Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 349178, 20 pages
doi:10.1155/2012/349178

*Research Article*

# A Filter Algorithm with Inexact Line Search

**Meiling Liu,[1] Xueqian Li,[2] and Qinmin Wu[3]**

[1] *Department of Mathematics, Tongji University, Shanghai 200092, China*
[2] *Business School, University of Shanghai for Science and Technology, Shanghai 200093, China*
[3] *School of Management, Fudan University, Shanghai 200433, China*

Correspondence should be addressed to Meiling Liu, liuml2004@163.com

Received 25 October 2011; Accepted 18 December 2011

Academic Editor: Dongdong Ge



A filter algorithm with inexact line search is proposed for solving nonlinear programming problems. The filter is constructed by employing the norm of the gradient of the Lagrangian function to the infeasibility measure. Transition to superlinear local convergence is showed for the proposed filter algorithm without second-order correction. Under mild conditions, the global convergence can also be derived. Numerical experiments show the efficiency of the algorithm.

## 1. Introduction

Fletcher and Leyffer [1] proposed filter methods in 2002 offering an alternative to traditional merit functions in solving nonlinear programming problems (NLPs). The underlying concept is that a trial point is accepted provided that there is a sufficient decrease of the objective function or the constraint violation. Filter methods avoid the difficulty of determining a suitable value of the penalty parameter in the merit function. The promising numerical results in [1] led to a growing interest in filter methods in recent years. Two variants of trust-region filter sequential quadratic programming (SQP) method were proposed by Fletcher et al. [2, 3]. Chin and Fletcher [4] developed filter method to sequential linear programming strategy that takes equality-constrained quadratic programming steps. Ribeiro et al. [5] proposed a general filter algorithm that does not depend on the particular method used for the step of computation. Ulbrich [6] argued superlinear local convergence of a filter SQP method. Ulbrich et al. [7] and Wächter and Biegler [8] applied filter technique to interior method and achieved the global convergence to first-order critical point. Wächter and Biegler [9, 10] proposed a line-search filter method and applied it to different algorithm framework. Gould et al. [11] and Shen et al. [12] developed new multidimensional filter technique.

Su and Pu [13] extended the monotonicity of the filter technique. Nie [14] applied filter method to solve nonlinear complementarity problems. In this paper, the global convergence is analyzed widely. However, it has been noted by Fletcher and Leyffer [1] that the filter approach can suffer from the Maratos effect as that of a penalty function approach. By the Maratos effect, a full step can lead to an increase of both infeasibility measure and objective function in filter components even if arbitrarily close to a regular minimizer. This makes the full step unacceptable for the filter and can prohibit fast local convergence.

In this paper, we propose a filter algorithm with inexact line-search for nonlinear programming problems that ensures superlinear local convergence without second-order correction steps. We use the norm of the gradient of the Lagrangian function in the infeasibility measure in the filter components. Moreover, the new filter algorithm has the same global convergence properties as that of the previous works [2, 3, 9]. In addition, since the sufficient decrease conditions in an SQP framework can usually make the algorithm complex and time-consuming, the presented method is a line-search method without using SQP steps. An inexact line-search criterion is used as the sufficient reduction conditions. In the end, numerical experiences also show the efficiency of the new filter algorithm.

This paper is organized as follows. For the main part of the paper, the presented techniques will be applied to general NLP. In Section 2, we state the algorithm mechanism. The convergent properties are shown in Section 3. The global and superlinear convergence are proved. Furthermore, the Maratos effect is avoided. Finally, Section 4 shows the effectiveness of our method under some numerical experiences.

## 2. Inexact Line-Search Filter Approach

### *2.1. The Algorithm Mechanism*

We describe and analyze the line-search filter method for NLP with equality constraints. State it as

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & c_i(x) = 0, \quad i \in E, \end{aligned} \tag{2.1}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraints $c_i$ are assumed to be continuously differentiable, and $E = \{i \mid i = 1, 2, \ldots, m\}$.

