,$$

$$\nabla g_i(x^k)^Td^k=-\sqrt{2}\frac{\delta_i^k}{\eta_i^k}\lambda_i^kg_i(x^k)-\frac{\beta_k}{\eta_i^k}\|d_0^k\|^\delta,\quad i\in I.$$

证明: 由 (6.2) 知:

$$\nabla f(x^k)^Td_0^k=-\sum_{i\in I}\frac{\sqrt{2}\theta_i^k}{\eta_i^k}(\pi_i^k)^2-d_0^{kT}H_kd_0^k<-d_0^{kT}H_kd_0^k.$$

结合 d^k 的定义和上式可得: $f(x^k)^T d^k \le \theta \nabla f(x^k)^T d_0^k < 0$.

由 (6.2),(6.3) 和 d^k 的定义有:

$$\nabla g_i(x^k)^T d^k = (1-\beta_k) g_i(x^k)^T d_0^k + \beta_k g_i(x^k)^T d_1^k = -\sqrt{2}\frac{\delta_i^k}{\eta_i^k}\lambda_i^k g_i(x^k) - \frac{\beta_k}{\eta_i^k}\|d_0^k\|^\delta.$$

引理 6.3.3 *步 4 中的线搜索是适定的.*

证明: 首先, 由 (6.5) 和泰勒展式可知:

$$\begin{aligned} a_k &\triangleq f(x^k + td^k + t^2\widetilde{d}^k) - f(x^k) - \alpha t \nabla f(x^k)^T d^k \\ &= \nabla f(x^k)^T (td^k + t^2\widetilde{d}^k) - \alpha t \nabla f(x^k)^T d^k + o(t) \\ &= (1-\alpha) t \nabla f(x^k)^T d^k + o(t). \end{aligned}$$

又 f 连续可微, $\nabla f(x^k)^T d^k < 0$, $\alpha \in (0, \frac{1}{2})$, 故存在 $\bar{t} > 0$, 使得任意 $t \in [0, \bar{t}]$, $a_k \le 0$.

其次, 由 (6.5), 泰勒展式和引理 6.3.2 可知:

$$\begin{aligned} b_i^k &\triangleq g_i(x^k + td^k + t^2\widetilde{d}^k) - \nu_i^k g_i(x^k) \\ &= (1-\nu_i^k) g_i(x^k) + t\nabla g_i(x^k)^T d^k + o(t) \\ &= (1-\nu_i^k - \sqrt{2}\tfrac{\delta_i^k}{\eta_i^k}\lambda_i^k t) g_i(x^k) - t\tfrac{\beta_k}{\eta_i^k}\|d_0^k\|^\delta + o(t). \end{aligned}$$

结合 ν_i^k 的定义可知: 存在 $\bar{t}_i > 0$, 使得任意 $t \in [0, \bar{t}_i]$, $b_i^k \le 0$.

令 $\widehat{t} = \min\{\bar{t}, \bar{t}_i,\ i \in I\}$ 可知结论成立.

下面将分析 6.2 节中算法的全局收敛性. 根据假设 $H2$, $H3$, 我们可假定存在一无限子集 K, 使得

$$x^k \to x^*,\ H_k \to H_*,\ d_0^k \to d_0^*,\ d_1^k \to d_1^*,\ \pi_k \to \pi_*,\ \widetilde{\pi}_k \to \widetilde{\pi}_*,\ \mu^k \to \mu^* (k \in K).$$

引理 6.3.4 *如果 $x^k \to x^*$, $k \in K$, 那么 $d_0^k \to 0$, $k \in K$.*

证明: 由 $\{f(x^k)\}$ 的单调下降性和 f 的连续性并结合 $x^k \to x^*(k \in K)$ 可知:

$$f(x^k) \to f(x^*),\ k \to \infty. \tag{6.6}$$

假设 $d_0^* \ne 0$, 则由步 5 知:

$$\mu_j^k \to \mu_j^* > 0,\ j \in I,\ k \in K.$$

从而由 (6.2) 易知: (d_0^*, π_*) 是下面线性方程组的唯一解.

$$V_k = \begin{bmatrix} H_* & A_* \\ \Gamma_* A_*^T & \Theta_* \end{bmatrix} \begin{bmatrix} d \\ \lambda \end{bmatrix} = \begin{bmatrix} -\nabla f(x^*) \\ 0 \end{bmatrix},$$

因此 $\nabla f(x^*)^T d_0^* < 0$. 又由 (6.3) 易知：$(d_1^*, \widetilde{\pi}_*)$ 是下面线性方程组的唯一解.

$$V_k = \begin{bmatrix} H_* & A_* \\ \Gamma_* A_*^T & \Theta_* \end{bmatrix} \begin{bmatrix} d \\ \lambda \end{bmatrix} = \begin{bmatrix} -\nabla f(x^*) \\ -\|d_0^*\|^\delta e \end{bmatrix}.$$

类似于 d^k 的定义，我们可定义 d_*, 模仿引理 6.3.2 的证明可得：

$$\nabla f(x^*)^T d^* < 0,\ \ \nabla g_i(x^*)^T d^* = -\sqrt{2}\frac{\delta_i^*}{\eta_i^*}\lambda_i^* g_i(x^*) - \frac{\beta_*}{\eta_i^*}\|d_0^*\|^\delta,\ \ i \in I.$$

类似于引理 6.3.3 的证明，易知步 4 中的线搜索所得步长 t_k 满足下面的关系：

$$t_k \ge t_* > 0,\ k \in K.$$

因此由 (6.5) 和上面的关系式，可得：

$$0 = \lim_{k\in K}(f(x^{k+1}) - f(x^k)) \le \lim_{k\in K} \alpha t_k \nabla f(x^k)^T d^k \le \alpha t^* \nabla f(x^*)^T d^* < 0.$$

上述矛盾表明：$d_0^k \to 0$.

类似于引理 5.3.2 的证明，可得如下引理.

