DOI: 10.13656/j.cnki.gxkx.2002.02.002
广西科学 Guangxi Sciences 2002, 9 (2): 84~ 90

Research Advance on the Superlinearly Convergent Feasible Methods I SQP Type Methods*
超线性收敛可行方法的研究进展I ：SQP类方法

Jian Jinbao Li Jianling Mo Xingde
简金宝** 黎健玲 莫兴德

(College of Math. and Physics, Guangxi Univ., 100 Daxuelu, Nanning, Guangxi, 530004, China)
(广西大学理学院 南宁市大学路 100号 530004)

Abstract The superlinearly convergent feasible methods for nonlinearly constrained optimization problems are an important research subject in both theory and practice. Recently the researches are conducted in a deepgoing and systematic way, and some new results are obtained, such as SQP type, sequential systems of linear equations type and explicit search directions type. The recent results about the feasible SQP type methods are introduced. The superlinear convergence of the previous algorithms is unitedly analyzed by a new SQP extension model.

Key words nonlinear constraint, optimization, feasible methods, superlinear convergence.

摘要 非线性约束最优化的超线性收敛可行方法是一个具有重要理论意义和实用价值的研究方向。最近该方向得到广泛而深入的研究,获得一系列新的研究成果,如 SQP类方法、序列线性方程组类方法和显式搜索方向类方法。本文介绍可行 SQP类方法的最近研究成果,并采用一个拓广的 SQP模型统一分析各种算法的超线性收敛性。

关键词 非线性约束 最优化 可行方法 超线性收敛性
中图法分类号 O221.2

The most basic feature and claim about an algorithm for optimal problems are that the algorithm should have convergence in theory, i. e. approximate the solution infinitely. However, the application of high technologies in the economic activities raises a higher demand on algorithms in rate of convergence. The researches on superlinearly convergent algorithms have not only strongly applied value, but also important theoretically significance to the development of the subject itself. Since most of the previous superlinearly convergent algorithms made use of the penalty functions as the effect functions, the iterative points and approximate solutions are infeasible. But some practical problems usually strictly demand that the iterative points and approximate solutions are feasible or partially feasible. In order to solve this contradiction, Panier and Tits[1] first put forward a superlinearly convergent feasible method in 1987, and then this subject was paid attention gradually. Many scholars[2~12] have made deepgoing systematically researches. According to the technique of producing the master search directions, these methods may be divided into three types: SQP type[1~10], sequential systems of linear equations type[11~15] and explicit search directions type[16,17].

In order to reveal the internal relations of these algorithms in theory and make the scholars and readers be interested in this subject and know more about the research situation and recent results, we are going to introduce these research results and progresses in two parts. The feasible methods of SQP type will be introduced in part I, and the methods of QP-free type in part II (a separated paper).

2001-12-04收稿
* Supported by the National Natural Science Fundation (N. 19801009) and Natural Science Fundation of Guangxi Province (N. 9811023& 9912027).
** Email jianjb@gxu.edu.cn

1 A SQP extension model and its convergence rate

Generally, a nonlinearly optimal model with general constraints can be expressed as follows.

$$\min\{f_0(x)\mid f_i(x)\leqslant 0, i\in I; f_j(x)=0, j\in J\}, \tag{1.1}$$

where $I=\{1,2,\cdots,m\}$, $J=\{m+1, m+2,\cdots,m+l\}$. In order to overcome effectively the Maratos effect, and adapt various initial points, many authors have made correction, improvement and extension[2-6] to the previous quadratic programming (QP). To describe and analyze unitedly these methods and their rate of convergence, a more wide-ranging kind of QP model is introduced in this section. For a current iterative point x^k, choose index sets $T_k, L_k^-, L_k^0 \subseteq L \triangleq I\cup J$, parameter c_k, and function

$$F_{c_k}=f_0(x)-c_k\sum_{j\in T_k} f_j(x). \tag{1.2}$$

Let B_k be a given approximate matrix, we construct a more wide-ranging QP as follows

$$\min \nabla F_{c_k}(x^k)^T d+\frac{1}{2}d^T B_k d \tag{1.3a}$$

$$\text{s.t.}\ f_i(x^k)+\nabla f_i(x^k)^T d\leqslant \eta_i^k, i\in L_k^-, \tag{1.4b}$$

$$f_j(x)+\nabla f_j(x^k)^T d=\eta_j^k, j\in L_k^0. \tag{1.5c}$$

Based on QP(1.3), combining the correction direction of overcoming the Maratos effect, and ensuring the feasible descent as well as the global convergence, we give a more general SQP type frame of solving Problem (1.1), and call it a sequential quadratic programming extension model (SQPEM).

Algorithm 1.1— SQPEM[18]

Step 1 Choose an initial point x^1 to satisfy certain conditions (as $x^1\in R^n$ or satisfying certain feasibility), initial matrix B_1, $k:=1$.

Step 2 For the current iterative point x^k, determine T_k, L_k^-, L_k^o and parameter c_k, η_j^k, find a $K-T$ point d_o^k of QP(1.3).

Step 3 Produce a correcting direction d_1^k by certain technique.

Step 4 Let $x^{k+1}=x^k+d_0^k+d_1^k$, produce a new matrix B_{k+1}, let $k:=k+1$, and go back to Step 2.