The corresponding Lagrangian function is

$$L(x, \lambda) = f(x) + \lambda^T c(x), \tag{2.2}$$

where the vector $\lambda$ corresponds to the Lagrange multiplier. The Karush-Kuhn-Tucker (KKT) conditions for (2.1) are

$$\begin{aligned} \nabla f(x) + \lambda^T \nabla c(x) &= 0, \\ c(x) &= 0. \end{aligned} \tag{2.3}$$

For a given initial estimate $x_0$, the line-search algorithm generates a sequence of iterates $x_k$ by $x_{k+1} = x_k + \alpha_k d_k$ as the estimates of the solution for (2.1). Here, the search direction $d_k$ is computed from the linearization at $x_k$ of the KKT conditions (2.3):

$$\begin{pmatrix} W_k & \nabla c(x_k) \\ \nabla c(x_k) & 0 \end{pmatrix} \begin{pmatrix} d_k \\ \lambda^+ \end{pmatrix} = -\begin{pmatrix} \nabla f(x_k) \\ c(x_k) \end{pmatrix}, \tag{2.4}$$

where the symmetric matrix $W_k$ denotes the Hessian $\nabla^2_{xx} L(x_k, \lambda_k)$ of (2.2) or a positive definite approximation to it.

After a search direction $d_k$ has been computed, the step size $\alpha_k \in (0, 1]$ is determined by a backtracking line-search procedure, where a decreasing sequence of step size $\alpha_k$ is tried until some acceptable criteria are satisfied. Generally, the acceptable criteria are constructed by a condition that if the current trial point $x_k$ can provide sufficient reduction of a merit function. The filter method proposed by Fletcher and Leyffer [1] offers an alternative to merit functions. In this paper, the filter notion is defined as follows.

*Definition 2.1.* A pair $(V_k, \mathrm{f}_k)$ is said to dominate another pair $(V_l, f_l)$ if and only if both $V_k \le V_l$ and $f_k \le f_l$.

Here, we define $V(x) := \|c(x)\|_2 + \|\nabla L(x)\|_2$ as the $l_2$ norm of the infeasibility measure. That is, we modify the infeasibility measure in filter, with this modification, the superlinear convergence is possibly derived. Strictly, it is a stationarity measure. However, we still call it infeasibility measure according to its function. In the rest of paper, the norm is always computed by $l_2$ norm excepting special noting.

*Definition 2.2.* A filter is a list of pairs $(V_l, f_l)$ such that no pair dominates any other. A point $(V_k, f_k)$ is said to be acceptable for inclusion in the filter if it is not dominated by any point in the filter.

When a pair $(V_k, f_k)$ is said to be acceptable to the filter, we also say the iterate $x_k$ is acceptable to the filter. In filter notion, a trial point $x_{k+1}$ is accepted if it improves feasibility or improves the objective function. So, it is noted that filter criteria is less demanding than traditional penalty function. When improving optimality, the norm of the gradient of the Lagrangian function will tend to zero, so it offers a more precise analysis for the objective function.

However, this simple filter concept is not sufficient to guarantee global convergence. Fletcher et al. [3] replace this condition by requiring that the next iterate provides at least as much progress in one of the measure $V$ or $f$ that corresponds to a small fraction of the current infeasibility measure. Here, we use the similar technique to our filter. Formally, we say that a trial point can be accepted to the current iterate $x_k$ or the filter if

$$V(x_k) \le \beta V(x_l) \tag{2.5a}$$

or

$$f(x_k) \le f(x_l) - \gamma V(x_k), \tag{2.5b}$$

for some fixed constants $\beta, \gamma \in (0,1)$, and $(V(x_l), f(x_l))$ are points in current filter. In practical implementation, the constants $\beta$ close to 1 and $\gamma$ close to 0. However, the criteria (2.5a) and (2.5b) may make a trial point always provides sufficient reduction of the infeasibility measure alone, and not the objective function. To prevent this, we apply a technique proposed in Wächter and Biegler [10] to a different sufficient reduction criteria. The switching condition is

$$\nabla f_k^T d_k < 0, \qquad -\alpha_k \left[\nabla f_k^T d_k\right]^{e_1} > \delta [V(x_k)]^{e_2}, \tag{2.6}$$

where $\delta > 0$, $e_2 > 1$, $e_1 \geq 2e_2$. If the condition (2.6) holds, we replace the filter condition (2.5b) as an inexact line-search condition, that is, the Armijo type condition

$$f(x_{k+1}) \leq f(x_k) + \eta \alpha_k \nabla f_k^T d_k, \tag{2.7}$$

where $\eta \in (0, 1/2)$ is a constant. If (2.6) holds but not (2.7), the trial points are still determined by (2.5a) and (2.5b).