引理 6.3.5 *对所有的 k, 存在正常数 κ 满足 $\|d^k - d_0^k\| \le \kappa \|d_0^k\|^\delta$.*

定理 6.3.1 *6.2 节中算法或有限步终止于问题 (P) 的 KKT 点，或产生一无穷点列 $\{x^k\}$, 使得其任一聚点是问题 (P) 的 KKT 点.*

证明: 由引理 6.3.2 知: 该定理的第一部分结论是显然的. 因此可假设 $x^k \to x^*$, $k \in K$. 由引理 6.3.4 知：$d_0^k \to d_0^*$, $k \in K$. 结合 (6.2),(6.3), d^k 的定义和引理 6.3.5 可得：

$$\nabla f(x^*) + A_* \lambda_* = 0,\ \ \lambda_i^* \theta_i^* = 0,\ i \in I.$$

如果对 $\forall i \in I$, 有 $g_i(x^*) < 0$, 则 $\theta_i^* > 0$, 从而 $\lambda_i^* = 0$, $\nabla f(x^*) = 0$. 即，x^* 是问题 (P) 的 KKT 点.

如果存在 $i_0 \in I$, 使得 $g_{i_0}(x^*) = 0$. 若 $\lambda_{i_0}^* \ge 0$, 则 x^* 是问题 (P) 的 KKT 点. 若 $\lambda_{i_0}^* < 0$, 由于 $g_{i_0}(x^k) < 0$, 故存在 $\overline{K} \subseteq K$, 使得 $g_{i_0}(x^{k+1}) > g_{i_0}(x^k)$, $k \in \overline{K}$, 结合 (6.5) 和 $\nu_{i_0}^k$ 的定义可得：$\nu_{i_0}^k = 0$, $k \in \overline{K}$, 从而 $\lambda_{i_0}^k \ge 0$, $k \in \overline{K}$. 这与 $\lambda_{i_0}^* < 0$ 矛盾！

6.4 收敛速度分析

本节我们将分析 6.2 节中算法的收敛速度. 假定 (x^*, λ^*) 是该算法所生成序列 $(x^k.\lambda^k)$ 的一个聚点，那么从全局收敛性定理知，(x^*, λ^*) 是问题 P 的 KKT

点. 下面我们做如下假设:

H4(i) 函数 $f(x), g_i(x)$ 都是二次连续可微的.

(ii) 在 (x^*, λ^*) 处严格互补条件成立且 $\lambda_j^* \le \overline{\mu}$, $\forall j \in I$.

H5 在 (x^*, λ^*) 处二阶充分条件条件成立, 即在子空间 $\{u | \nabla g_i(x^*)^T u = 0, \forall i \in I(x^*)\}$ 上 $\nabla^2_{xx} L(x^*, \lambda^*)$ 是正定的.

引理 6.4.1 *在上述假设下, 序列* $\{x^k\}$ *收敛到* x^*. *即,* $x^k \to x^* (k \to \infty)$.

证明: 从 (6.5) 和引理 6.3.2 可知:

$$f(x^{k+1}) \le f(x^k) + \alpha t_k \nabla f(x^k)^T d^k \le f(x^k) - \alpha\theta t_k d_0^{kT} H_k d_0^k.$$

考虑到 (6.6) 和 $H3$, 可得到 $t_k \|d_0^k\| \to 0$. 因此利用引理 6.3.5 可得, $t_k \|d_k\| \to 0$. 从而 $\|x^{k+1} - x^k\| \to 0$. 又 $H4, H5$ 成立, 类似于文献 [88] 中的性质 4.1 的证明, 可得 $\lim\limits_{k\to\infty} x^k = x^*$.

下面两个引理见文献 [105] 中的推论 4.3 和引理 4.4.

引理 6.4.2 *在上述假设下, 下列关系成立:*

$$d_0^k \to 0,\ \pi_k \to \pi_*,\ \widetilde{\pi}_k \to \widetilde{\pi}_*,\ \lambda_k \to \lambda_*,\ \mu^k \to \lambda_* (k \to \infty).$$

引理 6.4.3 *当* k *充分大时, 有* $I_k = I(x^*)$. *对* $i \in I(x^*)$, *有*

$$\eta_i^k \to 1,\ \theta_i^k \to 0, \frac{\lambda_i^k \theta_i^k}{g_i(x^k)} \to -\frac{1}{\sqrt{2}},\ \lambda_i^k \delta_i^k \to \frac{1}{\sqrt{2}}.$$

对 $i \notin I(x^*)$, *有* $\eta_i^k \to 0,\ \theta_i^k \to 1$.

为分析 6.2 节中算法的超线性收敛性, 我们在做如下假设.

H6 矩阵序列 $\{H_k\}$ 满足

$$\|P_k(H_k - \nabla^2_{xx} L(x^*, \lambda^*)) d^k\| = o(\|d^k\|),$$

其中 $P_k = I_n - N_k(N_k^T N_k)^{-1} N_k^T$, $N_k = \nabla g_{I(x^*)}(x^k)$.

引理 6.4.4 *(i) 当* k *充分大时,* d^k *可分解为* $d^k = P_k d^k + \widehat{d}^k$, *其中* $\widehat{d}^k = O(\|g_{I(x^*)}(x^k)\|) + o(\|d_0^k\|^2)$,

(ii) 当 k *充分大时,*

$$\|\widehat{d}^k\| = O(\max\{\|d^k\|^2, \max_{i \in I(x^*)} |\frac{\eta_i^k}{\sqrt{2}\delta_i^k \lambda_i^k} - 1| \|d^k\|\}) = o(\|d^k\|),$$

$$\|\widetilde{\lambda}^k\| = O(\max\{\|d^k\|^2, \max_{i \in I(x^*)} |\frac{\eta_i^k}{\sqrt{2}\delta_i^k \lambda_i^k} - 1| \|d^k\|\}) = o(\|d^k\|).$$

证明: (i) 由引理 6.3.2 可知:

$$\nabla g_i(x^k)^T d^k = -\sqrt{2}\frac{\delta_i^k}{\eta_i^k}\lambda_i^k g_i(x^k) - \frac{\beta_k}{\eta_i^k}\|d_0^k\|^\delta, \quad i \in I.$$

特别有

$$N_k^T d^k = h^k,$$

其中 h^k 是一个 $|I(x^*)|$ 维向量，其分量为:

$$-\sqrt{2}\frac{\delta_i^k}{\eta_i^k}\lambda_i^k g_i(x^k) - \frac{\beta_k}{\eta_i^k}\|d_0^k\|^\delta, \quad i \in I(x^*).$$

由引理 6.3.2 知:

$$\|h^k\| = O(\|g_{I(x^*)}(x^k)\|) + o(\|d_0^k\|^2).$$

又

$$d^k = P_k d^k + N_k(N_k^T N_k)^{-1} N_k^T d^k.$$

故令 $\widetilde{d}^k = N_k(N_k^T N_k)^{-1} N_k^T d^k$. 易知结论成立.
(ii) 由泰勒展式和 $\widetilde{g}^k$ 的定义知: 对 $i \in I_k$,

$$-\Psi_i^k - \widetilde{g}_i^k = -\max\{\|d^k\|^\tau, \max_{i\in I_k}|\frac{\eta_i^k}{\sqrt{2}\delta_i^k\lambda_i^k} - 1|^\gamma \|d^k\|^2\} - g_i(x^k) - \nabla g_i(x^k)^T d^k + O(\|d^k\|^2).$$

结合引理 6.3.1, 引理 6.3.2(ii), 引理 6.3.5 易知结论成立. 对 $i \notin I_k$, 可类似证明.

