A filter SQP algorithm without a feasibility restoration phase*

CHUNGEN SHEN[1,2], WENJUAN XUE[3] and DINGGUO PU[1]

[1]Department of Mathematics, Tongji University, China, 200092
[2]Department of Applied Mathematics, Shanghai Finance University, China, 201209
[3]Department of Mathematics and Physics, Shanghai University of Electric Power, China, 200090
E-mail: shenchungen@gmail.com

Abstract. In this paper we present a filter sequential quadratic programming (SQP) algorithm for solving constrained optimization problems. This algorithm is based on the modified quadratic programming (QP) subproblem proposed by Burke and Han, and it can avoid the infeasibility of the QP subproblem at each iteration. Compared with other filter SQP algorithms, our algorithm does not require any restoration phase procedure which may spend a large amount of computation. We underline that global convergence is derived without assuming any constraint qualifications. Preliminary numerical results are reported.

Mathematical subject classification: 65K05, 90C30.

Key words: filter, SQP, constrained optimization, restoration phase.

1 Introduction

In this paper, we consider the constrained optimization problem:

$$(\text{NLP}) \quad \begin{cases} \min & f(x) \\ \text{s.t.} & c_{\mathcal{E}}(x) = 0, \\ & c_{\mathcal{I}}(x) \le 0, \end{cases}$$

where $c_{\mathcal{E}}(x) = (c_1(x), c_2(x), \cdots, c_{m_1}(x))^T, c_{\mathcal{I}}(x) = (c_{m_1+1}(x), c_{m_1+2}(x), \cdots,$ $c_m(x))^T, \mathcal{E} = \{1, 2, \cdots, m_1\}, \mathcal{I} = \{m_1 + 1, m_1 + 2, \cdots, m\}, f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $c_i : \mathbb{R}^n \rightarrow \mathbb{R}, i \in \mathcal{E} \cup \mathcal{I}$ are twice continuously differentiable functions. The feasible set of the problem (NLP) is denoted by $X = \{x \in R^n | c_{\mathcal{E}}(x) = 0, \ c_{\mathcal{I}}(x) \leq 0\}$.

The sequential quadratic programming (SQP) method has been widely used for solving the problem (NLP). It generates a sequence $\{x_k\}$ converging to the desired solution by solving the quadratic programming (QP) subproblem:

$$\begin{cases} \min & q(d) = \nabla f(x_k)^T d + \frac{1}{2} d^T B_k d \\ \text{s.t.} & \nabla c_{\mathcal{E}}(x_k)^T d + c_{\mathcal{E}}(x_k) = 0, \\ & \nabla c_{\mathcal{I}}(x_k)^T d + c_{\mathcal{I}}(x_k) \leq 0, \\ & \|d\|_\infty \leq \rho, \end{cases} \tag{1.1}$$

where $\rho > 0$ is the trust region radius, $B_k \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix. At each iteration x_k, the solution d of the QP subproblem is regarded as a trial step and the next trial iterate has the form $x_k + d$. The acceptance of this trial step depends on whether the trial iterate $x_k + d$ makes some merit function descent. Generally, this merit function is a type of penalty function with some parameters whose adjustment can be problematic. Fletcher and Leyffer [7] proposed a trust region filter SQP method to solve the problem (NLP) instead of traditional merit function SQP methods. In addition, the computational results presented in Fletcher and Leyffer [7] are also very encouraging.

Recently, this topic got high importance in recent years (see [6, 17, 18, 22, 23]). Trust region filter SQP methods have been studied by Fletcher, Leyffer and Toint in [8] and by Fletcher, Gould, Leyffer, Toint and Wächter in [9]. In this latter paper, an approximate solution of the QP subproblem is computed and the trial step is decomposed into normal and tangential components. Gonzaga, Karas and Vanti [10] presented a general framework for filter methods where each step is composed of a feasibility phase and an optimality phase. Similar filter method was proposed by Ribeiro, Karas and Gonzaga in [21]. In all these papers only the global convergence of the proposed methods is analyzed. On the other hand, in [24], Ulbrich studied the local convergence of a trust region filter SQP method. Anyway, the components in the filter adopted in [24], differs from those in [7, 8, 9]. It should be underlined that the filters approach has been

used also in conjunction with the line search strategy (see Wächter and Biegler [26, 27]); with interior point methods (see Benson, Shanno and Vanderbei [5]; Ulbrich, Ulbrich and Vicente [25]) and with the pattern search method (see Audet and Dennis [1]; Karas, Ribeiro, Sagastizábal and Solodov [15]).

However, all filter algorithms mentioned above include the provision for a feasibility restoration phase if the QP subproblem becomes inconsistent. Although any method (e.g. [3, 4, 16]) for solving a nonlinear algebraic system of equalities and inequalities can be used to implement this calculation, a large amount of computation may be spent. In this paper, we incorporate the filter technique with the modified QP subproblem for solving general constrained optimization problems. The main feature of this paper is that there is no restoration phase procedure. Another feature is that global convergence of our algorithm is established without assuming any constraint qualifications.

This paper is organized as follows. In section 2, we describe how the modified QP subproblem is embedded in the filter algorithm. In section 3, it is proved that the algorithm is well defined, and is globally convergent to a stationary point. If Mangasarian-Fromowitz constraint qualification (MFCQ) holds at this stationary point, then it is a KKT point. In section 4, we report some numerical results.

2 Algorithm

It is well-known that the Karush-Kuhn-Tucker (KKT) conditions for the problem (NLP) are:

$$
\begin{cases}
\nabla_x L(x, \lambda, \mu) = 0, \ c_{\mathcal{E}}(x) = 0, \ c_{\mathcal{I}}(x) \leq 0, \\
\lambda \geq 0, \ \lambda^T c_{\mathcal{I}}(x) = 0.
\end{cases}
\tag{2.2}
$$

where

$$
L(x, \lambda, \mu) = f(x) + \mu^T c_{\mathcal{E}}(x) + \lambda^T c_{\mathcal{I}}(x),
\tag{2.3}
$$

is the Lagrangian function, $\mu \in R^{m_1}$ and $\lambda \in R^{m-m_1}$ are the multipliers corresponding to the constraints of the problem (NLP).

The SQP method generates iterates by solving the subproblem (1.1). However, the subproblem (1.1) may be inconsistent if x_k is not a feasible point for the problem (NLP). In order to avoid this bad situation, we solve a modified QP

subproblem instead of the subproblem (1.1). Before introducing the modified QP subproblem, we first solve the problem

$$
\begin{cases}
\min \quad q_l(d) = \sum_{i \in \mathcal{E}} |\nabla c_i(x_k)^T d + c_i(x_k)| \\
\qquad\qquad + \sum_{i \in \mathcal{I}} \max\left\{0, \nabla c_i(x_k)^T d + c_i(x_k)\right\}, \\
\text{s.t.} \quad \|d\|_\infty \le \sigma_k,
\end{cases} \tag{2.4}
$$

where the scalar $\sigma_k > 0$ is used to restrict d in norm. Let $\tilde{d}_k$ denote the solution of (2.4). If $\sigma_k \le \rho$ and $q_l(\tilde{d}_k) = 0$, then the QP subproblem (1.1) is consistent. The problem (2.4) can be reformulated as a linear program

$$
\begin{cases}
\displaystyle\min_{\substack{d \in \mathbb{R}^n, z_1 \in \mathbb{R}^{m_1}, \\ z_2 \in \mathbb{R}^{m-m_1}}} \quad e_{m_1}^T z_1 + e_{m-m_1}^T z_2 \\
\text{s.t.} \qquad\qquad -z_1 \le \nabla c_{\mathcal{E}}(x_k)^T d + c_{\mathcal{E}}(x_k) \le z_1, \\
\qquad\qquad\qquad \nabla c_{\mathcal{I}}(x_k)^T d + c_{\mathcal{I}}(x_k) \le z_2, \\
\qquad\qquad\qquad z_2 \ge 0, \\
\qquad\qquad\qquad \|d\|_\infty \le \sigma_k,
\end{cases} \tag{2.5}
$$

where vectors $e_{m_1} = (1, 1, \cdots, 1)^T \in \mathbb{R}^{m_1}$ and $e_{m-m_1} = (1, 1, \cdots, 1)^T \in \mathbb{R}^{m-m_1}$.

Now we define that $\bar{s}(x_k, \sigma_k)$ and $\bar{r}(x_k, \sigma_k)$ are equal to $\tilde{z}_2$ and $\nabla c_{\mathcal{E}}(x_k)^T \tilde{d}_k + c_{\mathcal{E}}(x_k)$, respectively, where $(\tilde{d}_k, \tilde{z}_1, \tilde{z}_2)$ denotes the solution of (2.5). At each iteration, we generate the trial step by solving the modified QP subproblem

$$
\underline{QP}(x_k, B_k, \sigma_k, \rho_k) \quad
\begin{cases}
\min \quad q(d) = \nabla f(x_k)^T d + \frac{1}{2} d^T B_k d \\
\text{s.t.} \quad \nabla c_{\mathcal{E}}(x_k)^T d + c_{\mathcal{E}}(x_k) = \bar{r}(x_k, \sigma_k), \\
\qquad\quad \nabla c_{\mathcal{I}}(x_k)^T d + c_{\mathcal{I}}(x_k) \le \bar{s}(x_k, \sigma_k), \\
\qquad\quad \|d\|_\infty \le \rho_k.
\end{cases}
$$

Here, at each iterate, we require that ρ_k be greater than σ_k. So, the choice of σ_k depends on ρ_k.