If a trial point $x_k$ can be accepted at a step size by (2.7), we refer to $x_k$ as an $f$ type iterate and the corresponding $\alpha_k$ as an $f$ step size.

## 2.2. The Algorithm Analysis

By the right part of switching condition (2.6), it ensures that the improvement to the objective function by the Armijo condition (2.7) is sufficiently large compared to the current infeasibility measure $V(x_k)$. Thus, if iterate points remote from the feasible region, the decrease of the objective function can be sufficient. By setting $e_2 > 1$, the progress predicted by the line model $-\alpha_k[\nabla f_k^T d_k]^{e_1}$ of $f$ can be a power of the infeasibility measure $V(x)$. The choice of $e_1 \geq 2e_2$ makes it possible that a full step can be accepted by (2.7) when it closes to a local solution.

In this paper, we denote the filter as a set $\mathcal{F}_k$ containing all iterates accepted by (2.5a) and (2.5b). During the optimization, if the $f$ type switching condition (2.6) holds and the Armijo condition (2.7) is satisfied, the trial point is determined by (2.7) not by (2.5a) and (2.5b), and the value of the objective function is strictly decreased. This can prevents cycling of the algorithm (see [10]).

If the linear system (2.4) is incompatible, no search direction $d_k$ can be found and the algorithm switches to a feasibility restoration phase. In it, we try to decrease the infeasibility measure to find a new iterate $x_{k+1}$ that satisfies (2.4) and is acceptable to the current filter. Similarly, in case $\nabla f_k^T d_k < 0$, the sufficient decrease criteria for the objective function in (2.5b) may not be satisfied. To inspect where no admissible step size $\alpha$ can be found and the feasibility restoration phase has to be invoked, we can consider reducing $\alpha$ and define

$$\alpha_k^{\min} = \begin{cases} \min\left\{1-\beta, -\dfrac{\gamma V(x_k)}{-\nabla f_k^T d_k}, \dfrac{\delta [V(x_k)]^{e_2}}{\left[-\nabla f_k^T d_k\right]^{e_1}}\right\}, & \text{if } \nabla f_k^T d_k < 0, \\ 1-\beta, & \text{otherwise.} \end{cases} \tag{2.8}$$

If the trial step size $\alpha < \alpha_k^{\min}$, the algorithm turns to the feasibility restoration phase.

By $\alpha_k^{\min}$, it is ensured that the algorithm does not switch to the feasibility restoration phase as long as (2.6) holds for a step size $\alpha < \alpha_k$ and that the backtracking line-search procedure is finite. Thus, for a trial point $x_k$, the algorithm eventually either delivers a new iterate $x_{k+1}$ or reverts to the feasibility restoration phase. Once finding a feasible direction, the algorithm still implements the normal algorithm.

Of course, the feasibility restoration phase may not always be possible to find a point satisfying the filter-accepted criteria and the compatible condition. It may converge to a nonzero local minimizer of the infeasibility measure and indicate that the algorithm fails. In this paper, we do not specify the particular procedure for the feasibility restoration phase. Any method for dealing with a nonlinear algebraic system can be used to implement a feasibility restoration phase.

### *2.3. The Algorithm*

We are now in the place to state the overall algorithm.

*Algorithm 2.3.*

*Step 1.* Given: starting point $x_0$, constants $V^{\max} = \max\{10^4, 1.2V(x_0)\}$, $\beta, \gamma \in (0,1)$, $\eta \in (0,1/2)$, $\delta > 0$, $e_1 \geq 2e_2$, $e_2 > 1$, $\tau \in (0,1)$.

*Step 2.* Initialize: $\mathcal{F}_0 = \{(V^{\max}, f(x_0))\}$, the iteration counter $k \leftarrow 0$.

*Step 3.* For $k = 0, 1, 2, \ldots$, stop if $x_k$ satisfies the KKT conditions (2.3).

*Step 4.* Compute the search direction $d_k$ from (2.4). If the system (2.4) is incompatible, go to the feasibility restoration phase in Step 7.

*Step 5.* Set $\alpha_0 = 1$, compute $\alpha_k^{\min}$.

(1) If $\alpha_k < \alpha_k^{\min}$, go to Step 7. Otherwise, compute the new trial point $x_{k+1} = x_k + \alpha_k d_k$.