引理 6.4.5 *假定 $H1 - H5$ 成立，则当 k 充分大时，步长 $t_k \equiv 1$.*

证明: 由 $H4$ 易知, $\nu_i^k = 0$, $i \in I$. 因此只需证明:

$$f(x^k + d^k + \widetilde{d}^k) \leq f(x^k) + \alpha \nabla f(x^k)^T d^k,$$

$$g_i(x^k + d^k + \widetilde{d}^k) < 0, \ i \in I.$$

当 $i \in I(x^*)$ 时，结合引理 6.3.2(ii) 和引理 6.3.5 知: $g_i(x^k) = O(\|d^k\|)$, $i \in I(x^*)$.
由 (6.4) 和引理 6.4.4 知:

$$\begin{aligned}
\nabla g_i(x^k)^T \widetilde{d}^k &= -\sqrt{2}\tfrac{\delta_i^k}{\eta_i^k} g_i(x^k)\widetilde{\lambda}^k - \Psi_i^k - g_i(x^k + d^k) \\
&= -\Psi_i^k + O(\max\{\|d^k\|^3, \max_{i\in I(x^*)}|\tfrac{\eta_i^k}{\sqrt{2}\delta_i^k\lambda_i^k} - 1|\|d^k\|^2\}) - g_i(x^k + d^k).
\end{aligned}$$

因此,

$$\begin{aligned}
g_i(x^k + d^k + \widetilde{d}^k) &= g_i(x^k + d^k) + \nabla g_i(x^k + d^k)^T \widetilde{d}^k + O(\|\widetilde{d}^k\|^2) \\
&= g_i(x^k + d^k) + \nabla g_i(x^k)^T \widetilde{d}^k + O(\|d^k\|\|\widetilde{d}^k\|) + O(\|\widetilde{d}^k\|^2) \\
&= -\Psi_i^k + O(\max\{\|d^k\|^3, \max_{i\in I(x^*)}|\tfrac{\eta_i^k}{\sqrt{2}\delta_i^k\lambda_i^k} - 1|\|d^k\|^2\}).
\end{aligned} \tag{6.7}$$

故当 k 充分大时， $g_i(x^k+d^k+\widetilde{d}^k)<0$, $i\in I(x^*)$. 当 $i\notin I(x^*)$ 时，由 $g_i(x)$ 的连续性和 $g_i(x^*)<0$ 易知结论成立.

下证 $f(x^k+d^k+\widetilde{d}^k)\le f(x^k)+\alpha\nabla f(x^k)^Td^k$ 对充分大的 k 成立.

由泰勒展式和引理 6.4.4 有：

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

从 (6.2),(6.3),(6.4) 和 d^k 的定义得：

$$H_kd^k+\nabla f(x^k)+\sum_{i\in I(x^*)}\lambda_i^k\nabla g_i(x^k)=0,$$

$$\nabla g_i(x^k)^Td^k=-\sqrt{2}\frac{\delta_i^k}{\eta_i^k}\lambda_i^kg_i(x^k)-\frac{\beta_k}{\eta_i^k}\|d_0^k\|^\delta,\ i\in I(x^*),\tag{6.8}$$

$$g_i(x^k+d^k)+\nabla g_i(x^k)^T\widetilde{d}^k=o(\|d^k\|^2),\ i\in I(x^*).\tag{6.9}$$

由上面第一式和引理 6.4.4 可得：

$$\nabla f(x^k)^Td^k=-(d^k)^TH_kd^k-\sum_{i\in I(x^*)}\lambda_i^k\nabla g_i(x^k)^Td^k,$$

$$\nabla f(x^k)^T\widetilde{d}^k=-\sum_{i\in I(x^*)}\lambda_i^k\nabla g_i(x^k)^T\widetilde{d}^k+o(\|d^k\|^2).$$

因此

$$\begin{aligned}f(x^k+d^k+\widetilde{d}^k)&=f(x^k)+\tfrac{1}{2}\nabla f(x^k)^Td^k+\tfrac{1}{2}(d^k)^T(\nabla_{xx}^2f(x^k)-H_k)d^k\\&-\tfrac{1}{2}\sum_{i\in I(x^*)}\lambda_i^k\nabla g_i(x^k)^Td^k-\sum_{i\in I(x^*)}\lambda_i^k\nabla g_i(x^k)^T\widetilde{d}^k+o(\|d^k\|^2).\end{aligned}\tag{6.10}$$

由泰勒展式和 (6.9) 得：

$$g_i(x^k)+\nabla g_i(x^k)^T(d^k+\widetilde{d}^k)+\frac{1}{2}(d^k)^T\nabla_{xx}^2g_i(x^k)d^k=o(\|d^k\|^2).$$

从 (6.8) 和上式知：

$$\begin{aligned}&-\frac{1}{2}\sum_{i\in I(x^*)}\lambda_i^k\nabla g_i(x^k)^Td^k-\sum_{i\in I(x^*)}\lambda_i^k\nabla g_i(x^k)^T\widetilde{d}^k\\&=\frac{1}{2}\sum_{i\in I(x^*)}\lambda_i^k\nabla g_i(x^k)^Td^k-\sum_{i\in I(x^*)}\lambda_i^k\nabla g_i(x^k)^T(d^k+\widetilde{d}^k)\\&=\sum_{i\in I(x^*)}(1-\sqrt{2}\tfrac{\delta_i^k}{2\eta_i^k})\lambda_i^kg_i(x^k)-\frac{1}{2}\sum_{i\in I(x^*)}\frac{\beta_k}{\eta_i^k}\lambda_i^k\|d^k\|^\delta\\&+\frac{1}{2}\sum_{i\in I(x^*)}\lambda_i^k(d^k)^T\nabla_{xx}^2g_i(x^k)d^k+o(\|d^k\|^2)\\&=\sum_{i\in I(x^*)}(1-\sqrt{2}\tfrac{\delta_i^k}{2\eta_i^k})\lambda_i^kg_i(x^k)+\frac{1}{2}\sum_{i\in I(x^*)}\lambda_i^k(d^k)^T\nabla_{xx}^2g_i(x^k)d^k+o(\|d^k\|^2).\end{aligned}\tag{6.11}$$