Define $\Phi(x_k, \sigma_k) = \sum_{i \in \mathcal{E}} |\bar{r}_i(x_k, \sigma_k)| + \sum_{i \in \mathcal{I}} \bar{s}_i(x_k, \sigma_k)$, where $\bar{r}_i(x_k, \sigma_k)$ and $\bar{s}_i(x_k, \sigma_k)$ denote the ith component of $\bar{r}(x_k, \sigma_k)$ and $\bar{s}(x_k, \sigma_k)$ respectively. Actually, $\Phi(x_k, \sigma_k) = q_l(\tilde{d}_k)$ with the scalar $\sigma_k > 0$. Obviously, $d = 0$ is

feasible for the problem (2.4) and then $q_l(\tilde{d}_k) \leq q_l(0) = V(x_k)$, where the constraint violation $V(x)$ is defined as

$$V(x) = \sum_{i \in \mathcal{E}} |c_i(x)| + \sum_{i \in \mathcal{I}} \max \{0, c_i(x)\}.$$

Therefore, $\Phi(x_k, \sigma_k) \leq V(x_k)$.

Let d denote the solution of $\underline{QP}(x_k, B_k, \sigma_k, \rho_k)$. Let the trial point be $\hat{x} = x_k + d$. The KKT conditions for the subproblem $\underline{QP}(x_k, B_k, \sigma_k, \rho_k)$ are

$$\begin{cases} \nabla f(x_k) + \sum_{i \in \mathcal{I}} \lambda_{k,i} \nabla c_i(x_k) + \sum_{i \in \mathcal{E}} \mu_{k,i} \nabla c_i(x_k) + B_k d + \lambda_k^u - \lambda_k^l = 0, \\ \lambda_{k,i} (\nabla c_i(x_k)^T d + c_i(x_k) - \bar{s}_i(x_k, \sigma_k)) = 0, \ i \in \mathcal{I}, \\ (d - \rho_k e_n)^T \lambda_k^u = 0, \ (d + \rho_k e_n)^T \lambda_k^l = 0, \\ \nabla c_i(x_k)^T d + c_i(x_k) - \bar{r}_i(x_k, \sigma_k) = 0, \ i \in \mathcal{E}, \\ \nabla c_i(x_k)^T d + c_i(x_k) \leq \bar{s}_i(x_k, \sigma_k), \ i \in \mathcal{I}, \ \|d\|_\infty \leq \rho_k, \\ \lambda_{k,i} \geq 0, \ i \in \mathcal{I}, \ \lambda_{k,j}^u \geq 0, \ \lambda_{k,j}^l \geq 0, \ j \in \{1, 2, \cdots, n\}, \end{cases} \qquad (2.6)$$

where $e_n = (1, 1, \cdots, 1)^T \in \mathbb{R}^n$, $\mu_k \in \mathbb{R}^{m_1}$, $\lambda_k \in \mathbb{R}^{m-m_1}$, $\lambda_k^l \in \mathbb{R}^n$ and $\lambda_k^u \in \mathbb{R}^n$.

In our filter, we consider pairs of values $(V(x), f(x))$. Definitions 1–4 below are very similar to those in [7, 8].

Definition 1. The iterate x_k dominates the iterate x_l if and only if $V(x_k) \leq V(x_l)$ and $f(x_k) \leq f(x_l)$. And it is denoted by $x_k \preceq x_l$.

Thus, if $x_k \preceq x_l$, the latter is of no real interest to us since x_k is at least as good as x_l with respect to the objective function's value and the constraint violation. Furthermore, if $x_k \preceq x_l$, we say that the pair $(V(x_k), f(x_k))$ dominates the pair $(V(x_l), f(x_l))$.

Definition 2. The kth filter is a list of pairs $\{(V(x_l), f(x_l))\}_{l<k}$, such that no pair dominates any other.

Let $\mathcal{F}_k$ denote the indices in the kth filter, i.e.,

$$\mathcal{F}_k = \{l < k : x_j \npreceq x_l \text{ for all } j \in \{0, 1, 2, \ldots, k-1\} \setminus \{l\}\}.$$

Filter methods accept a trial point $\hat{x} = x_k + d$ if its corresponding pair $(V(\hat{x}), f(\hat{x}))$ is not dominated by any other pair in the kth filter, neither the pair corresponding to x_k, i.e., $(V(x_k), f(x_k))$.

In a filter algorithm, one accepts a new pair $(V(\hat{x}), f(\hat{x}))$ if it cannot be dominated by other pairs in the current filter. Although the definition of filter is simple, it needs to be refined a little in order to guarantee the global convergence.

Definition 3.   A new trial point $\hat{x}$ is said to be "acceptable to x_l", if

$$V(\hat{x}) - V(x_l) \leq -\gamma_1 V(\hat{x}), \tag{2.7a}$$

$$\text{or} \quad f(\hat{x}) - f(x_l) < -\gamma_2 V(\hat{x}), \tag{2.7b}$$

is satisfied, where $\gamma_i \in \left(0, \frac{1}{2}\right)$, $i = 1, 2$ are two scalars.

Definition 4.   A new trial point $\hat{x}$ is said to be "acceptable to the kth filter" if $\hat{x}$ is acceptable to x_l for all $l \in \mathcal{F}_k$.

Therefore, to accept a new pair into the current filter, we should test the conditions defined in Definition 4. If the new trial point $\hat{x}$ is acceptable in the sense of Definitions 3 and 4, we may wish to add the corresponding pair to the filter. Meanwhile, any pair in the current filter dominated by the new pair is removed.

In order to restrict $V(x_k)$, it still needs an upper bound condition for accepting a point. The trial point $\hat{x}$ satisfies the upper bound condition if

$$V(\hat{x}) \leq \mathcal{U}_k \tag{2.8}$$

holds, where $\mathcal{U}_k$ is a positive scalar. Here $\mathcal{U}_k$ is updated at each iteration and it may converge to zero for some instances. We aim to control the constraint violation by setting $\mathcal{U}_k$. In current iteration k, if a trial step is accepted and $\Phi(x_k, \sigma_k) = 0$, then we will keep $\mathcal{U}_k$ unchanged in next iteration. Otherwise we will reduce $\mathcal{U}_k$ in next iteration. The detailed information about the update of $\mathcal{U}_k$ is given in Algorithm 2.1.

Denote

$$\Delta q(d) = q(0) - q(d) = -\nabla f(x_k)^T d - \frac{1}{2} d^T B_k d \tag{2.9}$$

as the quadratic reduction of $f(x)$, and

$$\Delta f(d) = f(x_k) - f(x_k + d) \tag{2.10}$$

as the actual reduction of $f(x)$. In our algorithm, if a trial point $x_k + d$ is acceptable to the current filter and x_k, and

$$\Delta q(d) > 0, \tag{2.11}$$

the sufficient reduction criterion of the objective function f should be satisfied, i.e.,

$$\Delta f(d) \geq \eta \Delta q(d), \quad \eta \in \left(0, \frac{1}{2}\right). \tag{2.12}$$

Now we are in the position to state our filter SQP algorithm.

Algorithm 2.1.

Step 1. Initialization.

Given initial point x_0, parameters $\eta \in \left(0, \frac{1}{2}\right)$, $\gamma_1 \in \left(0, \frac{1}{2}\right)$, $\gamma_2 \in \left(0, \frac{1}{2}\right)$, $\gamma \in (0, 0.1)$, $r \in (0, 1)$, $\mathcal{U}_0 > 0$, $\rho_{\min} > 0$, $\bar{\rho} > \rho_{\min}$, $\rho \in [\rho_{\min}, \bar{\rho}]$, $\sigma_0 \in (\gamma\rho, \rho)$. Set $k = 0$, Flag=1, put k into $\mathcal{F}_k$.

Step 2. Compute the search direction.

Compute the search direction d from the subproblem $\underline{QP}(x_k, B_k, \sigma_k, \rho)$. If $\rho \geq \rho_{\min}$, set $d_s = d$ and $\Phi_s = \Phi(x_k, \sigma_k)$.

Step 3. Check for termination.

If $d = 0$ and $V(x_k) = 0$, then stop;

Else if $V(x_k) > 0$ and $\Phi(x_k, \sigma_k) - V(x_k) = 0$, then stop.

Step 4. Test to accept the trial point.

4.1. Check acceptability to the filter.

If the upper bound condition (2.8) holds for $\hat{x} = x_k + d$, $\Phi(x_k, \sigma_k) = 0$, and $x_k + d$ is acceptable to both the kth filter and x_k, then go to 4.3.

Otherwise, if $\Phi(x_k, \sigma_k) = 0$, then set $\rho = \frac{1}{2}\rho$ and choose $\sigma_k \in (\gamma\rho, \rho)$, go to Step 2;

4.2. Compute t_k, the first number t of the sequence $\{1, r, r^2, \cdots, \}$ satisfying

$$V\left(x_k + t_k d_s\right) - V\left(x_k\right) \le t_k \eta \left(\Phi_s - V(x_k)\right). \qquad (2.13)$$

Set Flag=0, $d = d_s$, and go to 4.4.

4.3. Check for sufficient reduction criterion.

If $\Delta f(d) < \eta \Delta q(d)$ and $\Delta q(d) > 0$, then $\rho = \frac{1}{2}\rho$ and choose $\sigma_k \in (\gamma\rho, \rho)$, and go to Step 2;

otherwise go to 4.4.