(2) If the conditions (2.6) and (2.7) hold, accept the trial step and go to Step 6, otherwise set $x_k = x_{k+1}$, go to Step 5(3).

(3) In case where no $\alpha_k$ make (2.7) hold, if $x_{k+1}$ can be accepted to the filter, augment the filter by $\mathcal{F}_{k+1} = \mathcal{F}_k \cup \{(V(x_{k+1}), f(x_{k+1}))\}$, go to Step 6; Otherwise set $x_k = x_{k+1}$, go to Step 5(4).

(4) Compute $\alpha_{k+1} = \tau\alpha_k$, go back to Step 5(1).

*Step 6.* Increase the iteration counter $k \leftarrow k + 1$ and go back to Step 4.

*Step 7.* Feasibility restoration phase: by decreasing the infeasibility of $V$ to find a new iterate $x_{k+1}$ such that (2.4) is compatible. And if (2.7) holds at $x_{k+1}$, continue with the normal algorithm in Step 6; if (2.5a) and (2.5b) hold at $x_{k+1}$, augment the filter by $\mathcal{F}_{k+1} = \mathcal{F}_k \cup \{(V(x_{k+1}), f(x_{k+1}))\}$, and then continue with the normal algorithm in Step 6; if the feasibility restoration phase cannot find such a point, stop with insuccess.

*Remark 2.4.* In contrast to SQP method with trust-region technique, the actual step does not necessarily satisfy the linearization of the constraints.

*Remark 2.5.* Practical experience shows that the filter allows a large degree of nonmonotonicity and this can be advantageous to some problems.

*Remark 2.6.* To prevent the situation in which a sequence of points for which are $f$ type iterative point with $V_k \rightarrow \infty$ is accepted, we set an upper bound $V^{\max}$ on the infeasibility measure function $V$.

For further specific implementation details of Algorithm 2.3, see Section 4.

## 3. Convergence Analysis

### 3.1. Global Convergence

In this section, we give a global convergence analysis of Algorithm 2.3. We refer to the global convergence analysis of Wächter and Biegler [10] in some places. First, state the necessary assumptions.

*Assumption A1.* Let all iterates $x_k$ are in a nonempty closed and bounded set $S$ of $\mathbb{R}^n$.

*Assumption A2.* The functions $f$ and $c$ are twice continuously differentiable on an open set containing $S$.

*Assumption A3.* The matrix $W_k$ is positive definite on the null space of the Jacobian $\nabla c(x_k)$ and uniformly bounded for all $k$, and the Lagrange multiply $\lambda$ is bounded for all $k$.

Assumptions A1 and A2 are the standard assumptions. Assumption A3 plays an important role to obtain the convergence result and ensures that the algorithm is implementable.

For stating conveniently, we define a set $J = \{i \mid x_i \text{ is accepted to the filter}\}$. In addition, sometimes, it is need to revise $W_k$ to keep it positive definite by some updating methods such as damped BFGS formula [15] or revised Broyden's method [16].

From Assumptions A1–A3, we can get

$$\|d_k\| \leq M_d, \qquad m_W\|d_k\|^2 \leq d_k^T W_k d_k \leq M_W\|d_k\|^2, \qquad \|\lambda_k\| \leq M_\lambda, \tag{3.1}$$

where $M_d$, $M_\lambda$, $M_W$, and $m_W$ are constants.

**Lemma 3.1.** *Suppose Assumptions A1–A3 hold, if $\{x_{k_i}\}$ is a subsequence of iterates for which $\|d_{k_i}\| > \epsilon_1$ with a constant $\epsilon_1 > 0$ independent of $k_i$, then for constant $\epsilon_2 = (m_W/2)\epsilon_1$, if $V(x_{k_i}) \leq (m_W/2M_\lambda)\epsilon_1$, then*

$$\nabla f(x_{k_i})^T d_{k_i} \leq -\epsilon_2. \tag{3.2}$$

*Proof.* By the linear system (2.4) and (3.1),

$$\begin{aligned}\nabla f_{k_i}^T d_{k_i} &= -\lambda_{k_i}^T \nabla c(x_{k_i}) d_{k_i} - d_{k_i}^T W_{k_i} d_{k_i} \\ &= \lambda_{k_i}^T c(x_{k_i}) - d_{k_i}^T W_{k_i} d_{k_i} \\ &\leq M_\lambda V(x_{k_i}) - m_W\|d_{k_i}\|^2\end{aligned}$$

$$
\begin{aligned}
&\leq M_\lambda \frac{m_W}{2M_\lambda}\epsilon_1 - m_W\epsilon_1 \\
&= -\epsilon_2.
\end{aligned}
\tag{3.3}
$$

□

Lemma 3.1 shows that the search direction is a descent direction for the objective function when the trial points are sufficiently close to feasible region.