由假设 $H4$ 和引理 6.3.2 有:

$$\sum_{i\in I(x^*)}(1-\sqrt{2}\frac{\delta_i^k}{2\eta_i^k})\lambda_i^k g_i(x^k)+o(\|g_{I(x^*)}(x^k)\|)<0.$$

结合上式, (6.11) 和假设 $H6$, 由 (6.10) 可得:

$$\begin{aligned}
f(x^k+d^k+\widetilde{d}^k) &= f(x^k)+\tfrac{1}{2}\nabla f(x^k)^T d^k+\sum_{i\in I(x^*)}(1-\sqrt{2}\tfrac{\delta_i^k}{2\eta_i^k})\lambda_i^k g_i(x^k)\\
&+ \frac{1}{2}(d^k)^T(\nabla_{xx}^2 f(x^k)+\sum_{i\in I(x^*)}\lambda_i^k\nabla_{xx}^2 g_i(x^k)-H_k)d^k+o(\|d^k\|^2)\\
&= f(x^k)+\tfrac{1}{2}\nabla f(x^k)^T d^k+\sum_{i\in I(x^*)}(1-\sqrt{2}\tfrac{\delta_i^k}{2\eta_i^k})\lambda_i^k g_i(x^k)+o(\|g_{I(x^*)}(x^k)\|)\\
&+ \frac{1}{2}d^{kT}P_k(\nabla_{xx}^2 f(x^k)+\sum_{i\in I(x^*)}\lambda_i^k\nabla_{xx}^2 g_i(x^k)-H_k)d^k+o(\|d^k\|^2)\\
&\le f(x^k)+\tfrac{1}{2}\nabla f(x^k)^T d^k\\
&+ \frac{1}{2}\|d^k\|\|P_k(\nabla_{xx}^2 L(x^k,\lambda_k)-H_k)d^k\|+o(\|d^k\|^2)\\
&= f(x^k)+\tfrac{1}{2}\nabla f(x^k)^T d^k+o(\|d^k\|^2).
\end{aligned}$$

由引理 6.3.2 和假设 $H3$ 知:

$$\begin{aligned}
\nabla f(x^k)^T d^k &\le \theta\nabla f(x^k)^T d_0^k\\
&\le -\theta d_0^{kT}H_k d_0^k\\
&\le -\theta a\|d_0^k\|^2.
\end{aligned}$$

注意到 $\alpha\in(0,1)$, 引理 6.3.5 和上式可得：当 k 充分大时,

$$f(x^k+d^k+\widetilde{d}^k)\le f(x^k)+\alpha\nabla f(x^k)^T d^k.$$

类似于文献 [107] 中定理 5.2 的证明方法，可得如下定理.

定理 6.4.1 *若 $H1-H6$ 成立，则有该算法所生成的序列 $\{x^k\}$ 超线性收敛，即,* $\|x^{k+1}-x^*\|=o(\|x^k-x^*\|)$.

6.5 数值试验

本节，我们对 6.2 节中的算法和文献 [109] 中的算法进行了数值比较. 两种算法都用 MATLAB 6.5 编程，都在 Windows XP 的个人计算机实现. 其中 (6.2), (6.3) 和 (6.4) 都使用优化工具箱求解. 6.2 节中的算法用 Algo61 表示，文献 [109] 中的算法用 Algo62 表示.

整个实验中，我们使用如下 BFGS 公式:

$$H_{k+1}=H_k-\frac{H_k s^k(s^k)^T H_k}{(s^k)^T H_k s^k}+\frac{y^k(y^k)^T}{(s^k)^T y^k},$$

其中

$$y^k = \begin{cases} \widetilde{y}^k, & \text{if} \quad (s^k)^T\widetilde{y}^k \geq 0.2s^{kT}H_ks^k; \\ \vartheta_k\widetilde{y}^k + (1-\vartheta_k)H_ks_k, & \text{otherwise}; \end{cases}$$

$$s^k = x^{k+1} - x^k,\ \widetilde{y}^k = \nabla_x L(x^{k+1}, \lambda^k) - \nabla_x L(x^k, \lambda^k),$$

$$\vartheta_k = \frac{0.8s_k^T H_k s_k}{s_k^T H_k s_k - s_k^T \widetilde{y}^k},$$

整个实验中，令

$$H_0 = I_n, \quad \tau = 2.5, \quad \delta = 2.8, \quad \beta = 0.6, \quad \alpha = 0.55, \quad \theta = 0.35, \quad \gamma = 0.6.$$

表 6.1 的测试问题来自文献 [110] 和文献 [111]. 所选测试问题的初始点和文献 [110], 文献 [111] 中相同. 表 6.1 的列表示如下意义： **prob** 表示来自文献 [110] 和文献 [111] 的测试问题； **Ni** 表示求解的迭代次数； **Nf, Ng** 分别表示约束函数值和约束函数梯度值的迭代次数. **objective, dnorm** 和 **eps** 表示目标函数值，模 d^k 和精度 ϵ.