4.4. Accept the trial point.

$$\text{Set } \rho_k = \rho, \ d_k = d, \ x_{k+1} = \begin{cases} x_k + d_k, & \text{if Flag=1,} \\ x_k + t_k d_k, & \text{if Flag=0,} \end{cases}$$

and $\rho = \min \left\{ \max\{2\rho, \rho_{\min}\}, \bar{\rho}\right\}$.

Step 5. Augment the filter if necessary.

If Flag=1 and $\Delta q(d_k) \le 0$, then include $(V(x_k), f(x_k))$ in the kth filter.

If Flag=0, then set $\mathcal{U}_{k+1} = V(x_k + t_k d_k)$; otherwise, set $\mathcal{U}_{k+1} = \mathcal{U}_k$ and leave the filter unchanged.

Step 6. Update.

Compute B_{k+1}. Set $\sigma_{k+1} \in (\gamma\rho, \rho)$ and $k := k + 1$, Flag=1, go to Step 2.

Remark 1. The role d_s in Step 2 is to record a trial step d with $\rho \ge \rho_{\min}$. By Step 4.1, we know that if $\Phi(x_k, \sigma_k) > 0$, then both conditions in Step 4.1 are violated. A trial iterate $x_k + d$ with $\Phi(x_k, \sigma_k) > 0$ may be considered as a "worse" step which may be far away from the feasible region. So Step 4.2 is executed to reduce the constraint violation. The step d_s is used in Step 4.2 due to the descent property of d_s proved by Lemma 3.4. In addition, it should be interpreted that, at the beginning of iteration k, the pair $(V(x_k), f(x_k))$ is not in the current filter, but x_k must be acceptable to the current filter.

Remark 2. Our algorithm has three loops: the loop 2-(4.1)-2, the loop 2-(4.3)-2, and the loop 2-6-2. We observed that the radius ρ is allowed to reduce to a value less than $\rho_{\min}$ in the loops 2-(4.1)-2 and 2-(4.3)-2. But in the loop 2-6-2
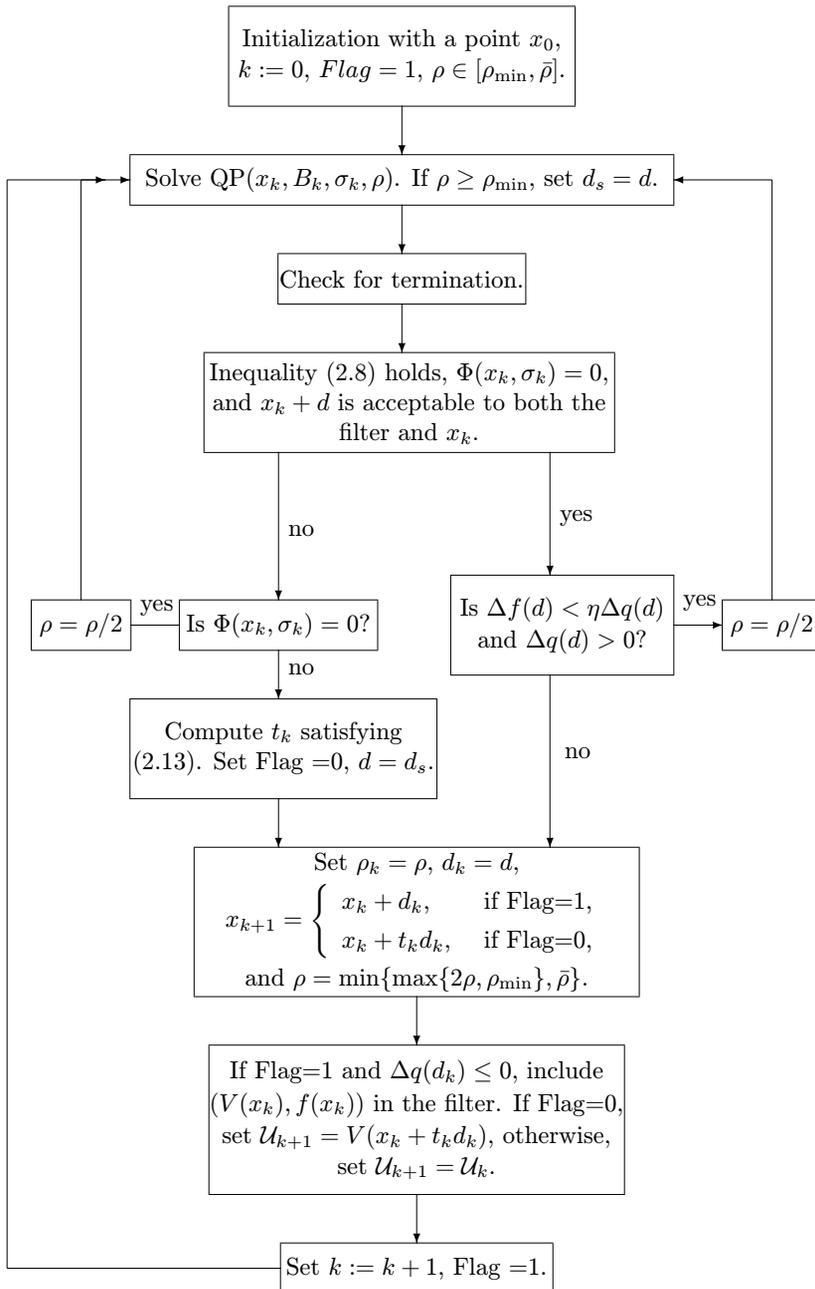
Figure 1 – Flowchart of Algorithm 2.1.

the radius ρ is not less than $\rho_{\min}$. At each iteration, the first trial radius ρ is greater than or equal to $\rho_{\min}$. Subsequently, trial radius ρ may not stop decreasing until either the filter acceptance criteria are satisfied or Step 4.2 is executed. Hence, ρ may be less than $\rho_{\min}$ during the execution process of the loops 2-(4.1)-2 and 2-(4.3)-2.

As for proving global convergence, we use the terminology firstly introduced by Fletcher, Leyffer and Toint [8]. We call d an f-*type step* if $\Delta q(d) > 0$, indicating that then the sufficient reduction criterion (2.12) is required. If d is accepted as the final step d_k in the kth iteration, we refer to k as an f-*type iteration*.

Similarly, we call d a V-*type step* if $\Delta q(d) \leq 0$. If d is accepted as the final step d_k in iteration k, we refer to k as a V-*type iteration*. In addition, if x_k is generated by Step 4.2, we also refer to it as a V-*type iteration*.

If $f(x_{k+1}) < f(x_k)$, then we regard the step d_k as an f *monotone step*. Obviously, an f-*type step* must be an f *monotone step*.

3   Global convergence

In this section, we prove the global convergence of Algorithm 2.1. Firstly, we give some assumptions:

(A1) Let $\{x_k\}$ be generated by Algorithm 2.1 and $\{x_k\}$, $\{x_k + d_k\}$ are contained in a closed and bounded set S of $\mathbb{R}^n$;

(A2) All the functions $f, c_i, i \in \mathcal{E} \cup \mathcal{I}$ are twice continuously differentiable on S.

(A3) The matrix B_k is uniformly positive definite and bounded for all k.

Remark 3.   Assumption (A1) is reasonable. It may be forced if, for example, the original problem involves a bounded box among its constraints.

Remark 4.   A consequence of Assumption (A3) is that there exist constants $\delta, M > 0$, independent of k such that $\delta \|y\|^2 \leq y^T B_k y \leq M \|y\|^2$ for all $y \in \mathbb{R}^n$. Assumptions (A1) and (A2) imply boundedness of $\|\nabla^2 c_i(x)\|$ $i \in \mathcal{E} \cup \mathcal{I}$, and

$\|\nabla^2 f(x)\|$ on S. Without loss of generality, we may assume $\|\nabla^2 c_i(x)\| \le M$, $i \in \mathcal{E} \cup \mathcal{I}$, $\|\nabla^2 f(x)\| \le M$, $\forall x \in S$.

Lemma 3.1. *Assume $\bar{x}$ is not a stationary point of V in the sense that $0 \notin \partial V(\bar{x})$, where ∂V denotes the Clarke subdifferential of V. Then there exists a scalar $\bar{\epsilon} > 0$ and a neighborhood $N(\bar{x})$ of $\bar{x}$, such that $\Phi(x, \sigma) - V(x) < -\bar{\epsilon}$ for all $\sigma \ge \gamma \rho_{\min}$ and all $x \in N(\bar{x})$.*

Proof. By [2, Lemma 2.1], $0 \notin \partial V(\bar{x})$ implies $\Phi(\bar{x}, \gamma \rho_{\min}) - V(\bar{x}) < 0$, where γ and $\rho_{\min}$ are from Algorithm 2.1. By the continuity of the function $\Phi(\cdot, \gamma \rho_{\min}) - V(\cdot)$ on $\mathbb{R}^n$, there exists a neighborhood $N(\bar{x})$ and a scalar $\bar{\epsilon} > 0$ such that $\Phi(x, \gamma \rho_{\min}) - V(x) < -\bar{\epsilon}$ whenever $x \in N(\bar{x})$. The condition $\sigma \ge \gamma \rho_{\min}$ together with the definition of Φ yields $\Phi(x, \sigma) - V(x) \le \Phi(x, \gamma \rho_{\min}) - V(x)$. Therefore, $\Phi(x, \sigma) - V(x) < -\bar{\epsilon}$ holds for all $\sigma \ge \gamma \rho_{\min}$ and all $x \in N(\bar{x})$. $\square$

Remark 5. It can be seen that $\bar{\epsilon}$ depends on $\bar{x}$ such that $0 \notin \partial V(\bar{x})$ (i.e., $\bar{\epsilon} = \bar{\epsilon}(\bar{x})$), because it must satisfy $\Phi(\bar{x}, \sigma) - V(\bar{x}) < -\bar{\epsilon}$.