**Lemma 3.2.** *Suppose Assumptions A1–A3 hold, and that there exists an infinite subsequence* $\{x_{k_i}\}$ *of* $\{x_k\}$ *such that conditions* (2.6) *and* (2.7) *hold. Then*

$$\lim_{k\to\infty} V(x_{k_i}) = 0. \tag{3.4}$$

*Proof.* From Assumptions A1 and A2, we know that $\nabla f$ is bounded. Hence, it has with (3.1) that there exists a constant $M_m > 0$ such that

$$\left|\nabla f_k^T d_k\right| \leq M_m. \tag{3.5}$$

By (2.6) it has

$$\delta[V(x_{k_i})]^{e_2} < -\alpha_{k_i}\left[\nabla f_{k_i}^T d_{k_i}\right]^{e_1} \leq M_m^{e_1}\alpha_{k_i}. \tag{3.6}$$

As $1-1/e_1 > 0$, we have

$$\left(\frac{\delta}{M_m^{e_1}}\right)^{1-1/e_1}[V(x_{k_i})]^{e_2-e_2/e_1} < \alpha_{k_i}^{1-1/e_1}. \tag{3.7}$$

Then by (2.7) and (3.7),

$$
\begin{aligned}
f(x_{k_i+1}) - f(x_{k_i}) &\leq \eta\alpha_{k_i}\nabla f_{k_i}^T d_{k_i} \\
&< -\eta\delta^{1/e_1}\alpha_{k_i}^{1-1/e_1}[V(x_{k_i})]^{e_2/e_1} \\
&< -\eta\delta^{1/e_1}\left(\frac{\delta}{M_m^{e_1}}\right)^{1-1/e_1}[V(x_{k_i})]^{e_2}.
\end{aligned}
\tag{3.8}
$$

Hence, for $c_1 := \eta\delta^{1/e_1}(\delta/M_m^{e_1})^{1-1/e_1}$, an integer $K$ and all $j = 1, 2, \ldots$,

$$f(x_{K+j}) = f(x_K) + \sum_{k_i=K}^{K+j-1}\left(f(x_{k_i+1}) - f(x_{k_i})\right) < f(x_K) - c_1\sum_{k_i=K}^{K+j-1}[V(x_{k_i})]^{e_2}. \tag{3.9}$$

Since $f(x_{K+j})$ is bounded below as $j \to \infty$, the series on the right hand side in the last line of (3.8) is bounded, then implies the conclusion. □

**Lemma 3.3.** *Let* $\{x_{k_i}\} \subset \{x_k\}$ *be an infinite subsequence of iterates so that* $(V(x_{k_i}), f(x_{k_i}))$ *is entered into the filter. Then*

$$\lim_{i\to\infty} V(x_{k_i}) = 0. \tag{3.10}$$

*Proof.* Here, we refer to the proof of [2, Lemma 3.3]. If the conclusion is not true, there exists an infinite subsequence $\{k_j\} \subset \{k_i\} \subset J$ such that

$$V\left(x_{k_j}\right) \geq \epsilon, \tag{3.11}$$

for all $j$ and for some $\epsilon > 0$. This means that no other $(V, f)$ pair can be added to the filter at a later stage within the region

$$\left[V_{k_j} - (1-\beta)\epsilon, V_{k_j}\right] \times \left[f_{k_j} - \gamma\epsilon, f_{k_j}\right], \tag{3.12}$$

or with the intersection of this region with

$$S_0 = [0, V^{\max}] \times \left[f^{\min}, \infty\right] \tag{3.13}$$

for some constants $f^{\min} \leq f(x_k)$. Now, the area of each of these regions is $(1-\beta)\gamma\epsilon^2$. Hence, the set $S_0 \cup \{(V, f) \mid f \leq M_f\}$ is completely covered by at most a finite number of such regions, for any $M_f \geq f^{\min}$. Since the pairs $(V_{k_j}, f_{k_j})$ keep on being added to the filter, $f_{k_j}$ tends to infinity when $i$ tends to infinity. Without loss of generality, assume that $f_{k_{j+1}} \geq f_{k_j}$ for all $j$ is sufficiently large. But (2.5a) and (3.11) imply that

$$V_{k_{j+1}} \leq \beta V_{k_j}, \tag{3.14}$$

so $V_{k_j} \to 0$, which contradicts (3.11). Then, this latter assumption is not true and the conclusion follows. □