表 6.1 数值结果

Algo	Prob	Ni	Nf	Ng	objective	dnorm	eps
Algo61	HS12	10	41	7	-30.0000	2.6980e-006	1e-06
Algo62	-	9	37	6	-30.0000	3.9052e-007	1e-06
Algo61	HS29	12	70	8	-22.6274	7.4282e-007	1e-06
Algo62	-	14	98	10	-22.6274	1.6131e-007	1e-06
Algo61	HS31	8	196	17	-6.0000	4.1021e-007	1e-06
Algo62	-	10	208	19	-6.0000	1.1714e-007	1e-06
Algo61	HS34	58	2452	130	-0.8340	8.2056e-015	1e-06
Algo62	-	60	2756	135	-0.8340	4.2665e-012	1e-06
Algo61	HS35	9	196	19	0.1111	7.0825e-010	1e-06
Algo62	-	14	236	25	0.1111	9.1633e-010	1e-06
Algo61	HS43	11	180	28	-44.0000	4.4855e-008	1e-06
Algo62	-	12	189	28	-44.0000	3.0538e-007	1e-06
Algo61	HS100	19	344	76	680.6301	2.5531e-006	1e-06
Algo62	-	19	442	56	680.6301	1.1857e-007	1e-06
Algo61	HS113	34	1594	66	24.3062	8.5715e-007	1e-06
Algo62	-	33	1464	58	24.3062	2.3525e-007	1e-06
Algo61	S264	12	234	19	-44.1134	1.6354e-007	1e-06
Algo62	-	14	243	22	-44.1134	5.3506e-008	1e-06
Algo61	S388	15	245	26	-5821.1000	6.1715e-007	1e-06
Algo62	-	14	231	23	-5821.1000	6.4045e-008	1e-06

结　论

本文研究非线性约束优化问题的求解. 我们提出几种序列二次规划 (SQP) 算法, 建立相应算法的收敛性, 并对所给算法进行数值实验.

第 2 章在文献 [83, 84, 85] 的基础上, 结合积极集估计技术, 提出了一个积极集可行点 SQP 算法, 该算法的主要优点在于: 算法的主搜索方向由一个低维的凸二次规划 (2.3) 确定, 而且不需求解任何子问题来修正 (2.3) 中参数 σ_k, 只需取 $\sigma_k = \|d^{k-1}\|^{\nu}$, 其中 d^{k-1} 为前一个迭代点处的主搜索方向. 为克服 Maratos 效应, 我们通过求解一个低维的最小二乘问题得到高阶修正方向. 在适当的条件下, 我们证明算法具有全局收敛性和超线性收敛性.

第 3 章提出一个求解极大极小问题的修正的 SQP 算法. 该算法的主搜索方向通过求解一个二次规划得到. 为克服 Maratos 效应, 不同于文献 [95][98], 我们通过求解一个线性方程组得到高阶修正方向, 无需求解二次规划子问题. 在较弱的条件下, 我们证明该算法是全局收敛和一步超线性收敛的.

第 4 章提出一个求解非线性约束优化问题的可行点 SQP 算法. 在该算法的每一个迭代步, 分别通过求解二次规划子问题和线性方程组得到一个下降方向和一个可行方向, 在此基础上, 我们构造一个可行下降方向. 为避免 Maratos 效应, 我们通过解一线性方程组得到高阶修正方向. 在适当的条件下, 该算法被证明是全局收敛和超线性收敛的. 与已有算法相比, 本章提出算法的优点是: 算法中线性方程组不涉及乘子估计, 所求解的两个线性方程组的系数矩阵相同且比文献 [102] 中系数矩阵的结构简单, 因而可减少计算量.

第 5 章以文献 [106] 中算法为基础, 提出一个求解非线性不等式约束优化问题的非内点型可行点 QP-free 算法. 这个算法不要求迭代点必须是可行域的内点. 在算法的每一个迭代步, 为得到搜索方向, 只须求解四个系数相同的线性方程组. 在适当的条件下, 我们建立该算法的全局收敛性和超线性收敛定理.

第 6 章以文献 [105] 中算法为基础, 提出一个改进的内点型可行点 QP-free 算法, 在该算法的每一个迭代步, 为得到搜索方向, 只需解三个系数矩阵相同的线性方程组, 而且在无严格互补条件下得到系数矩阵序列的一致非奇异性和近似乘子序列的有界性, 且其全局收敛性分析不受稳定点数目有限的限制.

在本文研究工作的基础上, 我们提出如下值得进一步研究的问题:

1. 本文第 3 章所提算法能否推广到求解约束极大极小问题?

2. 本文第 6 章所提算法中的线性方程组的系数矩阵涉及近似乘子, 能否设计一系数矩阵不含近似乘子且在无严格互补条件下也证明了系数矩阵序列的一致非奇异性和近似乘子序列的有界性?

3. 本文各算法的收敛性都是在非退化假设和严格互补条件下讨论的, 能否把

稳定 SQP 算法的思想用于上述算法？

4. 由于实际中的问题大多是大规模问题，能否利用本文思想，针对大规模问题的特点设计出有效的算法？

参 考 文 献

[1] 李董辉，童小娇，万中．数值最优化．北京：科学出版社． 2005, 201-222

[2] 袁亚湘，孙文瑜．最优化理论与方法．北京：科学出版社． 1997, 521-558

[3] 席少霖．非线性最优化方法．北京：高等教育出版社． 1992, 346-404

[4] Nocedal J., Wright S.J. Numerical optimization. Berlin: Springer, 1998, 528-573

[5] Wilson,R.B. A simplicial algorithm for concave programming. PhD theisis,Harvard university, Graduate School of Business Administration,1963.

[6] Han,S.P. Superlinearly convergent variable metric algorithms for general nonlinear programming problems. Mathematical Programming, 1976, 11: 263-282

[7] Han,S.P. Globally convergent method for nonlinear programming. Journal of Optimization Theory and Application, 1977, 22: 297-309

[8] Powell,M.J.D. Fast algorithms for nonlinear constrained optimization calculations. Waston,G.A.(ed).Numerical Analysis,Springer,Berlin, 1977, 144-157

[9] Boggs,P.T.,Tolle,J.W. Sequential quadratic programming, Acta Numerica, 1995, 4: 1-51

[10] Fukushima,M.,Luo,Z.Q.,Pang,J.S. A Globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints. Computational Optimization and Application, 1998, 10: 5-34

[11] Flecther,R.,Leyffer,S.,Ralph,D.,Scholtes,S. Local convergent of SQP method for mathematical programs with equilibrium constraints. Numerical Analysis Report, Department of Mathematics, University of Dundee,Dundee,Scotland,2001

[12] Jiang,H.Y.,Ralph,D., Smooth SQP methods for mathematical programs with nonlinear complementarity constraints. SIAM Journal on Optimization, 2000, 10: 779-808

[13] Bock,H.G.,Egartner,W.,Kappis,W.,Schulz,V. Practical shape optimization for turbine and compressor Blades by the use of PRSQP methods. Optimization and Engineering, 2002, 3: 395-414

[14] Ito,K.,Kunisch,K. Augmented Lagrangian SQP methods for nonlinear optimal constrol problems of tracking type. SIAM Journal on Control and Optimization, 1996, 34: 874-891