Lemma 3.2. *Let $d_k = 0$ be a feasible point of $\underline{QP}(x_k, B_k, \sigma_k, \rho_k)$. Then x_k is a stationary point of $V(x)$. Moreover, if $x_k \in X$, then x_k is a KKT point of the problem (NLP).*

Proof. Since $d_k = 0$ implies $\Phi(x_k, \sigma_k) - V(x_k) = 0$, it follows from [2, Lemma 2.1] that x_k is a stationary point of $V(x)$. If $V(x_k) = 0$ and $d_k = 0$, then it follows from [2, Lemma 2.2] that x_k is a KKT point for the problem (NLP). $\square$

Lemma 3.3. *Let Assumptions (A1)-(A3) hold and d be a feasible point of the subproblem $\underline{QP}(x, B, \sigma, \rho)$, then we have*

$$V(x + td) - V(x) \le t(\Phi(x, \sigma) - V(x)) + \frac{1}{2}t^2 mn M \rho^2, \qquad (3.14)$$

for $t \in [0, 1]$.

Proof.   By Taylor Expansion formula, the feasibility of d and Assumption (A2), we have that for $i \in \mathcal{E}$ and $t \in [0, 1]$,

$$
\begin{aligned}
|c_i(x + td)| &= |c_i(x) + t\nabla c_i(x)^T d + \frac{1}{2}t^2 d^T \nabla^2 c_i(z_i)d| \\
&\leq (1 - t)|c_i(x)| + t|(c_i(x) + \nabla c_i(x)^T d)| + \frac{1}{2}t^2 nM\rho^2, \quad (3.15)
\end{aligned}
$$

and for $i \in \mathcal{I}$ and $t \in [0, 1]$,

$$
\begin{aligned}
c_i(x + td) &= c_i(x) + t\nabla c_i(x)^T d + \frac{1}{2}t^2 d^T \nabla^2 c_i(z_i)d \\
&\leq (1 - t)c_i(x) + t(c_i(x) + \nabla c_i(x)^T d) + \frac{1}{2}t^2 nM\rho^2, \quad (3.16)
\end{aligned}
$$

where the vector z_i is between x and $x + td$. The term $\frac{1}{2}t^2 nM\rho^2$ in (3.15) and (3.16) is derived because

$$
|d^T \nabla^2 c_i(z_i)d| \leq nM\rho^2,
$$

where we use that $\| \cdot \|^2 \leq n\| \cdot \|_\infty^2$. Formulae (3.15) and (3.16) combining with definitions of $V(x)$ and $\Phi(x, \sigma)$ yield (3.14). $\square$

The following lemma shows that the loop in Step 4.2 is finite.

Lemma 3.4.   *Let Assumptions (A1)-(A3) hold, $\eta \in \left(0, \frac{1}{2}\right)$ and $\bar{x}$ satisfies $0 \notin \partial V(\bar{x})$. Then, there exist a scalar $\bar{t} > 0$ and a neighborhood $N(\bar{x})$ of $\bar{x}$ such that for any $x \in N(\bar{x})$ and any d feasible for the problem $\underline{QP}(x, B, \sigma, \rho)$ with $\gamma \rho_{\min} \leq \sigma \leq \rho \leq \bar{\rho}$, it holds that*

$$
V(x + td) - V(x) \leq t\eta(\Phi(x, \sigma) - V(x)), \quad (3.17)
$$

for all $t \in (0, \bar{t}]$.

Proof.   It follows from Lemma 3.1 that there exists a neighborhood $N(\bar{x})$ and $\bar{\epsilon}(\bar{x}) > 0$ such that $\Phi(x, \sigma) - V(x) < -\bar{\epsilon}(\bar{x})$ whenever $x \in N(\bar{x})$. Combining this with (3.14), we have

$$
V(x + td) - V(x) \leq t\eta(\Phi(x, \sigma) - V(x)) - t(1 - \eta)\bar{\epsilon}(\bar{x}) + \frac{1}{2}t^2 mnM\rho^2.
$$

Hence, the inequality (3.17) holds for all $t \in (0, \bar{t}(\bar{x})]$, where

$$\bar{t}(\bar{x}) := \frac{2(1 - \eta)\bar{\epsilon}(\bar{x})}{mnM\bar{\rho}^2}. \qquad \square$$

Hence, if $0 \notin \partial V(x_k)$, (2.13) follows taking $\bar{x} = x = x_k$ and $d = d_s$. It implies that the loop in Step 4.2 is finite.

The following lemma shows that the iterate sequence $\{x_k\}$ approaches a stationary point of $V(x)$ when Step 4.2 of Algorithm 2.1 is invoked infinitely many times.

Lemma 3.5 *If Step* 4.2 *of Algorithm* 2.1 *is invoked infinitely many times, then there exists an accumulation point $\bar{x}$ of $\{x_k\}$ such that the sequence $\{U_k\}$ converges to $V(\bar{x})$, where $0 \in \partial V(\bar{x})$, i.e., $\bar{x}$ is a stationary point of $V(x)$.*

Proof. Since Step 4.2 is invoked infinitely many times. By Step 5 U_{k+1} is reset by $V(x_k + t_k d_k)$ infinitely. From Step 2 and 4.2, we note that if Step 4.2 is invoked at iteration k, the radius ρ in $\mathrm{QP}(x_k, B_k, \sigma_k, \rho)$ associated with d_s is greater than or equal to $\rho_{\min}$. Define $\mathcal{K} = \{k \mid U_{k+1} \text{ is reset by } V(x_k + t_k d_k)\}$. Obviously, $\mathcal{K}$ is an infinite set. The inequality (2.13) ensures $V(x_{k+1}) = U_{k+1}$ for all $k \in \mathcal{K}$. The upper bound condition (2.8) ensures $V(x_{k+1}) \leq U_k = U_{k+1}$ for all $k \notin \mathcal{K}$. Therefore, $V(x_k) \leq U_k$ for all k. This together with (2.13) yields $U_{k+1} \leq V(x_k) \leq U_k$ for all $k \in \mathcal{K}$. For $k \notin \mathcal{K}$, $U_{k+1} = U_k$. Therefore, $\{U_k\}$ is a monotonically decreasing sequence and also has a lower bound zero. Then there exists an accumulation point $\bar{x}$ of $\{x_k\}$ such that $\{U_{k+1}\}_{\mathcal{K}} \to V(\bar{x})$. If $0 \notin \partial V(\bar{x})$, by Lemma 3.1, there exists a neighborhood $N(\bar{x})$ of $\bar{x}$ and $\bar{\epsilon}(\bar{x}) > 0$, such that

$$\Phi(x_k, \sigma_k) - V(x_k) < -\bar{\epsilon}(\bar{x}), \qquad (3.18)$$

whenever $x_k \in N(\bar{x})$. This together with (2.13) yields

$$U_{k+1} - U_k \leq U_{k+1} - V(x_k) < -t_k \eta \bar{\epsilon}(\bar{x}), \qquad (3.19)$$

for $k \in \mathcal{K}$ and $x_k \in N(\bar{x})$. By the mechanism of Algorithm 2.1 and Lemma 3.4, we have $t_k \geq r\bar{t}(\bar{x})$. It follows with (3.19) that

$$U_{k+1} - U_k < -r\eta\bar{t}(\bar{x})\bar{\epsilon}(\bar{x})$$

for $k \in \mathcal{K}$ and $x_k \in N(\bar{x})$. Letting $k \in \mathcal{K}$ tend to infinity in above inequality, the limit in the left-hand side is zero while the limit in the right-hand side is less than zero, which is a contradiction. Therefore $0 \in \partial V(\bar{x})$.   □

Lemma 3.6.   *Consider sequences $\{V(x_k)\}$ and $\{f(x_k)\}$ such that $f(x_k)$ is monotonically decreasing and bounded below. If for all k, either $V(x_{k+1}) - V(x_k) \leq -\gamma_1 V(x_{k+1})$ or $f(x_k) - f(x_{k+1}) \geq \gamma_2 V(x_{k+1})$ holds, where constants γ_1 and γ_2 are from (2.7a) and (2.7b), then $V(x_k) \to 0$, for $k \to +\infty$.*

Proof.   See [8, Lemma 1].   □

Lemma 3.7.   *Suppose Assumptions $(A1)$-$(A3)$ hold. If there exists an infinite sequence of iterates $\{x_k\}$ on which $(V(x_k), f(x_k))$ is added to the filter, where $V(x_k) > 0$ and $\{f(x_k)\}$ is bounded below, then $V(x_k) \to 0$ as $k \to +\infty$.*

Proof.   Since inequalities (2.7a) and (2.7b) are the same as [8, (2.6)], the conclusion follows from Lemma 3.6 and [8, Corollary].   □

Lemma 3.8.   *Suppose Assumptions $(A1)$-$(A3)$ hold. Let d be a feasible point of $\underline{QP}(x_k, B_k, \sigma_k, \rho)$. If $\Phi(x_k, \sigma_k) = 0$, it then follows that*

$$\Delta f(d) \geq \Delta q(d) - n\rho^2 M, \tag{3.20}$$

$$V(x_k + d) \leq \frac{1}{2} mn\rho^2 M. \tag{3.21}$$

Proof.   The proof of this lemma is very similar to the proof [8, Lemma 3]. By Taylor's theorem, we have

$$f(x_k + d) = f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T \nabla^2 f(y) d,$$

where y denotes some point on the line segment from x_k to $x_k + d$. This together with (2.9) and (2.10) implies

$$\Delta f(d) = \Delta q(d) + \frac{1}{2} d^T \left(B_k - \nabla^2 f(y) \right) d.$$