Next, we show that if $\{x_k\}$ is bounded, there exists at least one limit point of the iterative points is a first-order optimal point for (2.1).

**Lemma 3.4.** *Suppose Assumptions A1–A3 hold. Let* $\{x_{k_i}\}$ *be a subsequence with* $\nabla f_{k_i}^T d_{k_i} < -\epsilon_2$ *for a constant* $\epsilon_2 > 0$ *independent of* $k_i$*. Then, there exists a constant* $\overline{\alpha} > 0$ *so that for all* $k_i$ *and* $\alpha_{k_i} < \overline{\alpha}$*,*

$$f(x_{k_i} + \alpha_{k_i} d_{k_i}) - f(x_{k_i}) \leq \eta\alpha_{k_i} \nabla f_{k_i}^T d_{k_i}. \tag{3.15}$$

*Proof.* From Assumptions A1 and A2, $d^T\nabla^2 f(x)d \le c_f\|d\|^2$ for some constant $c_f > 0$. Thus, it follows from the Taylor Theorem and (3.1) that

$$\begin{aligned} f(x_{k_i} + \alpha_{k_i} d_{k_i}) - f(x_{k_i}) - \alpha_{k_i}\nabla f_{k_i}^T d_{k_i} \\ = \alpha_{k_i}^2 d_{k_i}^T \nabla^2 f(y_1) d_{k_i} \le c_f \alpha_{k_i}^2 \|d_{k_i}\|^2 \le \alpha_{k_i}(1-\eta)\epsilon_2 \\ \le -(1-\eta)\alpha_{k_i}\nabla f_{k_i}^T d_{k_i}, \end{aligned} \tag{3.16}$$

if $\alpha_{k_i} \le \overline{\alpha_{k_i}} := (1-\eta)\epsilon_2/c_f M_d^2$, where $y_1$ denotes some point on the line segment from $x_{k_i}$ to $x_{k_i} + \alpha_{k_i} d_{k_i}$. Then the conclusion follows. □

**Lemma 3.5.** *Suppose Assumptions A1–A3 hold, and the filter is augmented only a finite number of times, then*

$$\lim_{k\to\infty} \|d_k\| = 0. \tag{3.17}$$

*Proof.* Since the filter is augmented only a finite number of times, there exists an integer $K_1$ so that for all iterates $\{x_k\}_{k>K_1}$ the filter is not augmented. If the claim is not true, there must exist a subsequence $\{x_{k_i}\}$ and a constant $\epsilon > 0$ so that $\|d_{k_i}\| \ge \epsilon$ for all $i$. Then by Lemma 3.1, it has $\nabla f_{k_i}^T d_{k_i} \le -\epsilon_2$ for all $k_i \ge K_2$, $K_2$ is some integer and $K_2 > K_1$. And from Lemmas 3.2 and 3.4, it has $V(x_{k_i}) \le \epsilon$ and

$$f(x_{k_i+1}) - f(x_{k_i}) \le \eta\alpha_{k_i}\nabla f_{k_i}^T d_{k_i} \le -\alpha_{k_i}\eta\epsilon_2. \tag{3.18}$$

Since $f(x_{k_i})$ is bounded below and monotonically decreasing for all $k \ge K_2$, one can conclude that $\lim_{i\to\infty}\alpha_{k_i} = 0$. This means that for $k_i > K_2$ the step size $\alpha = 1$ has not been accepted. So, we can get a $\alpha_{k_i} < 1$ such that a trial point $x_{k_i+1} = x_{k_i} + \alpha_{k_i} d_{k_i}$ satisfies

$$\left(V(x_{k_i+1}), f(x_{k_i+1})\right) \notin \mathcal{F}_{k_i} \tag{3.19}$$

or

$$f(x_{k_i+1}) - f(x_{k_i}) > \eta\alpha_{k_i}\nabla f_{k_i}^T d_{k_i}. \tag{3.20}$$