Assumption 1.2

(i) Functions $f_o(x), f_j(x)\in C^2, j\in L$;

(ii) If the accumulation point x^* of the sequence $\{x^k\}$ produced by Algorithm 1.1 is a $K-T$ point of Problem(1.1), then the second-order sufficent conditions for optimality hold at $K-T$ pair (x^*, u^*);

(iii) The strict complementarity condition of Problem(1.1) holds at $K-T$ pair (x^*, u^*).

Assumption 1.3

(i) The total sequence $\{x^k\}$ generated by Algorithm 1.1 converges to a $K-T$ point x^* of Problem (1.1), and the non-degenerate condition holds at x^*;

(ii) The following relations are satisfied

$$\lim_{k\to\infty} d_0^k=0, \lim_{k\to\infty}\eta_j^k=0(\forall j), \|d_1^k\|=o(\|d_0^k\|^2). \tag{1.6}$$

(iii) Let $L_k=\{j\in L_k^-\cup L_k^o\mid f_j(x^k)+\nabla f_j(x^k)^t d_o^k=\eta_j^k\}$, and $L_*=\{j\in L\mid f_j(x^*)=0\}$, suppose that the following relations hold for k large enough,

$$L_k\equiv L_*, T_k\subseteq L_*. \tag{1.7}$$

Let u^* be a $K-T$ multiplier of Problem (1.1) corresponding to x^*, denote

$$g_j(x)=\nabla f_j(x), j\in L; N(x)=(g_j(x), j\in L), A_k=(g_j(x^k), j\in L_*), \tag{1.8}$$

$$R_k=E_n-A_k(A_k^T A_k)^{-1}A_k^T,$$

$$B(x^*, u^*)=\nabla^2 f_0(x^*)+\sum_{j\in L} u_j^*\nabla^2 f_j(x^*). \tag{1.9}$$

Theorem 1.4[18] Suppose that Assumption 1.2(i, ii) and Assumption 1.3 hold, then (i) If

$$\|R_k(B_k-B(x^*, u^*))R_k d_0^k\|=o(\|d_0^k\|), \tag{1.10}$$

$$\|H^k=(\eta_j^k, j\in L_*)\|=o(\|d_0^k\|), \tag{1.11}$$

then Algorithm 1.1 is two-step superlinearly convergent, i.e.

$$\|x^{k+1}-x^*\|=o(\|x^{k-1}-x^*\|). \tag{1.12}$$

(ii) Algorithm 1.1 is one-step superlinearly convergent, i.e

$$\|x^{k+1}-x^*\|=o(\|x^k-x^*\|), \tag{1.13}$$

if and only if Relation(1.9) and

$$\|R_k(B_k-B(x^*, u^*))d_0^k\|=o(\|d_0^k\|), \tag{1.14}$$

hold true.

In many practical algorithms, Assumption 1.2 is usually supposed true, and Assumption 1.3 is satisfied by choosing suitable correcting direction d_1^k and line search technique.

2 The feasible methods with feasible initial points

The main idea of various feasible methods is how to construct the correcting direction d_1^k and select effect function as well as search method for determining step size, so that Algorithm 1.1 possesses global convergence and so on, meanwhile, the step size 1 can be accepted in a neighbourhood of x^*. In this paper, except for special explanation, it is supposed that $J=T_k=L_k^o$

$= \varnothing, e = (1, 1, \cdots, 1)^T$. The basic assumption of Problem(1.1) is given below.

Assumption 2.1 Some non-degenerate constraint conditions(NDCC) for Problem (1.1) hold true.

Earlier feasible SQP method was given by Panier and Tits[1] in 1987, which can be described below.

Algorithm 2.2— Panier-Tits method[1]

Step 1 Select parameters $M > 0, \alpha \in (0, \frac{1}{2}), \beta \in (0,1), \nu > 2, \omega > 2, \tau \in (2,3)$. Let $\mu_i^1 \equiv 0, \forall i \in I; x^1 \in X \triangleq \{x \in R^n | f_i(x) \leqslant 0, i \in I\}$, $B_1 \in R^{n \times n}$, $k = 1$.

Step 2 Let $T_k = L_k^o \equiv \varnothing, L_k^- \equiv I$, calculate a $K-T$ point d_o^k of QP(1.3). If $d_o^k = 0$, stop, and x^k is a $K-T$ point of Problem(1.1); if QP(1.3) has no $K-T$ point or $\|d_o^k\| > M$ or $\|B_k d_o^k\| > \|d_o^k\|^{\frac{1}{2}}$, go to Step 4.

Step 3 Compute d_1^k.

(i) Solve the quadratic programming given below to obtain a $K-T$ point d^k,

$$\min g_0(x^k)^T d + \frac{1}{2} d^T B_k d,$$

$$\text{s.t. } f_i(x^k) + g_i(x^k)^T d \leqslant -\|d_0^k\|^\nu, \forall i \in I.$$

If d^k exists, and

$\|d^k\| \leqslant M, \theta_k = g_0(x^k)^T d^k \leqslant \min\{-\|d_0^k\|^\omega, -\|d^k\|^\omega\}$,

then go to (ii), othwerwise go to Step 4.

(ii) Solve the following quadratic programming

$$\min \{\frac{1}{2}\|d\|^{\frac{1}{2}} | f_j(x^k + d^k) + g_j(x^k)^T d = -\|d_0^k\|^\tau, j \in I_k\},$$

to get a solution $\tilde{d}^k$, where I_k is the active constraint set of quadratic programming in (i). If $\tilde{d}^k$ exists, and $\|\tilde{d}^k\| \leqslant \|d^k\|$, go to Step 5; otherwise, go to Step 4.