Then (3.20) follows from the boundedness of $\nabla^2 f(y)$ and B_k, and $\|d\|_\infty \le \rho$. Since $\Phi(x_k, \sigma_k) = 0$, $\underline{QP}(x_k, B_k, \sigma_k, \rho)$ reduces to (1.1). For $i \in \mathcal{E} \cup \mathcal{I}$, it follows that

$$c_i(x_k + d) = c_i(x_k) + \nabla c_i(x_k)^T d + \frac{1}{2} d^T \nabla^2 c_i(y_i) d,$$

where y_i denotes some point on the line segment from x_k to $x_k + d$. By feasibility of d,

$$|c_i(x_k + d)| \le \frac{1}{2} n\rho^2 M, \ i \in \mathcal{E}$$

and

$$c_i(x_k + d) \le \frac{1}{2} n\rho^2 M, \ i \in \mathcal{I}.$$

follow in a similar way. It follows with the definition of $V(x)$ that (3.21) holds. $\square$

Lemma 3.9. *Suppose Assumptions $(A1)$-$(A3)$ hold. Let d be a feasible point of $\underline{QP}(x_k, B_k, \sigma_k, \rho)$ with $\Phi(x_k, \sigma_k) = 0$. Then $x_k + d$ is acceptable to the filter if $\rho^2 \le \frac{2\tau_k}{mnM(1+\gamma_1)}$, where $\tau_k := \min\limits_{k \in \mathcal{F}_k} \{V(x_k)\}$.* $\square$

Proof. It follows from (3.21) that $V(x_k + d) - \tau_k \le -\gamma_1 V(x_k + d)$ holds if $\rho^2 \le \frac{2\tau_k}{mnM(1+\gamma_1)}$. By the definition of τ_k, (2.7a) is satisfied. Hence $x_k + d$ is acceptable to the filter. $\square$

In order to prove that the iterate sequence generated from Algorithm 2.1 converges to a KKT point for the problem (NLP), some constraint qualification should be required, such as the well-known MFCQ. Thus we review its definition as follows.

Definition 5 (See [2]).   MFCQ is said to be satisfied at x, with respect to the underlying constraint system $c_\mathcal{E}(x) = 0, c_\mathcal{I}(x) \le 0$, if there is a $z \in \mathbb{R}^n$ such that the gradients $\nabla c_i(x), \ i \in \mathcal{E}$ are linearly independent and the following systems

$$\nabla c_i(x)^T z = 0, \ i \in \mathcal{E},$$
$$\nabla c_i(x)^T z < 0, \ i \in \{i : c_i(x) = 0, \ i \in \mathcal{I}\}$$

are satisfied.

Proposition 3.10. *Suppose that MFCQ is satisfied at $x^* \in X$, then there exists a neighborhood $N(x^*)$ of x^* such that*

1. *MFCQ is satisfied at every point in $N(x^*) \cap X$;*

2. *inequality*

$$\sup_k \left\{ \|\mu_k\|_1 + \|\lambda_k\|_1 : x_k \in N(x^*), \sigma_k \in (\rho_{\min}, \rho_k), \rho_k \in (\rho_{\min}, \bar{\rho}] \right\} < +\infty.$$

    holds, where the vectors λ_k, μ_k generated by $\underline{QP}(x_k, B_k, \sigma_k, \rho_k)$ are multiplier vectors associated with $x_k \in N(x^*)$.

Proof.   The first result is established in [20, Theorem 3]. The second result is established in [2, Theorem 5.1].   □

If MFCQ does not hold at a feasible x^*, then the second statement of Proposition 3.10 cannot be guaranteed. All feasible points of the problem (NLP) at which MFCQ does not hold will be called non-MFCQ points.

The following lemma shows that (1.1) is consistent when x_k approaches a feasible point at which MFCQ holds and both the quadratic reduction and the actual reduction of the objective function have sufficient reduction. Its proof is similar to that of [8, Lemma 5].

Lemma 3.11.   *Suppose Assumptions $(A1)$-$(A3)$ hold and let $x^* \in S$ be a feasible point of problem (NLP) at which MFCQ holds but which is not a KKT point. Then there exists a neighborhood N° of x^* and positive constants ϵ, ν and $\overline{\kappa}$, such that for all $x_k \in S \cap N^\circ$ and all ρ for which*

$$\nu V(x_k) \leq \rho \leq \overline{\kappa}, \tag{3.22}$$

it follows that $\underline{QP}(x_k, B_k, \sigma_k, \rho)$ with $\Phi(x_k, \sigma_k) = 0$ has a feasible solution d at which the predicted reduction satisfies

$$\triangle q(d) \geq \frac{1}{3}\rho\epsilon, \tag{3.23}$$

the sufficient reduction condition (2.12) holds, and the actual reduction satisfies

$$\triangle f(d) \geq \gamma_2 V(x_k + d). \tag{3.24}$$

Proof. Since x^* is a feasible point at which MFCQ holds but it is not a KKT point, it follows that the vectors $\nabla c_i(x^*)$, $i \in \mathcal{E}$ are linearly independent, and there exists a vector s^* that satisfies

$$\nabla f(x^*)^T s^* < 0, \tag{3.25}$$

$$\nabla c_i(x^*)^T s^* = 0, \; i \in \mathcal{E}, \tag{3.26}$$

$$\nabla c_i(x^*)^T s^* < 0, \; i \in \mathcal{A}(x^*) = \{i : c_i(x^*) = 0, \; i \in \mathcal{I}\}, \tag{3.27}$$

where $\|s^*\| = 1$. Let $\nabla c_{\mathcal{E}}(x_k)^+ = \left(\nabla c_{\mathcal{E}}(x_k)^T \nabla c_{\mathcal{E}}(x_k)\right)^{-1} \nabla c_{\mathcal{E}}(x_k)^T$ and $\nabla c_{\mathcal{E}}(x_k)$ denote the matrix with columns $\nabla c_i(x_k)$, $i \in \mathcal{E}$. It follows from linear independence and continuity that there exists a neighborhood of x^* in which $\nabla c_{\mathcal{E}}(x_k)^+$ is bounded. Let $P = -\nabla c_{\mathcal{E}}(x_k)^{+T} c_{\mathcal{E}}(x_k)$ and $s = \left(I - \nabla c_{\mathcal{E}}(x_k)\nabla c_{\mathcal{E}}(x_k)^+\right)s^* / \|\left(I - \nabla c_{\mathcal{E}}(x_k)\nabla c_{\mathcal{E}}(x_k)^+\right)s^*\|$ if $\mathcal{E}$ is not empty, otherwise $P = 0$ and $s = s^*$. Let $p = \|P\|$. It follows from (3.25) and (3.27) by continuity that there exists a (smaller) neighborhood $\mathcal{N}^*$ and a constant $\epsilon > 0$ such that

$$\nabla f(x_k)^T s^* < -\epsilon \quad \text{and} \quad \nabla c_i(x_k)^T s^* < -\epsilon, \; i \in \mathcal{A}(x^*), \tag{3.28}$$

when $x_k \in \mathcal{N}^*$. By definition of P, it follows that $p = \mathcal{O}(V(x_k))$, and thus we can choose the constant ν in (3.22) sufficiently large so that $\rho > p$ for all $x_k \in \mathcal{N}^*$.

We now consider the solution of (1.1). The line segment defined by

$$d_\alpha = P + \alpha(\rho - p)s, \quad \alpha \in [0, 1], \tag{3.29}$$

for a fixed value of $\rho > p$. Since $\rho > p$, and P and s are orthogonal, it implies

$$\|d_1\|_\infty \leq \|d_1\| = \sqrt{p^2 + (\rho - p)^2} = \sqrt{\rho^2 - 2\rho p + 2p^2} \leq \rho. \tag{3.30}$$

From (3.29) and the definitions of P, s, d_α satisfies the equality constraints $c_{\mathcal{E}}(x_k) + \nabla c_{\mathcal{E}}(x_k)^T d_\alpha = 0$ of (1.1) for all $\alpha \in [0, 1]$.

If $x_k \in \mathcal{N}^* \cap S$ and $i \in \mathcal{I}\backslash\mathcal{A}(x^*)$, then there exists positive constants $\bar{c}$ and $\bar{a}$, independent of ρ, such that

$$c_i(x_k) \leq -\bar{c} \quad \text{and} \quad \nabla c_i(x_k)^T s \leq \bar{a}$$

for all vectors s such that $\|s\|_\infty \leq 1$, by the continuity of $c_i(x_k)$ and bounded-ness of $\nabla c_i(x_k)$ on S. It follows that

$$c_i(x_k) + \nabla c_i(x_k)^T d \leq -\bar{c} + \rho\bar{a}, \quad i \in \mathcal{I}\backslash\mathcal{A}(x^*)$$

for all vectors d such that $\|d\|_\infty \leq \rho$. Therefore, inactive constraints do not affect the solution of (1.1) if ρ satisfies $\rho \leq \bar{c}/\bar{a}$.