[15] Flecher,R. Practical mathod of optimization(2nd). Chicester-New York: Wiley, 1987

[16] Tone,K. Revisions of constrained approximation in the succssive QP mwthod for nonlinear programming problems. Mathematical Programming, 1983, 24: 114-152

[17] Schittkowski,K. The nonlinear programming method of Wilson,Han and Powell with augmented lagrangian type line search function. Numerische Mathematics, 1981, 38: 83-127

[18] Spellucci,P. A new technique for inconsistent QP problems in the SQP methods. Mathematical Methods of Operations Research, 1998, 47: 355-400

[19] Spellucci,P. An SQP method for general nonlinear programmings using only equality constrained subproblems. Mathematical Programming, 1998, 82: 413-448

[20] Burke,J.V., Han,S.P. A robust SQP method. Mathematical Programming, 1989, 43: 277-303

[21] Zhou,G.L. A modified SQP method and its global convergence. Journal of Global Optimization,1997, 11: 193-205

[22] 徐以凡，郑应平. 处理退化问题的一类 SQP 算法. 系统科学与数学， 2001, 21: 257-263

[23] Zhang,J.L., Zhang X.S. A robust SQP method with for optimization with inequality constraints. Journal of Computational Mathematics, 2003, 21:247-256

[24] Mo,J.T., Zhang,K.C.,Wei,Z.X. A variant of SQP method for inequality constrained optimization and its global convergence. Journal Computations and Applied Mathematics, 2006, 197: 270-281

[25] Lucidi,S. New results on a continuously differentiable exact penalty function. SIAM Journal on Optimization, 1992, 2: 558-574

[26] Facchinei,F. Robust recursive quadratic programming algorithm model with global and superlinear convergence properties. Journal of Optimization Theory and Applications, 1997, 92: 543-579

[27] Maratos,N. Exact penalty functions for finite dimensional and control optimization problems. PhD thesis, University of Science and Technology, London, 1978

[28] Pillo,G.D.,Facchinei,F.,Grippo,L. An RQP algorithm using a differentiable exact penalty function for inequality constrained problems. Mathematical Programming, 1992, 55: 49-68

[29] Powell,M.J.D.,Yuan,Y. A recursive quadratic programming algorithm that uses differentiable exact penalty function. Mathematical Programming, 1986, 35: 265-278

[30] Ramma,M.V. An exact duality theory for semidefite programming and its complexity implications. Mathematical Programming, 1997, 77: 129-162

[31] Yang,B.T.,Zhang,K.C.,You,Z.Y. A successive quadratic programming method that uses new corrections for search direction. Journal of Computational and Applied Mathematics, 1996, 71: 15-31

[32] Zhang,J.L., Zhang,X.S. An SQP method based on smoothing penalty function for nonlinear optimization with inequality constraints. Journal of Systems Science and Complexity, 2001, 14: 212-217

[33] 赖炎连，贺国平. 二阶修正的约束变尺度算法. 系统科学与数学， 1990, 10: 216-227

[34] 高自友，吴方. 非线性约束条件下的 SQP 可行方法. 应用数学学报， 1995, 18: 579-590

[35] 简金宝. SQP 技术与广义投影结合的次可行方向法. 高校应用数学学报， 1996, 11: 65-74

[36] Xu,Y.F.,Wang,W. A feasible and superlnear algorithm for inequality constrained minimization problems. ACTA Mathematicae Applicatae Sinica, 2000, 16: 36-41

[37] Zhang,J.L.,Wang,C.Y. A modified SQP method and its superliear convergence. OR Transactions, 2000, 4: 32-40

[38] 张菊亮，章祥荪. 一个等式约束问题的 SQP 方法及其收敛性. 应用数学学报， 2001, 24: 1-9

[39] Bonnans,J.F.,Panier,E.R.,Tits,A.T.,Zhou J. Avoiding the Maratos effect by means of nonmonotone line search II:Inequality constraints problems-Feasible iterates. SIAM Journal on Numerical Analysis, 1992, 29: 1187-1202

[40] Chamberlain,R.M., Powell,P.J.M.,Lemarechal,C.,Pedersen,H.C. The watchdog technique for forcing convergence in algorithm for constrained optimization. Mathematical Programming, 1982, 16: 1-17

[41] Gomes,F.A.M.,Maciel,M.C.,Martinez,J.M. Nonlinear programming algorithm using trust regions and augmented lagrangians with nonmonotone penalty parameters. Mathematical Programming, 1999, 84: 161-200

[42] He,G.P.,Diao,B.Q.,Gao,Z.Y. An SQP algorithm with nonmonotone line search for general nonlinear constrained optimization problem. Journal of Computational Mathematics, 1997, 15: 179-182

[43] Panier,E.R.,Tits,A.T.,Zhou J. Avoiding the Maratos effect by means of nonmonotone line search I:General constraints problems. SIAM Journal on Numerical Analysis, 1991, 28: 1183-1195

[44] 王薇，徐以凡，赖炎连. 一类非单调搜索的 SQP 算法. 运筹学学报， 1998, 2: 56-63

[45] Xu,Y.F.,Wang,W. A mixed superlinearly convergent algorithm with nomonotone search for constrained optimizations, Applied Mathematics A Journal Chinese Uiversity, Ser.B,2000, 15: 211-219

[46] Xu,Y.F.,Wang,W.,Gao,Z.Y. The algorithm of sequential KKT equations by nonmonotone search for arbitrary initial point. Computational Optimization and applications, 2001, 18: 221-232

[47] Zhang,J.L.,Zhang X.S. A modified SQP method with nonmonotone linesearch technique. Journal of Global Optimization, 2001, 21: 201-218

[48] Wright,S.J.Superlinear convergence of a stablized SQP method to a degenerate solution. Computational Optimization and Applications, 1998, 11: 253-275

[49] Wright,S.J.Constraints identification and algorithm stablized for degenerate nonliear programs. Mathematical programming, 2003, 95: 137-160

[50] Hager,W.W. Stablized Sequential Quadratic Programming. Computational Optimization and Applications, 1999, 12: 253-273

[51] Li,D.H.,Qi,L.Q. A stablized SQP Method via linear equations. Applied mathematical technique report AMR00/5, The University of New South Wales, 2000

[52] Fletcher,R., Leyffer,S. Nonlinear programming without a penalty function. Mathematical Programming, 2002, 91: 239-269