Step 4 Generate a "first-order" feasible descent direction q^k by an auxiliary method, let $d^k = q^k, \tilde{d}^k = 0$, $\theta_k = g_0(x^k)^T d^k$.

Step 5 Line search. Compute the first number λ_k of sequence $\{1, \beta, \beta^2, \cdots\}$ satisfying

$$f_0(x^k + \lambda d^k + \lambda^2 \tilde{d}^k) \leqslant f_0(x^k) + \alpha\lambda\theta_k, f_i(x^k + \lambda d^k + \lambda^2 \tilde{d}^k) \leqslant 0, \forall i \in I.$$

Step 6 Let $x^{k+1} = x^k + \lambda_k d^k + \lambda_k^2 \tilde{d}^k, k := k+1$, and generate a new matrix B_{k+1} by some methods, then go back to Step 2.

In addition to ensure the global convergence of the sequence points generated by the auxiliary method, the "first-order" feasible descent direction guaranteed Algorithm 2.2 to possess strong convergence under suitable assumptions. Its characters were given in References[1~3].

Algorithm 2.3— Jian method[2]

It is obvious that Algorithm 2.2 needs to solve 3 ~ 4 QPs at each iteration(According to Reference[1], one QP needs to be solved in order to get the "first-order" feasible descent direction q^k), the amount of computation is too much and the construction is complex. Jian J B[2] had made a modification for Algorithm 2.2 in 1992.

Step 1 Select parameters $\alpha \in (0, \frac{1}{2}), \beta \in (0, 1), \omega > 2, \tau \in (2,3)$. Let $\mu_i^1 \equiv 0, \forall i \in I; x^1 \in X \triangleq \{x \in R^n | f_i(x) \leqslant 0\}, B_1 \in R^{n \times n}, k = 1$.

Step 2 Let $T_k = L_k^o \equiv \varnothing, L_k^- \equiv I$, calculate a $K-T$ point d_o^k of QP(1.3). If $d_o^k = 0$, stop, and x^k is a $K-T$ point of Problem(1.1); if QP(1.3) has no $K-T$ point or $\|B_k d_0^k\| > \|d_0^k\|^{\frac{1}{2}}$, go to Step 5.

Step 3 Compute revised direction d_1^k: compute

$I_k = \{j \in I | f_j(x^k) + g_j(x)^T d_0^k = 0\}, A_k = (g_j(x^k)), j \in I_k)$,

$$d_1^k = \begin{cases} -A_k(A_k^T A_k)^{-1}(\|d_0^k\|^\tau + \tilde{f}^k), & \text{if } \det(A_k^T A_k) \neq 0; \\ 0, & \text{if } \det(A_k^T A_k) = 0, \end{cases}$$

$\tilde{f}^k = (f_j(x^k + d_0^k), j \in I_k)$.

Step 4 Seeking search. If

$\theta_k = g_0(x^k)^T d_0^k \leqslant \min\{-\|d_0^k\|^\omega, -\|d_0^k + d_1^k\|^\omega\}$,

$f_0(x^k + d_0^k + d_1^k) \leqslant f_0(x^k) + \alpha\theta_k, f_j(x^k + d_0^k + d_1^k) \leqslant 0, \forall j \in I$,

then let $\lambda_k = 1$, go to Step 7; else go to Step 5.

Step 5 It is similar to Step 4 of Algorithm 2.2, but let $d_0^k = q^k, d_1^k = 0$.

Step 6 Line search. Compute the first number λ_k of the sequence $\{1, \beta, \beta^2, \cdots\}$ satisfying

$$f_0(x^k + \lambda d_0^k) \leqslant f_0(x^k) + \alpha\lambda g_0(x^k)^T d_0^k, f_j(x^k + \lambda d_0^k) \leqslant 0, \forall j \in I.$$

Step 7 Let $x^{k+1} = x^k + \lambda_k d_0^k + \lambda_k d_1^k$, the others are the same as Step 6 of Algorithm 2.2.

Algorithm 2.4— Gao-Wu method[8]

In 1995, Gao Z Y and Wu F modified the method in Reference[1] by the ε-pivoting operation, and they used the new updating formula given by Pantoja-Mayne[19] in 1991 to produce matrix B_{k+1}, thus Relation(1.14) holds true, and the one-step superlinear convergence was attainable. The main steps are given below.

Steps 1 and 2 They are the same as Steps 1, 2 of Algorithm 2.2.

Step 3 Compute d_1^k. Find parameter $\varepsilon_k > 0$ by the

ε-pivoting operation such that det$(A_k^T A_k) \geq \varepsilon$, where $A_k = (g_j(x^k), j \in J_{\varepsilon_k})$, $J_{\varepsilon_k} = \{j \in I \mid 0 \leq -f_j(x^k) \leq \varepsilon_k\}$.

(i) Compute

$$d^k = d_0^k + A_k(A_k^T A_k)^{-1}(-\|d_0^k\|^{\nu} e).$$

If

$$\theta_k = g_0(x^k)^T d^k > \min\{-\|d_0^k\|^{\sigma}, -\|d^k\|^{\sigma}\},$$

go to Step 4.