For active inequality constraints $i \in \mathcal{A}(x^*)$, it follows from (3.28) and (3.29) that $c_i(x_k) + \nabla c_i(x_k)^T d_1 = c_i(x_k) + \nabla c_i(x_k)^T P + (\rho - p)\nabla c_i(x_k)^T s \leq c_i(x_k) + \nabla c_i(x_k)^T P - (\rho - p)\epsilon \leq 0$ if $\rho \geq p + (c_i(x_k) + \nabla c_i(x_k)^T P)/\epsilon$. We obtain from the definition of P that

$$\exists \nu_0 > 0 \text{ s.t.} \frac{1}{\epsilon}\Big(c_i(x_k) + \nabla c_i(x_k)^T P\Big) \leq \nu_0 V(x_k).$$

Thus we can choose the constant ν in (3.22) sufficiently large so that $c_i(x_k) + \nabla c_i(x_k)^T d_1 \leq 0$, $i \in \mathcal{A}(x^*)$. Therefore d_1 is feasible in (1.1) with respect to the active inequality constraints, and hence to all the constraints by above results. Combining with the fact that (1.1) is equivalent to $\underline{QP}(x_k, B_k, \sigma_k, \rho)$ with $\Phi(x_k, \sigma_k) = 0$, we obtain that $\underline{QP}(x_k, B_k, \sigma_k, \rho)$ with $\Phi(x_k, \sigma_k) = 0$ is consistent for all $x_k \in \mathcal{N}^*$ and all ρ satisfying (3.22) for any value of $\bar{\kappa} \leq \bar{c}/\bar{a}$.

Next we aim to obtain a bound on the predicted reduction $\Delta q(d)$. We note that $q(0) - q(d_1) = -\nabla f(x_k)^T(P + (\rho - p)s) - \frac{1}{2}d_1^T B_k d_1$. Using (3.28), bounds on B_k and P, and $\rho > p = \mathcal{O}(V(x_k))$, we have

$$q(0) - q(d_1) \geq (\rho - p)\epsilon - \frac{1}{2}\rho^2 nM + \mathcal{O}(V(x_k))$$

$$= \rho\epsilon - \frac{1}{2}\rho^2 nM + \mathcal{O}(V(x_k)).$$

If $\rho < \frac{\epsilon}{nM}$, then $q(0) - q(d_1) \geq \frac{1}{2}\rho\epsilon + \mathcal{O}(V(x_k))$. Since d_1 is feasible and $p = \mathcal{O}(V(x_k))$, it follows that the predicted reduction (2.9) satisfies

$$\Delta q(d) \geq \Delta q(d_1) \geq \frac{1}{2}\rho\epsilon + \mathcal{O}(V(x_k)) \geq \frac{1}{2}\rho\epsilon - \xi V(x_k)$$

for some ξ sufficiently large and independent of ρ. Hence, (3.23) is satisfied if $\rho \geq 6\xi V(x_k)/\epsilon$. This condition can be achieved by making the constant ν in (3.22) sufficiently large.

Next we aim to prove (3.24). By (3.20) and (3.23), we have

$$\frac{\triangle f(d)}{\triangle q(d)} \geq 1 - \frac{n\rho^2 M}{\triangle q(d)} \geq 1 - \frac{3n\rho^2 M}{\rho\epsilon} = 1 - \frac{3n\rho M}{\epsilon}.$$

Then, if $\rho \leq (1-\eta)\epsilon/(3nM)$, it follows that (2.12) holds. By (2.12), (3.21) and (3.23), we have $f(x_k) - f(x_k + d) - \gamma_2 V(x_k + d) = \triangle f(d) - \gamma_2 V(x_k + d) \geq \frac{1}{3}\eta\rho\epsilon - \frac{1}{2}\gamma_2 mn\rho^2 M \geq 0$ if $\rho \leq \frac{2}{3}\eta\epsilon/(\gamma_2 mnM)$. Therefore, we may define the constant $\bar{\kappa}$ in (3.22) to be the least of $\frac{2}{3}\eta\epsilon/(\gamma_2 mnM)$ and the values $(1-\eta)\epsilon/(3nM)$, $\rho < \frac{\epsilon}{nM}$ and $\bar{c}/\bar{a}$, as required earlier in the proof. $\quad\square$

Lemma 3.12. *Suppose Assumptions $(A1)$-$(A3)$ hold and let $x^* \in S$ be a feasible point of problem (NLP) at which MFCQ holds but which is not a KKT point. Then there exists a neighborhood N° of x^* and a positive constant ν, such that for all $x_k \in S \cap N^\circ$, all ρ and all σ_k for which*

$$\nu V(x_k) \leq \sigma_k < \rho, \tag{3.31}$$

it follows that $\Phi(x_k, \sigma_k) = 0$ and $\underline{QP}(x_k, B_k, \sigma_k, \rho)$ has a feasible solution d at which the predicted reduction satisfies

$$\triangle q(d) > 0. \tag{3.32}$$

Proof. By Lemma 3.11, there exists a neighborhood N^1 of x^* and positive constants $\nu, \bar{\kappa}$ such that for all $x_k \in S \cap N^\circ$ and all ρ and σ_k for which $\nu V(x_k) \leq \rho \leq \bar{\kappa}$, it follows that the QP subproblem (1.1) has a feasible solution d at which the predicted reduction satisfies (3.32). Since the global optimality of d ensures that $\triangle q(d)$ decreases monotonically as ρ decreases, the predicted reduction satisfies (3.32) whenever $\nu V(x_k) \leq \rho$.

From the earlier proof, if $\nu V(x_k) \leq \sigma_k \leq \bar{\kappa}$, we also have that the QP subproblem (1.1) is consistent by taking $\rho = \sigma_k$. It means that the problem (2.4) has the optimal value 0. If σ_k increases to a value larger than $\bar{\kappa}$, then the feasible region of the problem (2.4) with $\rho = \sigma_k$ is also enlarged correspondingly and the optimal value is still 0. Therefore, $\Phi(x_k, \sigma_k) = 0$ whenever $\nu V(x_k) \leq \sigma_k$.

The above two conclusions complete the proof. $\quad\square$

As a last preparation for the proof of Theorem 3.14, we proceed as in [8] and show that the loops 2-(4.1)-2 and 2-(4.3)-2 terminate finitely.

Lemma 3.13.   *Suppose Assumptions (A1)-(A3) hold, then the loops 2-(4.1)-2 and 2-(4.3)-2 terminate finitely.*

Proof.   If x_k is a KKT point for the problem (NLP), then $d = 0$ is the solution of QP$(x_k, B_k, \sigma_k, \rho)$, and Algorithm 2.1 terminates, so do the loops 2-(4.1)-2 and 2-(4.3)-2. If $V(x_k) > 0$ and $\Phi(x_k, \sigma_k) - V(x_k) = 0$, then stop by Step 3 of Algorithm 2.1. In fact, it follows from [2, Lemma 2.1] that $V(x_k) > 0$ and $0 \in \partial V(x_k)$, which means that x_k is a stationary point of $V(x)$ and not feasible for the problem (NLP). If both above situations do not occur, and the loops 2-(4.1)-2 and 2-(4.3)-2 do not terminate finitely, then $\rho \to 0$ from Algorithm 2.1. There are two cases to be considered.

Case (i). $V(x_k) > 0$ and $0 \notin \partial V(x_k)$.

(a) If $i \in \mathcal{E}$ and $c_i(x_k) > 0$, then for all d ($\|d\|_\infty \le \rho$),

$$c_i(x_k) + \nabla c_i(x_k)^T d \ge c_i(x_k) - \rho\|\nabla c_i(x_k)\| > 0, \tag{3.33}$$

if either $\|\nabla c_i(x_k)\| = 0$ or $\rho < \frac{c_i(x_k)}{\|\nabla c_i(x_k)\|}$. Thus, for sufficiently small ρ, the equality constraints cannot be satisfied and (1.1) is inconsistent. Therefore, $\Phi(x_k, \sigma_k) > 0$.

(b) If $i \in \mathcal{I}$ and $c_i(x_k) > 0$, a similar conclusion is obtained.

(c) If $i \in \mathcal{E}$ and $c_i(x_k) < 0$, a similar conclusion is also obtained.

So, for some ρ sufficiently small, $\Phi(x_k, \sigma_k) > 0$ and $0 \notin \partial V(x_k)$. By Step 4, the procedure executes at 4.2 of Step 4. It follows from Lemma 3.4 that (2.13) holds. By the mechanism of Algorithm 2.1, the loop 2-(4.1)-2 terminates finitely.

Case (ii). $V(x_k) = 0$. Then $\Phi(x_k, \sigma_k) = 0$.

For the inactive constraints at x_k, by a similar argument, it will still be inactive for sufficiently small ρ. Thus, we only need to consider the active constraints.

Since x_k is not a KKT point, there exists a vector s, $\|s\| = 1$, and a scalar $\epsilon > 0$ such that $\nabla f(x_k)^T s < -\epsilon$, $\nabla c_i(x_k)^T s = 0$, $i \in \mathcal{E}$ and $\nabla c_i(x_k)^T s \le 0$, $i \in \mathcal{A}(x_k) = \{i : c_i(x_k) = 0, i \in \mathcal{I}\}$. We note that $q(0) - q(\rho s) = -\rho\nabla f(x_k)^T s -$

$\frac{1}{2}\rho^2 s^T B_k s$. Using bound on B_k, we have $q(0) - q(\rho s) \geq \rho\epsilon - \frac{1}{2}\rho^2 nM$. If $\rho < \frac{\epsilon}{nM}$, then $q(0) - q(\rho s) \geq \frac{1}{2}\rho\epsilon$. Since ρs is feasible for $\underline{QP}(x_k, B_k, \sigma_k, \rho)$, it follows that the predicted reduction (2.9) satisfies

$$\Delta q(d) \geq \Delta q(\rho s) \geq \frac{1}{2}\rho\epsilon,$$

where d is the solution of $\underline{QP}(x_k, B_k, \sigma_k, \rho)$. If $\rho \leq \frac{(1-\eta)\epsilon}{2nM}$, it follows with (3.20) that $\Delta f(d) \geq \eta \Delta q(d) > 0$. So the sufficient reduction condition (2.12) for an f-*type* iteration is satisfied. Moreover, by (3.21), we have

$$
\begin{aligned}
f(x_k) - f(x_k + d) - \gamma_2 V(x_k + d) &= \Delta f(d) - \gamma_2 V(x_k + d) \\
&\geq \frac{1}{2}\eta\rho\epsilon - \frac{1}{2}\gamma_2 mn\rho^2 M \\
&\geq 0
\end{aligned}
$$

if $\rho \leq \eta\epsilon/(\gamma_2 mnM)$. Thus, $x_k + d$ is acceptable to x_k. Of course, the upper bound condition (2.8) is satisfied.