[53] Fletcher,R., Leyffer,S.Global convergence of trust SQP-filter algorithms for general nonlinear programming. SIAM Journal on Optimization, 1999, 13: 635-659

[54] Ulbrich,S. On the superlinear local convergence of a filter-SQP. Mathematical Programming, 2004, 100: 217-245

[55] *Wächter*,A.,Biegler,L.T. Line search fliter method for nonlinear programming: Motivation and global convergence. SIAM Journal on Optimization, 2005, 16: 1-31

[56] *Wächter*,A.,Biegler,L.T. Line search fliter method for nonlinear programming: Local convergence. SIAM Journal on Optimization, 2005, 16: 32-48

[57] 简金宝，薛声家. 非线性约束最优化一族超线性收敛的可行方法. 数学研究与评论， 1999, 19: 135-140

[58] 简金宝. 非线性约束最优化超线性与二次收敛算法的研究. 博士学位论文，西安，西安交通大学， 2000

[59] Qi,L.Q.,Yang,Y.F. Globally and supearlikearly QP-free algorithm for nonlinear constrained optimization. Journal of Optimization Theory and Applications, 2002, 113: 297-323

[60] Bonnans,J.F. Local analysis of Newton type methods for variational inequality and nonlinear programming. Applied Mathematics and Optimization, 1994, 29: 161-186

[61] 高自友，贺国平，赖炎连. 具有相容子问题的序列二次规划新算法. 中国科学 A 辑， 1996, 26: 991-1001

[62] 高自友，吴方．非线性约束条件下一个超线性收敛的可行方法．数学学报，1997, 40: 895-900

[63] Biggs,M.C.Recursive quadratic programming based on penalty functions for constrained optimization. in L.C.W.Dixon(eds), Nonlinear Optimization: Theory and Algorithm, 1980

[64] Herskovits,J.A two-stage feasible directions algorithm for nonlinear constrained optimization. Mathematical Programming, 1986, 36: 19-38

[65] Ling,C., Qi,L.Q., Zhou,G.L., Wu,S.Y. Global convergence of a robust smoothing SQP method for semi-infinite programming. Journal of Optimization Theory and Applications, 2006, 129: 147-164

[66] Facchinei,F., Lazzari,C. Local feasible QP-free algorithms for the constrained minimization of SC^1 functions. Journal of Optimization Theory and Applications, 2003, 119: 281-316

[67] Lazzari,C. New feasible QP-free algorithms for the constrained minimization of SC^1 functions. PhD Thesis, Department of computer and system science, University of Rome-La Sapienza, 2002

[68] Polak,E., Mayne,D.Q. A superlinearly convergent algorithm for constrained optimization problems. Mathematical Programming, 1982, 16: 45-61

[69] Boggs,P.T.,Tolle,J.W. Sequential quadratic programming for large scale nonlinear optimization. Journal of Computational and Applied Mathematics, 2000, 124: 123-137

[70] Gill,P.E.,hurray,W.,Saunders,M.A. SNOPT: An SQP algorithm for large scale constrained optimization. SIAM Journal on Optimization, 2002, 12: 979-1006

[71] Schulz,V.H.,Bock,H.G. Partially reduced SQP methods for large scale nonlinear optimization problems. Nonlinear analysis, Theory, Methods and Applications, 1997, 30: 4723-4734

[72] Facchinei,F.,Liuzzi,G., Lucidi,S. A truncated Newton method for the solution of large scale inequality constrained minimization problems. Computational Optimization and Applications, 2003, 25: 85-122

[73] Dennis,J.E.,Heinkenschloss,M.,Vicente,L.N. Trust region interior point SQP algorithm for a class of nonlinear programming problems. SIAM Journal on Control Optimization, 1998, 36: 1750-1794

[74] Heinkenschloss,M.,Vicente,L.N. Analysis of inexact trust region SQP algorithm. SIAM Journal on Optimization, 2001, 12: 283-302

[75] Psiaki,M.L.,Park,K.Augmented lagrangian nonlinear programming algorithm that uses SQP and trust region technique. Journal of Optimization Theory and Application, 1995, 86: 311-325

[76] 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: 273-282

[77] M.heinkenschloss, M.Ulbrich, S.Ulbrich. Superlinear and Quadratic convergence of affine scaling interior point Newton methods for problems with simple bounds without strict complementarity assumpotion. Mathematical Programming, 1999, 86: 615-635

[78] Durazzi,C.,On the Newton interior point method for nonlinear programming problems. Journal of Optimization Theory and Applications, 2000, 104: 73-90

[79] Yamashita,H.,Yabe,H. Superlinear and Quadratic convergence of some primal dual interior point methods for constrained optimization. Mathematical Programming, 1996, 75: 377-397

[80] Qi,L.Q.,Yang,Y.F. A globally and superlinearly convergent SQP algorithm for nonlinear constrained optimization. Journal of Global Optimization, 2001, 21: 157-184

[81] Wiest,E.J.,Polak,E. A generalized quadratic programming based phase I-phase II method for inequality constrained optimization. Applied Mathematica and Optimization, 1992, 26: 223-252

[82] Qi,L.Q., Wei,Z.X. On the constant positive linear dependence condition and its application to SQP methods. SIAM Journal on Optimization, 2000, 10: 963-981

[83] Birge,J.R., Qi,L., Wei,Z. A variant of the Topkis-Veinott method for solving inequality constrained optimization problems. Applied Mathematics and Optimization, 2000, 41: 309-330

[84] Lawrence,C.T.,Tits,A.L. A computationally efficient feasible sequential quadratic programming algorithm. SIAM Jouanal on Optimization, 2001, 11: 1092-1118

[85] Kostreva,M.M., Chen,X. A superlinearly convergent method of feasible directions. Applied Mathematics and Computation, 2000, 116: 245-255

[86] Facchinei.F.,Fischer.A. and Kanzow.C. On the accurate identification of active constraints. SIAM Journal on Optimization, 1999, 9: 14-32

[87] 高自友, 贺国平, 吴方. 任意初始点下的序列线性方程组方法. 中国科学 A 辑, 1997, 27: 24-33

[88] Panier,E.R.,Tits,A.L. A superlinearly convergent feasible method for the solution of inequality constrained optimization problems. SIAM Journal on Control and Optimization, 1987, 25: 934-950