(ii) Compute $I_k = \{j \in I \mid f_j(x^k) + g_j(x^k)^T d_0^k = 0\}$ and

$$\tilde{d}^k = A_k(A_k^T A_k)^{-1}(-\|d_0^k\|^{\tau} e - \tilde{f}(x^k + d^k)), \tilde{f}_j(x^k + d^k) = \begin{cases} f_j(x^k + d^k), j \in I_k, \\ 0, j \in J_{\varepsilon_k} \setminus I_k, \end{cases}$$

if $\|\tilde{d}^k\| \geq \|d^k\|$, let $\tilde{d}^k = 0$. Go to Step 5.

Steps 4, 5 and 6 They are the same as Steps 4, 5, 6 of Algorithm 2.2.

Now we discuss unitedly the convergence of Algorithms 2.2～2.4 as follows.

Assumption 2.5 The iteration sequence $\{x^k\}$ possesses an accumlation point x^*.

Assumption 2.6 $\lim_{k \to \infty} B_k = B_*$, B_* is positive definite in the subspace $\Omega = \{d \mid g_j(x^*)^T d = 0, j \in L_*\}$, and Relation (1.14) holds true.

Theorem 2.7 (i) Suppose that Assumptions 1.2(i) and 2.1 hold. Then Algorithms 2.2～2.4 are all globally convergent, i.e. each accumlation point x^* of iterative sequence $\{x^k\}$ is a K-T point of Problem (1.1).

(ii) If Assumptions 2.1, 2.5 and 1.2 (i, ii) hold, then Algorithms 2.2～2.4 are all strongly convergent, i.e. $\lim_{k \to \infty} x^k = x^*$.

(iii) If Assumptions 1.2, 2.1, 2.5 and 2.6 all hold, then Assumption 1.3 holds too. Furthermore, Algorithms 2.2～2.4 are all one-step superlinearly convergent.

Proof References [1, 2, 8] proved respectively that the step size of all Algorithms 2.2～2.4 equal one when k large enough, and the auxiliary "first-order" direction q^k needs not to be calculated. In Algorithms 2.2, 2.4, the iteration formulae are given as $x^{k+1} = x^k + d^k + \tilde{d}^k = x^k + d_0^k + (d^k - d_0^k + \tilde{d}^k) \overset{\text{def}}{=} x^k + d_0^k + d_1^k$, and it's not difficult to know that $\|d_1^k\| = o(\|d_0^k\|^2)$ from the corresponding proof in References [1, 2, 8].

On the other hand, the other requirements of Assumption 1.3 had been proved in the relevant references. So we can confirm Theorem 2.7 from Theorem 1.4.

In 1996, Gao[10] also gave a one-step superlinearly convergent feasible method by using two systems of linear equations to produce the updating direction d_1^k.

Remark 1 In the relevant references, such as References [1, 2], under Relation (1.10), only one got two-step superlinear convergence, but with the appearance of new updating formulae satisfying Relation (1.14) (such as Reference [19]), the convergence rate of these algorithms can be one-step superlinear.

Algorithm 2.8— Gao-Wu method[9]

Because matrix B_k needn't to be positive definite in Algorithms 2.2～2.4, the auxiliary direction needs to be computed in a finite number of iterations. In 1997, under the assumption that B_k is uniformly positive definite, Gao and Wu[9] presented a curve search feasible method without the auxiliary direction, it is given as follows.

Step 1 Select $x^1 \in X$, a positive definite matrix B_1, parameters $\alpha \in (0, 1)$, $\beta \in (0, \frac{1}{2})$, $\tau \in (2, 3)$; and functions $\lambda_j^k \equiv 0, \forall j \in I$, let $k = 1$.

Step 2 Let $T_k = L_k^o \equiv \emptyset$, $L_k^- \equiv I$, solve QP(1.3) to get a K-T point d_0^k, let I_k be its active constraint set. If $d_0^k = 0$, then x^k is a K-T point of Problem (1.1), stop.

Step 3 Compute direction.

$d^k = d_0^k - \delta_k Q_k^T e$, $\tilde{d}^k = -Q_k^T(\|d_0^k\|^{\tau} e + \tilde{f}^k)$, where $\delta_k = \dfrac{\|d_0^k\|(d_0^k)^T B_k d_0^k}{2\|d_0^k\| \cdot |e^T \pi^k| + 1}$, $\pi^k = -Q_k g_0(x^k)$, $\tilde{f}_j^k = \begin{cases} f_j(x^k + d^k), j \in I_k, \\ 0, j \in I \setminus I_k, \end{cases}$

$Q_k = (N(x^k)^T N(x^k) + D_k)^{-1} N(x^k)^T$, $D_k = \text{diag}(-f_j(x^k), j \in I)$.

If $\|\tilde{d}^k\| > \|d^k\|$, let $\tilde{d}^k = 0$.

Step 4 Line search. It is the same as Step 5 of Algorithm 2.2.

Step 5 Generate a new positive definite matrix B_{k+1}, others are similar to Step 6 of Algorithm 2.2.

The convergence and convergence rate of Algorithm 2.8 can be seen from Theorem 3.6 afterwards. In 1999, Jian-Xue[7] presented a class of feasible methods containning some of parameters for solving nonlinear inequality constrained optimization problems. At each iteration, the algorithm generates the search directions by solving only one QP in small scale. The total algorithm can be seen in Reference [7].