Finally, it follows with Lemma 3.9 that for a sufficiently small ρ, an f-*type* iteration is generated and the loop 2-(4.3)-2 terminates finitely. $\square$

Lemma 3.13 together with Lemma 3.4 implies that Algorithm 2.1 is well-defined. We are now able to adapt [8, Theorem 7] to Algorithm 2.1.

Theorem 3.14. *Let Assumptions* $(A1)$-$(A3)$ *hold.* $\{x_k\}$ *is a sequence generated by Algorithm* 2.1*, then there is an accumulation point that is a KKT point for the problem (NLP), or a non-MFCQ point for the problem (NLP) or a stationary point of* $V(x)$ *that is infeasible for the problem (NLP).*

Proof.   Since the loops 2-(4.1)-2 and 2-(4.3)-2 are finite, we only need to consider that the loop 2-6-2 is infinite. All iterates lie in S, which is bounded, so it follows that the iteration sequence has at least one accumulation point.

Case (i). Step 4.2 of Algorithm 2.1 is invoked finitely many times. Then Step 4.2 of Algorithm 2.1 is not invoked for all sufficiently large k.

Sub-case (i). There are infinite V-*type* steps in the main iteration sequence. Then, from Lemma 3.7, $V(x_k) \to 0$ and $\tau_k \to 0$ on this subsequence. Moreover,

there exists a subsequence indexed by $k \in S_\phi$ of V-type iterations for which $x_k \to x^*$, $V(x_k) \to 0$ and $\tau_{k+1} < \tau_k$. One consequence is that x^* is a feasible point for the problem (NLP). If MFCQ is not satisfied at x^*, then x^* is a non-MFCQ point for the problem (NLP). We therefore assume that MFCQ is satisfied at x^* and consider the assumption that x^* is not a KKT point (to be contradicted).

From Lemma 3.12, we know that $\Phi(x_k, \sigma_k) = 0$ and the step d is an f-type step for all sufficiently large k if $\rho \geq \gamma V(x_k)$. So, $x_k + d$ is acceptable to the filter for sufficiently large k from Lemma 3.9 if $\rho^2 \leq \frac{2\tau_k}{mnM(1+\gamma_1)}$. Thus, by Lemma 3.11 we deduce that if ρ satisfies

$$\nu V(x_k) \leq \rho \leq \min \left\{ \sqrt{\frac{2\tau_k}{mnM(1+\gamma_1)}}, \kappa \right\}, \tag{3.34}$$

then k is an f-type iterate for sufficiently large k.

Now we need to show that a value of $\rho \geq \nu V(x_k)$ can be found in the loop 2-(4.3)-2 such that k is an f-type iterate for sufficiently large k. Since $\tau_k \to 0$ when $k(\in S_\phi) \to +\infty$, the range (3.34) becomes

$$\nu V(x_k) \leq \rho \leq \sqrt{\frac{2\tau_k}{mnM(1+\gamma_1)}}. \tag{3.35}$$

From the definition of S_ϕ, we know that $\tau_{k+1} = V(x_k) < \tau_k$. Because of the square root, the upper bound in (3.35) can be greater than twice the lower bound. From Algorithm 2.1, a value $\rho \geq \rho_{\min}$ is chosen at the beginning of each iteration, then it will be greater than the upper bound in (3.35) for sufficiently large k. We can see that successively halving ρ in the loops 2-(4.1)-2 and 2-(4.3)-2 will eventually locate in the range of (3.35), or the right of this interval. Lemma 3.12 implies that it is not possible for any value of $\rho \geq \nu V(x_k)$ to produce an V-type step. Thus if $k(\in S_\phi)$ is sufficiently large, an f-type iteration is generated, which contradicts the definition of S_ϕ. So x^* is a KKT point.

Sub-case (ii). There are finite V-type steps in the main iteration sequence. Then, there exists a positive integer K, such that for all $k > K$, x_k is an f-type iteration. So $\{f(x_k)\}_{k>K}$ is strictly monotonically decreasing. It follows from Lemma 3.6 that $V(x_k) \to 0$ and hence, any accumulation point x^* of the main iteration

sequence is a feasible point. Since $f(x)$ is bounded, we get that

$$\sum_{k > K} \Delta f(d_k) < +\infty. \tag{3.36}$$

It follows that $\Delta f(d_k) \to 0, k > K$ as $k \to +\infty$.

Now we assume that MFCQ is satisfied at x^* and x^* is not a KKT point. Similarly to sub-case (i), when

$$\nu V(x_k) \leq \rho \leq \bar{\kappa}, \tag{3.37}$$

f-type iterations are generated. The right-hand side of (3.37) is a constant, independent of k. Since the upper bound of (3.37) is a constant and the lower bound converges to zero, the upper bound must be more than twice the lower bound. So a value of ρ will be located in this interval, or a value to the right of this interval. Hence, $\rho \geq \min\{\frac{1}{2}\bar{\kappa}, \rho_{\min}\}$. Since the global optimality of d ensures that $\Delta q(d)$ decreases monotonically as ρ decreases, by Lemma 3.11, $\Delta q(d) \geq \frac{1}{3}\epsilon \min\{\frac{1}{2}\bar{\kappa}, \rho_{\min}\}$ holds even if ρ is greater than $\bar{\kappa}$. This together with (2.12) yields that $\Delta f(d_k) \geq \frac{1}{3}\eta\epsilon \min\{\frac{1}{2}\bar{\kappa}, \rho_{\min}\}$, which is a contradiction. Thus, x^* is a KKT point.

Case (ii). Step 4.2 of Algorithm 2.1 is invoked infinitely many times. Then it follows from Lemma 3.5 that there exists an accumulation point x^* of $\{x_k\}_{\mathcal{K}}$ such that $0 \in \partial V(x^*)$, where $\mathcal{K}$ is an infinite index set satisfying that any $k \in \mathcal{K}$ is a V-type iteration number and $\mathcal{U}_{k+1} = V(x_k + t_k d_k)$. If $V(x^*) > 0$, then x^* is a stationary point of $V(x)$ that is infeasible for the problem (NLP). If $V(x^*) = 0$ and MFCQ fails to be satisfied at x^*, then x^* is a non-MFCQ point for the problem (NLP). If $V(x^*) = 0$ and MFCQ holds at x^*, then we prove that x^* is a KKT point for the problem (NLP).

Suppose x^* is not a KKT point for the problem (NLP), but $V(x^*) = 0$, and MFCQ holds at x^*. Then, from Lemma 3.5, $\mathcal{U}_k \to 0$ on this sequence. Therefore, for all $k \in \mathcal{K}$, $V(x_k) \leq \mathcal{U}_k \to 0$ and $\mathcal{U}_{k+1} = V(x_k + t_k d_k)$. By Step 4.4, at any new iteration the first trial radius ρ is greater than or equal to $\rho_{\min}$. So $\sigma_k \geq \gamma \rho_{\min}$ is true from the fact $\sigma_k \in (\gamma\rho, \rho)$. Combining this with the fact $V(x_k) \to 0$ as $k(\in \mathcal{K}) \to \infty$ yields $\nu V(x_k) \leq \sigma_k < \rho$ for all sufficiently

large k. It follows with Lemma 3.12 that $\Phi(x_k, \sigma_k) = 0$ and (3.32) is satisfied for all sufficiently large $k \in \mathcal{K}$. Similar to Sub-case (i) of Case (i), k is an f-type iterate for sufficiently large $k \in \mathcal{K}$. Hence Step 4.2 could not be invoked at this iteration and $\mathcal{U}_{k+1} = \mathcal{U}_k$ which contradicts the definition of $\mathcal{K}$. So x^* is a KKT point. $\square$

4   Numerical results

We give some numerical results of Algorithm 2.1 coded in Matlab for the constrained optimization problems. The details about the implementation are described as follows:

(a) Termination criteria. Algorithm 2.1 stops if

$$V(x_k) \le \varepsilon \qquad \text{and} \qquad \|\nabla f(x_k) + \nabla c_{\mathcal{E}}(x_k)\mu_k + \nabla c_{\mathcal{I}}(x_k)\lambda_k\|_\infty \le \varepsilon.$$

(b) Update B_k. Initiate $B_0 = I$, where I is the identity matrix with appropriate dimension. Update B_k by the BFGS formula with Powell's modifications [19], which is described as following:

set

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k},$$

where

$$y_k = \begin{cases} \hat{y}_k, & \hat{y}_k^T s_k \ge 0.2 s_k^T B_k s_k, \\ \theta^k \hat{y}_k + (1 - \theta^k) B_k s_k, & \text{otherwise,} \end{cases}$$

and

$$\begin{cases} s_k = x_{k+1} - x_k, \\ \hat{y}_k = \nabla f(x_{k+1}) - \nabla f(x_k) + \big(\nabla c_{\mathcal{E}}(x_{k+1}) - \nabla c_{\mathcal{E}}(x_k)\big)\mu_k \\ \qquad + \big(\nabla c_{\mathcal{I}}(x_{k+1}) - \nabla c_{\mathcal{I}}(x_k)\big)\lambda_k, \\ \theta^k = 0.8 \, s_k^T B_k s_k / (s_k^T B_k s_k - s_k^T \hat{y}_k). \end{cases}$$

(c) The parameters are chosen as: $\rho_0 = 5.0$, $\rho_{\min} = 10^{-4}$, $\eta = 0.1$, $\gamma_1 = \gamma_2 = 2 \times 10^{-4}$, $\varepsilon = 10^{-6}$, $\mathcal{U}_0 = 10 \max\{1, V(x_0)\}$, $\sigma_k = 0.9\rho$ for all $k = 0, 1, 2 \cdots$ and all $\rho > 0$.