[89] 简金宝. 最优化两个拓广的 SQP 和 SSLE 算法模型及其超线性和二次收敛性. 高校应用数学学报, 2001, 16: 435-444

[90] Mayne,D.Q.,Polak,E.,Sangiovanni-vincenteli,A. Computer-aided design via optimization: A review. Automatics, 1982, 18: 147-154

[91] Polak,E.,Mayne,D.Q,Stimler,D.M. Control system design via semi-infinite optimization: A review. Proceedings IEEE, 1984, 72: 1777-1794

[92] Luskan,L. A compact variable metric algorithm for nonlinear minimax approximation. Computing, 1986: 36, 355-373

[93] Polak,E.,Mayne,D.Q.,Higgins,J.E. Superlinearly convergent algorithm for minmax problems. Journal of Optimization Theory Application, 1991, 69: 407-439

[94] Polyak,R.A. Smooth optimization methods for minimax problems. SIAM Journal on Control and Optimization, 1988, 26: 1274-1286

[95] Zhou,J.L.,Tits,A.L. Nonmonotone line search for minimax problems. Journal of Optimization Theory and Application, 1993, 76: 455-476

[96] Zhou,J.L.,Tits,A.L. An SQP algorithm for finely discretized continous minimax problems and other minimax problems with many objetive functions. SIAM Journal on Numerical Analysis, 1996, 36: 461-487

[97] 薛毅. 求解 minimax 优化问题的 SQP 方法. 系统科学与数学, 2002, 22: 355-364

[98] 朱志斌，张可村. Minimax 问题的一个超线性收敛的 SQP 算法. 数值计算与计算机应用, 2005, 3: 161-176

[99] Yu,Y.H.,Gao,L. Nonmonotone line search algorithm for constrained minimax problems. Journal of Optimization Theory and Application, 2002, 115: 419-446

[100] Vardi,A. New minimax algorithm. Journal of Optimization Theory and Application, 1992, 75: 613-634

[101] Panier,E.R.,Tits,A.L. On combining feasibility descent and superlinearly convergence in inequality constrained optimization. Mathematical Programming, 1993, 59: 261–276

[102] Zhu,Zh.b. A simple feasible SQP algorithm for inequality constrained optimization. Applied Mathematics and Computation, 2006, 182: 987–998

[103] Panier,E.R.,Tits,A.L., Herskovits,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

[104] Gao,Z.Y., He,G.P., Wu,F. Sequential systems of linear equations with general constraints. Journal of Optimization Theory and Applications, 1997, 27: 24-33

[105] Qi,H.D., Qi,Q.L. A new QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization. SIAM Journal On Optimization, 2000, 11: 113-132

[106] Yang, Y.F., Li,D.H., Qi.L. A feasible sequential linear equation method for inequality constrained optimization. SIAM Journal on Optimization, 2002, 13: 1222-1244

[107] Facchinei,F., Lucidi,S. Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minimization problems. Journal of Optimization Theory and Applications, 1995, 85: 256-289

[108] Lucidi,S. New results on a continuously differentiable penalty function. SIAM Journal on Optimization, 1992, 2: 558-574

[109] Zhu,Zh.b. An interior point type QP-free algorithm with superlinear convergence for inequality constrained optimization. Applied Mathematical Modelling, 2007, 31: 1201-1212

[110] Hock,W.,Schittkowski,K. Test Examples for Nonlinear Programming Codes, Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, Berlin Heidelberg New York, 1981.

[111] Hock,W.,Schittkowski,K. More Test Examples for Nonlinear Programming Codes. Springer-Verlag, Berlin Heidelberg New York, 1987.

附录 攻读学位期间所发表的学术论文目录

[1] Qingjie Hu, Yunhai Xiao and Yu Chen, An Active Set Sequential Quadratic Programming Algorithm for Nonlinear Optimization, Bulletin of the Australian Mathematical Society, 74 (2006), pp. 69-84.

[2] Qingjie Hu, Jinbao Jian, Haiyan Zheng and Chunming Tang, Semilocal E-convexity and Semilocal E-convex Programming, Bulletin of the Australian Mathematical Society, 75 (2007), pp. 59-74.

[3] Qingjie Hu and Yu Chen, Nondifferentiable Multiobjective Programming under Type I Semi-d-univexity, Advances in Theoretical and Applied Mathematics, 2 (2007), pp.65-75.

[4] Jinbao Jian, Qingjie Hu, Chunming Tang and Haiyan Zheng, A Sequential Quadratically Constrained Quadratic Programming Method of Feasible Directions, Applied Mathematics and Optimization, 56 (2007), pp. 343-363.

[5] Qingjie Hu and Juzhou Hu, A Sequential Quadratics Programming Algorithm for Nonlinear Minimax Problems, Bulletin of the Australian Mathematical Society, 76 (2007), pp. 353-368.

[6] Qingjie Hu, Wanyou Chen and Yunhai Xiao, An Improved Active Set Feasible SQP Algorithm for the Solution of Inequality Constrained Optimization Problems, Journal of Computational Analysis and Applications (已接受).

致　谢

在本文即将完稿之际，回想三年来的求学生涯，感慨万千．本文得以顺利完成，除了本人的刻苦钻研、勤奋努力外，还得到许多老师，同学及亲友的帮助与关心．在此，谨向各位致以我真诚的谢意．

首先感谢我的导师李董辉教授．三年来，他给与我无微不至的关怀，学习上给与我耐心细致的指导．从论文的选题到具体的研究工作，都得到了李老师的悉心指导与反复修改，并提出了许多富有创意的建议．李老师不但传授给我专业知识，同时教会了我许多做人的道理．李老师渊博的知识、精辟的见解和开拓创新的精神使我受益终身．在此，我向李老师致以最崇高的敬意．

感谢湖南大学数学与计量经济学院曾金平教授三年来对我的关心与鼓励．

感谢湖南大学数学与计量经济学院的领导和老师三年来对我的关心与帮助．

感谢各位师兄弟 (姐妹) 以及其他同学的关心与帮助．在这充满团结、友爱的气氛里，他们的关心与帮助时时陪伴着我，让我的生活丰富多彩．

感谢广西大学简金宝教授对我的关心与鼓励．

感谢湖南商学院信息学院的领导和老师对我的关心与帮助．

最后，我要深深感谢我的家人及亲戚对我的支持与理解．特别要感谢我的妻子陈玉女士和我的儿子胡博睿，感谢他们对我的理解、支持和鼓励，感谢他们对我的付出．

谨已此文献给所有关心和支持我的人们．