3 Subfeasible and strongly subfeasible methods

Though the feasible methods[1,2,7～9] overcome the shortcomings of penalty function methods that the iterative points are infeasible, the calculation of a feasible initial point increases the amount of computation of these algorithms since a feasible point is usually obtained by solving an auxiliary optimization problem. To overcome this new problem, Jian J B[20] set up two new kinds of methods, i. e. subfeasible (direction) methods and strongly subfeasible (direction) methods with arbitrary initial point. Furthermore, Jian studied superlinearly convergent algorithms in these ideas in References [4~6]. These algorithms are given below.

Algorithm 3.1— Jian subfeasible method[5]

Step 1 Select parameter $k \in [0,1]$, arbitrary initial point $x^1 \in R^n$, functions $\mathfrak{h}_j^k = Q(x^k) = \max\{0, f_j(x^k), j \in I\}, j \in I$. Others are similar to Step 1 of Algorithm 2.3.

Step 2 It is similar to Step 2 of Algorithm 2.3, but it doesn't stop until $\|d_0^k\| + Q(x^k) = 0$

Step 3 Compute updating direction d_1^k. Let

$I_k = \{j \in I | f_j(x^k) + g_j(x^k)^T d_0^k = \mathfrak{h}_j^k = Q(x^k)\}$, $A_k = (g_j(x^k), j \in I_k)$,

if $\det(A_k^T A_k) = 0$, then go to Step 5, else compute

$d_1^k = -A_k(A_k^T A_k)^{-1}\{(\|d_0^k\|^f - Q(x^k) + \|d_0^k\| Q(x^k))e + \mathcal{J}^k\}$,

$\_^k = (\_^k_j, j \in I_k) = -(A_k^T A_k)^{-1} A_k^T (g_0(x^k) + B_k d_0^k)$, $r_k = \sum_{j \in I_k} \_^k_j$.

Step 4 Do seeking search. Let $s_k = 0$, if $Q(x^k) = 0$; $s_k = 1$, If $Q(x^k) > 0$. if

$(1 - s_k) g_0(x^k)^T d_0^k \leqslant (1 - s_k)\min\{-\|d_0^k\|^W, -\|d_0^k + d_1^k\|^W\}$, $f_0(x^k + d_0^k + d_1^k) \leqslant f_0(x^k) + r_k\|d_0^k\|^k Q(x^k) + \mathrm{T} g_0(x^k)^T d_0^k$, $f_j(x^k + d_0^k + d_1^k) - Q(x^k) \leqslant -\mathrm{T}(\|d_0^k\|^f + Q(x^k)^W)$, $\forall j \in I$,

then let $\lambda_k = 1$, go to Step 7; Otherwise, go to Step 5.

Step 5 Calculate a solution (θ_k, q^k) of the following linear program problem,

$\min \theta$,

s.t. $g_0(x^k)^T q \leqslant Q(x^k) + \theta$, $f_j(x^k) + g_j(x^k)^T q \leqslant Q(x^k) + \theta$, $j \in I$, $-1 \leqslant q_j \leqslant 1$, $j = 1, 2, \cdots, n$.

If $\theta_k = 0$, then x^k is a K-T point of Problem (1.1), stop. Otherwise, let $d_0^k = q^k$, $d_1^k = 0$, go to Step 6.

Step 6 Do line search. Find the first number λ_k of the sequence $\{1, \mathrm{U}, \mathrm{U}^2, \cdots\}$ satisfying

$(1 - s_k)\{f_0(x^k + \lambda d_0^k) - f_0(x^k) - \lambda \mathrm{T}\theta_k\} \leqslant 0$,

$f_j(x^k + \lambda d_0^k) - Q(x^k) \leqslant \lambda \mathrm{T} s_k \theta_k$, $\forall j \in I$.

Step 7 It is the same as Step 7 of Algorithm 2.3.

The non-degenerate Assumption 2.1 for Algorithm 3.1 is that the vectors $\{g_j(x) j: f_j(x) = Q(x)\}$ are independent for each $x \in R^n$. Note that $Q(x^k) = 0$, if and only if $x^k \in X$, and in view of Steps 4 and 6, we know that the iterative point $x^k \in X(\forall k \geqslant t)$ whenever some $x^t \in X$.

Theorem 3.2 (i) Let Assumptions 1.2 (i) and 2.1 be satisfied. Then Algorithm 3.1 is globally convergent.

(ii) Suppose that $\lim_{k \to \infty} x^k = x^*$, $Q(x^k) = o(\|d_0^k\|)$, and Assumptions 1.2, 2.1 and 2.6 all hold true. Then Algorithm 3.1 does not go to Steps 5 and 6 for k large enough, the step size $\lambda_k \equiv 1$, and it is one-step superlinearly convergent.

Under the assumption that matric set $\{B_k\}$ are uniformly positive definite, References [4, 6] had moved away the auxiliary direction q^k to make the algorithm more perfect and independent in theory.

Algorithm 3.3— Jian subfeasible method without auxiliary direction[4]

For $x \in R^n$, define $Q(x) = \max\{0, f_j(x), j \in I\}$. The non-degenerate Assumption 2.1 is that vectors $\{g_j(x): f_j(x) = Q(x)\}$ are linearly independent. Let

$D(x) = \mathrm{diag}(Q(x) - f_j(x), j \in I)$, $Q(x) = (N(x)^T N(x) + D(x))^{-1} N(x)^T$.