Firstly, the following examples is from [29]. Since EXAMPLE 3 in [29] is unbounded below, we do not give it here. The numerical results of other examples are described in the following.

EXAMPLE 1

$$\begin{aligned} \min \quad & f(x) = x - \tfrac{1}{2} + \tfrac{1}{2}\cos^2 x, \\ \text{s.t.} \quad & x \geq 0. \end{aligned} \tag{4.38}$$

$x^* = 0$, $f(x^*) = 0$. The iteration number of Algorithm 2.1 is 2.

EXAMPLE 2

$$\begin{aligned} \min \quad & f(x) = \sum_{i=1}^{4} x_i^2, \\ \text{s.t.} \quad & 6 - \sum_{i=1}^{4} x_i^2 \leq 0. \end{aligned} \tag{4.39}$$

$x^* = (1.224745, 1.224745, 1.224745, 1.224745)^T$, $f(x^*) = 6$. The iteration number of Algorithm 2.1 is 6.

EXAMPLE 4

$$\begin{aligned} \min \quad & f(x) = \tfrac{4}{3}\left(x_1^2 - x_1 x_2 + x_2^2\right)^{\frac{3}{4}} - x_3, \\ \text{s.t.} \quad & -x_i \leq 0, \ i = 1, 2, 3, \\ & x_3 \leq 2. \end{aligned} \tag{4.40}$$

$x^* = (0, 0, 2)^T$, $f(x^*) = -2$. The iteration number of Algorithm 2.1 is 4.

Compared with the results in [29, 28], the computation at each iteration in this paper is less than those in them.

Except for above numerical experiments, we also test some examples from [12]. We compare these numerical results with those in [13]. The detailed results of the numerical tests on these problems are summarized in Table 1, where NIT, NF, and NG represent the numbers of iterations, function, and gradient calculations, respectively. The problems are numbered in the same way as in Hock and Schittkowski [12]. For example, HS022 represents problem 22 in Hock and Schittkowski [12].

				Algorithm 2.1	Algorithm in [13]
Problem	n	m_1	m	NIT-NF-NG	NIT-NF-NG
HS007	2	1	1	10-11-11	9-18-10
HS014	2	1	2	5-6-6	5-6-6
HS022	2	0	2	4-6-5	5-7-7
HS038	4	0	8	24-29-25	38-64-39
HS043	4	0	3	11-16-12	12-15-15
HS052	5	3	3	6-7-7	5-9-9
HS063	3	2	5	8-9-9	7-8-8
HS086	5	0	15	5-7-6	4-7-5
HS113	10	0	8	13-22-14	14-20-15

Table 1 – Numerical results.

For Problem HS022, $\Phi(x_0) = 0.3941$ implies that QP problem is inconsistent at the first iteration. After this, all QP problems are consistent in the subsequent iterations. Similarly, for Problem HS063, only $\Phi(x_0) = 1.25$ implies the same situation as Problem HS022. Except for Problem HS022 and HS063, for all the other problems in Table 1, their QP problems are consistent in all iterations.

The above analysis shows that our algorithm deals with inconsistent QP problem effectively and is comparable to the algorithm in [13]. So, the numerical tests confirm the robustness of our algorithm.

REFERENCES

[1] C. Audet and J.E. Dennis. *A pattern search filter method for nonlinear programming without derivatives.* SIAM Journal on Optimization, **14** (2004), 980–1010.

[2] J.V. Burke and S.P. Han. *A robust sequential quadratic programming method.* Mathematical Programming, **43** (1989), 277–303.

[3] S. Bellavia, M. Macconi and B. Morini. *STRSCNE: a scaled trust region solver for constrained nonlinear equations.* Computational Optimization and Applications, **28** (2004), 31–50.

[4] S. Bellavia, M. Macconi and B. Morini. STRSCNE: http://ciro.de.unifi.it/STRSCNE/, (2007).

[5] H.Y. Benson, D.F. Shanno and R. Vanderbei. *Interior-point methods for nonconvex nonlinear programming jamming and numerical test.* Mathematical Programming, **99** (2004), 35–48.

[6] C.M. Chin and R. Fletcher. *On the global convergence of an SLP-filter algorithm that takes EQP steps.* Mathematical Programming, **96** (2003), 161–177.

[7] R. Fletcher and S. Leyffer. *Nonlinear programming without a penalty function.* Mathematical Programming, **91** (2002), 239–269.

[8] R. Fletcher, S. Leyffer and Ph.L. Toint. *On the global convergence of a filter-SQP algorithm.* SIAM Journal on Optimization, **13** (2002), 44–59.

[9] R. Fletcher, N.I.M. Gould, S. Leyffer, Ph.L. Toint and A. Wächter. *Global convergence of a trust region SQP-filter algorithms for general nonlinear programming.* SIAM Journal on Optimization, **13** (2002), 635–659.

[10] C.C. Gonzaga, E. Karas and M. Vanti. *A globally convergent filter method for nonlinear programming.* SIAM Journal on Optimization, **14** (2004), 646–669.

[11] J. Gauvin. *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming.* Mathematical Programming, **12** (1977), 136–138.

[12] W. Hock and K. Schittkowski. *Test Examples for Nonlinear Programming Codes.* Lecture Notes in Econom. and Mathematical Systems 187, Springer-Verlag, Berlin (1981).

[13] X.W. Liu and Y.X. Yuan. *A robust algorithm for optimization with general equality and inequality constraints.* SIAM Journal on Scientific Computing, **22** (2000), 517–534.

[14] F. John. *Extremum problems with inequalities as subsidiary conditions.* Studies and Essays Presented to R. Courant on his 60th Birthday (Interscience, New York, NY 1948), 187–204.

[15] E.W. Karas, A. Ribeiro, C. Sagastizábal and M. Solodov. *A bundle-filter method for nonsmooth convex constrained optimization.* Mathematical Programming, **116** (2009), 297–320.

[16] M. Macconi, B. Morini and M. Porcelli. *Trust-region quadratic methods for nonlinear systems of mixed equalities and inequalities.* Applied Numerical Mathematics, to appear.

[17] P.Y. Nie. *A filter method for solving nonlinear complementarity problems.* Applied Mathematics and Computation, **167** (2005), 677–694.

[18] P.Y. Nie. *Sequential penalty quadratic programming filter methods for nonlinear programming.* Nonlinear Analysis: Real World Applications, **8** (2007), 118–129.

[19] M.J.D. Powell. A fast algorithm for nonlinearly constrained optimization calculations, in Numerical Analysis, Proceedings, Biennial conference, Dundee, Lecture Notes In Math. 630, G.A. Waston, ed., Springer-Verlag, Berlin, New York (1977), 144–157.

[20] R.M. Robinson. *Stability theory for systems of inequalities, part II: differential nonlinear systems.* SIAM Journal on Numerical Analysis, **4** (1976), 497–513.

[21] A.A. Ribeiro, E.W. Karas and C.C. Gonzaga. *Global convergence of filter methods for nonlinear programming.* SIAM Journal on Optimization, **19** (2008), 1231–1249.

[22] C. Sainvitu and Ph.L. Toint. *A filter-trust-region method for simple-bound constrained optimization.* Optimization Methods and Software, **22** (2007), 835–848.

[23] C. Shen, W. Xue and D. Pu. *Global convergence of a tri-dimensional filter SQP algorithm based on the line search method.* Applied Numerical Mathematics, **59** (2009), 235–250.

[24] S. Ulbrich. *On the superlinear local convergence of a filter-SQP method.* Mathematical Programming, **100** (2004), 217–245.

[25] M. Ulbrich, S. Ulbrich and L.N. Vicente. *A globally convergent primal-dual interior filter method for nonconvex nonlinear programming.* Mathematical Programming, **100** (2004), 379–410.

[26] A. Wächter and L.T. Biegler. *Line search filter methods for nonlinear programming: motivation and global convergence.* SIAM Journal on Optimization, **16** (2005), 1–31.

[27] A. Wächter and L.T. Biegler. *Line search filter methods for nonlinear programming: local convergence.* SIAM Journal on Optimization, **16** (2005), 32–48.

[28] J.L. Zhang and X.S. Zhang. *A modified SQP method with nonmonotone linsearch technique.* Journal of Global Optimization, **21** (2001), 201–218.

[29] G.L. Zhou. *A modified SQP method and its global convergence.* Journal of Global Optimization, **11** (1997), 193–205.