Step 1 Choose an arbitrary initial point $x^1 \in R^n$, a symmetric positive definite matrix B_1, parameters $\mathrm{U}, \mathrm{X} \in (0,1)$, $\mathrm{T} \in (0, 0.5)$, $e > 0$; functions $\mathfrak{h}_j^k \equiv Q(x^k)$, $\forall j \in I$. Let $k = 1$.

Step 2 Let $T_k = L_k^0 = \emptyset$, $L_k^- = I$. Solve a K-T point d_o^k of QP(1.3), if $\|d_0^k\| + Q(x^k) = 0$, then x^k is a K-T point of Problem (1.1), stop.

Step 3 Denote the active constraint set of QP (1.3) at d_0^k by I_k, calculate

$$d^k = d_0^k - \mathfrak{f}_k Q(x^k)^T e, \tilde{d}^k = \begin{cases} 0, \text{if} \| -Q(x^k)^T(\tilde{\mathcal{J}}^k - Q(x^k)\tilde{e}^k)\| > \|d_0^k\|, \\ -Q(x^k)^T(\tilde{\mathcal{J}}^k - Q(x^k)\tilde{e}^k), \text{ others.} \end{cases}$$

$$\mathfrak{d}_k = \min\{\frac{1}{2}, \|d_0^k\|^X\}, \mathfrak{f}_k = \frac{\mathfrak{d}_k (d_0^k)^T B_k d_0^k + Q(x^k)^e}{1 + \|Q(x^k)^T g_0(x^k)\|}$$

$$\tilde{e}_j^k = \begin{cases} 1, j \in I_k, \\ 0, j \in I \backslash I_k, \end{cases} \quad \tilde{\mathcal{J}}_j^k = \begin{cases} f_j(x^k + d_0^k), j \in I_k, \\ 0, j \in I \backslash I_k, \end{cases}$$

Step 4 Find the first number λ_k of the sequence $\{1, \mathrm{U}, \mathrm{U}^2, \cdots\}$ satisfying

$f_0(x^k + \lambda d^k + \lambda^2 \tilde{d}^k) \leqslant f_0(x^k) + \mathrm{T}\lambda g_0(x^k)^T d^k + \lambda Q(x^k)^W$,

$f_j(x^k + \lambda d^k + \lambda^2 \tilde{d}^k) - Q(x^k) \leqslant -\mathrm{T}\lambda \mathfrak{f}_k$, $j \in I$.

Step 5 Let $x^{k+1} = x^k + \lambda_k d^k + \lambda_k^2 \tilde{d}^k$, and generate a new symmetric positive definite matrix B_{k+1}. Let

$k := k+1$, go to Step 2.

Algorithm 3.4 Jian strongly subfeasible method[6]

For $x \in R^n$, we denote

$I_-(x) = \{j \in I | f_j(x) \leqslant 0\}$, $I_+(x) = I \backslash I_-(x)$, $\varphi(x) = \max\{0, f_j(x), j \in I\}$,

$I(x) = \{j \in I | j \in I_-(x), f_j(x) = 0;\ \text{or}\ j \in I_+(x), f_j(x) = \varphi(x)\}$.

The non-degenerate Assumption 2.1 is that vectors $g_j(x) | j \in I(x)\}$ are linearly independent for any $x \in R^n$. Jian[6] in 1996 further used the idea of strongly subfeasible (direction) methods to study SQP type algorithm, such that the feasibility is monotonously non-decreasing during the process of iteration, i.e., $I_-(x^k) \subseteq I_-(x^{k+1})$.

Step 1 Choose an arbitrary initial point $x^1 \in R^n$, a symmetric positive definite matrix B_1, parameters $c, \chi \in (0,1)$, $e > 0$. Let $k = 1$.

Step 2 Let $T_k = L_k^0 = \varnothing$, $L_k^- = I$, $I_-^k = I_-(x^k)$, $I_+^k = I_+(x^k)$, and functions

$\mathrm{H}_j^k = \varphi(x^k)$, if $j \in I_+^k$, $\mathrm{H}_j^k = 0$, if $j \in I_-^k$.

Calculate a K-T point d_0^k of QP(1.3), if $\|d_0^k\| + \varphi(x^k) = 0$, then x^k is a K-T point of Problem (1.1), stop.

Step 3 Calculate directions p^k, d_1^k:

$p^k = d_0^k - \delta_k Q_k^T e$, $d_1^k = -Q_k^T\{\delta_k e + \tilde{f}^k - \mathrm{H}(x^k)\tilde{e}^k\}$,

where

$U_k = \sum_{j \in I} \max\{0, \mu_j^k\}$, $V_k = \min\{c, \|d_0^k\|^{\chi}\}$, $\delta_k = \dfrac{V_k (d_0^k)^T B_k d_0^k + \varphi(x^k)^e}{1 + U_k}$,

$I_k = L_+^k \cup L_-^k$, $L_+^k = \{j \in I_+(x^k) | f_j(x^k) + g_j(x^k)^T d_0^k = \varphi(x^k)\}$,

$L_-^k = \{j \in I_-(x^k) | f_j(x^k) + g_j(x^k)^T d_0^k = 0\}$,

$$D_j^k = \begin{cases} \varphi(x^k) - f_j(x^k), & j \in I_+^k, \\ -f_j(x^k), & j \in I_-^k, \end{cases}$$

$$\tilde{e}_j^k = \begin{cases} 1, & j \in L_+^k, \\ 0, & j \in I \backslash L_+^k, \end{cases} \quad \tilde{f}_j^k = \begin{cases} f_j(x^k + d_0^k), & j \in I_k, \\ 0, & j \in I \backslash I_k, \end{cases}$$

$D_k = \mathrm{diag}(D_j^k, j \in I)$, $\mu^k = -Q_k g_0(x^k)$, $Q_k = (N(x^k)^T N(x^k) + D_k)^{-1} N(x^k)^T$.

Step 4 Line search. (i) If

$f_0(x^k + d_0^k + d_1^k) - f_0(x^k) \leqslant 0.25 g_0(x^k)^T p^k$,

$f_j(x^k + d_0^k + d_1^k) - \varphi(x^k) \leqslant -0.5\delta_k$, $j \in I_+^k$; $f_j(x^k + d_0^k + d_1^k) \leqslant 0$, $j \in I_-^k$,

then let $\lambda_k = 1$, $d^k = d_0^k + d_1^k$, go to Step 5;

(ii) Find the first number λ_k of sequence $\{1, 2^{-1}, 2^{-2}, \cdots\}$ satisfying

$f_0(x^k + \lambda p^k) - f_0(x^k) \leqslant 0.5\lambda (g_0(x^k)^T p^k + \varphi(x^k)^e)$,

$f_j(x^k + \lambda p^k) - \varphi(x^k) \leqslant -0.5\lambda\delta_k$, $j \in I_+^k$; $f_j(x^k + \lambda p^k) \leqslant 0$, $j \in I_-^k$. Let $d^k = p^k$, go to Step 5.

Step 5 Let $x^{k+1} = x^k + \lambda_k d^k$, produce a new positively definite matrix B_{k+1}. Let $k := k+1$, go to Step 2.

To analyse unitedly the convergence and superlinear rate of convergence of Algorithms 2.8, 3.3, 3.4, suppose,

Assumption 3.5 Sequences x^k and $\{d_0^k\}$ are bounded, $\{B_k\}$ is uniformly positive definite.

Theorem 3.6 (i) Let Assumptions 1.2(i), 2.1, 3.5 be satisfied. Then Algorithms 2.8, 3.3 and 3.4 are all globally convergent.

(ii) Assume that Assumptions 1.2 (i, ii), 2.1 and 3.5 hold. Then Algorithms 2.8, 3.3 and 3.4 are all strongly convergent.

(iii) Suppose, in addition Assumptions 1.2 (i, ii), 2.1 and 3.5, that Relations (1.11) and (1.14) hold true. Then Algorithms 2.8, 3.3 and 3.4 are all one-step superlinearly convergent.

4 Feasible method for general constraints

The algorithms introduced above are all limited to optimal problems with inequality constraints, i.e. $J = \varnothing$. In 1995, Jian[3] extended the SQP method and presented a "feasible descent" algorithm for equality and inequality constrained optimization problems. The basic idea is to convert Problem (1.1) into an inequality constrained problem. It is showed in the following.

$$\min\{F_c(x) \stackrel{\text{def}}{=} f_0(x) - c\sum_{j \in J} f_j(x) | f_j(x) \leqslant 0, j \in L = I \cup J\}.$$

Algorithm 4.1[3]

Denote

$X_+ = \{x \in R^n | f_j(x) \leqslant 0, j \in L\}$, $I(x, \varepsilon) = \{j \in I | -f_j(x) \leqslant \varepsilon\} \cup J$.

The non-degenerate Assumption 2.1 is that vectors $g_j(x) | j \in I(x, 0)\}$ are linearly independent for any $x \in X_+$.

Step 1 Choose initial point $x^1 \in X_+$ and matrix B_1, parameters $M, \mathrm{W}, \varepsilon, \chi > 0$; $V, W > 2$; $f \in (2,3)$, $T \in (0, 0.5)$, $U \in (0,1)$, $k = 1$. $\mathrm{H}_j \equiv 0$, $\forall j \in L$.

Step 2 Use pivoting operation to generate $L_k^- = I(x^k, \mathrm{W})$ such that $\det(N_k^T N_k) \geqslant \mathrm{W}$, where $N_k = (g_j(x^k), j \in L_k^-)$.

Step 3 Update parameter c, Calculate

$\mu_k = -Q_k g_0(x^k)$, $Q_k = (N_k^T N_k)^{-1} N_k^T$, $\tilde{\alpha}_k = \min\{\mu_j^k, j \in J\}$,

$$\alpha_k = \begin{cases} c_{k-1}, & \alpha_{k-1} \geqslant -\tilde{\alpha}_k + X, \\ \max\{X - \tilde{\alpha}_k, \alpha_{k-1} + X\}, & \alpha_{k-1} < -\tilde{\alpha}_k + X \end{cases}$$

Step 4 Let $T_k = J, L_k^0 = \emptyset, L_k^- = I(x^k, W_k), c = \alpha_k$, solve QP(1.3) to get a $K-T$ point d_0^k. If $d_0^k = 0$, then x^k is a $K-T$ point of Problem (1.1), stop; if d_0^k does not exist or $\|d_0^k\| > 1$, then go to Step 7.

Step 5 Compute

$$d^k = d_0^k - \|d_0^k\|^V Q_k^T e, \theta_k = \nabla F_{c_k}(x^k)^T d^k.$$

If

$$\|d^k\| \leqslant M, \theta_k \leqslant \min\{-\|d_0^k\|^W, -\|d^k\|^W\},$$

then go to Step 6. Otherwise, go to Step 7.

Step 6 Calculate $I_k = \{j \in L_k^- \mid f_j(x^k) + g_j(x^k)^T d_0^k = 0\}$ and

$$\tilde{d}^k = -Q_k^T(\|d^k\|^f e + \tilde{f}^k), \tilde{f}_j^k = \begin{cases} f_j(x^k + d^k), j \in I_k, \\ 0, j \in L_k^- \setminus I_k, \end{cases}$$

if $\|\tilde{d}^k\| > \|d^k\|$, then let $\tilde{d}^k = 0$. Go to Step 8.

Step 7 Find a "first-order" feasible descent direction d^k, let $\tilde{d}^k = 0, \theta_k = \nabla F_{c_k}(x^k)^T d^k$.

Step 8 Calculate the first number λ_k of the sequence $\{1, U, U^2, \cdots\}$ satisfying

$$F_{c_k}(x^k + \lambda d^k + \lambda^2 \tilde{d}^k) \leqslant F_{c_k}(x^k) + T\lambda\theta_k, f_j(x^k + \lambda d^k + \lambda^2 \tilde{d}^k) \leqslant 0, \forall j \in L.$$

Step 9 It is the same as Step 6 in Algorithm 2.2.

The convergence and superlinear rate of convergence of Algorithm 4.1 are similar to Theorem 2.7.

References

1 Panier E R, Tits A L. A superlinearly convergent feasible method for the solution of inequality constrained optimization problems. SIAM J Control and Optimization, 1987, 25 (4): 934~ 950.

2 Jian J B, Tau P H. A two-step superlinearly convergent-feasible method for nonlinear programming problems. In: Wang Yinqing. OR and Decision Marking. Vol I. Chengdu (China): Chengdu University of Science and Technology Press, 1992. 461~ 467.

3 Jian J B. A superlinearly convergent feasible descent algorithm for nonlinear optimization. Journal of Mathematics (China), 1995, 15(3): 319~ 326.

4 Jian J B. A subfeasible direction method combining SQP technique with generalized projection. Applied Mathematics (Ser A) (China), 1996, 11(1): 65~ 74.

5 Jian J B. A superlinearly convergent combined phase I-phase II subfeasible method. J Sys Sci and Sys Eng, 1994, 3(2): 104~ 111.

6 Jian J B. A strong subfeasible directions algorithm with superlinear convergence. J Sys Sci and Sys Eng, 1996, 5(3): 287~ 295.

7 Jian J B, Xue S J. A class of superlinearly convergent feasible methods for nonlineraly constrained optimization. Journal of Mathematical Research and Exposition, 1999, 19 (1): 135~ 140.

8 Gao Z Y, Wu F. A SQP feasible method for nonlinearl constraints. Acta Mathematicae Applicatae Sinica (Ser A), 1995, 18(4): 579~ 590.

9 Gao Z Y, Wu F. A superlinearly feasible method for nonlinearl constraints. Acta Mathematica Sinica (Ser A), 1997, 40(6): 895~ 900.

10 Gao Z Y. A superlineaely convergent feasible method for nonlinear programming with nonlinear inequality constraints(Part I): Algorithm. J of Northern Jiaotong University, 1996, 20(1): 51~ 60.

11 Gao Z Y, Wu F, Lai Y L. A superlinearly convergent algorithm of sequential systems of equations for nonlinear optimization. Chinese Science Bulletin, 1994, 39(9): 774~ 777.

12 Gao Z Y, He G P, Wu F. A new method for nonlinear optimization problems-sequential systems of linear equations algorithm. In: Du D Z, Zhang X S, Cheng K. Operations Research and Its Applications. Lecture Notes in Operation Research 1. Beijing: World Publishing Corporation, 1995. 64~ 73.

13 Panier E R, Tits A L, Herskovits J N. A QP-free global convergent, locally superlinearly convergent algorithm for inequality constrained optimization. SIAM J Control and Optimization, 1988, 26(4): 788~ 811.

14 Zhang K C, Jian J B. A superlinearly convergent generalized projection-sequential systems of linear equations algoritnm. J of Guangxi University, 1999, 24(1): 9~ 14.

15 Jian J B. A generalized projection-successive linear equations algorithm for nonlinearly equality and inequality constrained optimization and its rate of convergence. Appl Mathematics (Ser B) (China), 1997, 12(4): 343~ 354.

16 Jian J B. A fast convergent directly feasible directions method for nonlinearly inequality constrained optimization. J of Sys Sci and Sys Eng, 1997, 6(1): 5~ 14.

17 Jian J B. A superlinearly convergent directionly strongly subfeasible directions method for nonlinear inequality constrained optimization a generalized projection variable metric method. J of Sys Sci and Sys Eng, 1997, 6(4): 418 ~ 429.

18 Jian J B. Two extension models of SQP and SSLE algorithms for optimization and their superlinear and quadrtical convergence. Applied Mathematics (Ser A) (China), 2001, 16(4): 435~ 444.

19 Pantoja J F A De O, Mayne D Q. Exact penalty function algorithm with simple updating of the penalty parameters. JOTA, 1991, 69(1): 441~ 467.

20 Jian J B. On subfeasible and strongly subfeasible directions of methods. In: Yue M Y. Optimization Theory and Applications. Xi'an University of Electronic and Technology of China Press, 1994. 113~ 118.

(责任编辑: 蒋汉明)