Let $V^{\min} = \min\{V_k \mid (V_k, f_k) \in \mathcal{F}_k\}$. From Lagrange's Theorem, it has

$$\begin{aligned} V(x_{k_i+1}) &= V(x_{k_i}) + \alpha_{k_i}\nabla V(y_2)^T d_{k_i} \\ &\le V(x_{k_i}) + \alpha_{k_i}\left\|\nabla V(y_2)\right\|\|d_{k_i}\| \\ &\le V(x_{k_i}) + c_V\alpha_{k_i} M_d, \end{aligned} \tag{3.21}$$

for some constant $c_V$, where $y_2$ denotes some point on the line segment from $x_{k_i}$ to $x_{k_i} + \alpha d_{k_i}$. Since $\lim_{i\to\infty}\alpha_{k_i} = 0$ and $\lim_{i\to\infty} V(x_{k_i}) = 0$ by Lemmas 3.2 and 3.3, it has $V(x_{k_i+1}) < V^{\min}$

for $k_i$ sufficiently large, so (3.19) is not true. In case (3.20), since $\alpha_{k_i} \to 0$ for sufficiently large $k_i$, we have $\alpha_{k_i} \le \overline{\alpha}$ with $\overline{\alpha}$ from Lemma 3.4, that is, (3.20) can not be satisfied. Then the conclusion follows. □

**Lemma 3.6.** *Suppose Assumptions A1–A3 hold. Let $\{x_{k_i}\}$ be a subsequence of $\{x_k\}$ with $\nabla f_{k_i}^T d_{k_i} \le -\epsilon_2$ for a constant $\epsilon_2 > 0$ independent of $k_i$. Then, there exists trial points can be accepted to the filter.*

*Proof.* The mechanisms of Algorithm 2.3 ensure that the first iterate can be accepted to the filter. Next, we can assume that $(V(x_k), f(x_k))$ is acceptable to the $k$th filter and $(V(x_l), f(x_l)) \in \mathcal{F}_k, l < k$. If $\alpha_{k_i} \le c_2 := \epsilon_2 / M_d^2 c_f$, it has

$$\alpha_{k_i}^2 \le \frac{\alpha_{k_i}\epsilon_2}{M_d^2 c_f} \le \frac{-\alpha_{k_i}\nabla f_{k_i}^T d_{k_i}}{c_f \|d_{k_i}\|^2}, \tag{3.22}$$

that is, $\alpha_{k_i}\nabla f_{k_i}^T d_{k_i} + c_f \alpha_{k_i}^2 \|d_{k_i}\|^2 \le 0$, so by (3.16)

$$f(x_{k_i} + \alpha_{k_i} d_{k_i}) \le f(x_{k_i}). \tag{3.23}$$

Similarly, if $\alpha_{k_i} \le c_3 V(x_{k_i}) \le V(x_{k_i}) / \|d\|^2 c_V$, with $c_3 := 1/M_d^2 c_V$ and $c_V$ from Lemma 3.5, it has

$$\begin{gathered} -\alpha_{k_i} V(x_{k_i}) + c_V \alpha_{k_i}^2 \|d_{k_i}\|^2 \le 0, \\ V(x_{k_i} + \alpha_{k_i} d_{k_i}) \le V(x_{k_i}). \end{gathered} \tag{3.24}$$

Hence, we have

$$\begin{gathered} f(x_{k_i+1}) = f(x_{k_i} + \alpha_{k_i} d_{k_i}) \le f(x_{k_i}) \le f(x_l) - \gamma V(x_l), \\ V(x_{k_i+1}) = V(x_{k_i} + \alpha_{k_i} d_{k_i}) \le V(x_{k_i}) \le \beta V(x_l). \end{gathered} \tag{3.25}$$

The claim then follows from (3.25). □

The last Lemma 3.6 shows, for case $V(x_k) > 0$, Algorithm 2.3 either accepts a new iterate to the filter or switches to the feasibility restoration phase. For case $V(x_k) = 0$ and the algorithm does not stop at a KKT point, then $\nabla f_k^T d_k < 0$, $\alpha_k^{\min} = 0$, and the Armijo condition (2.7) is satisfied for sufficiently small step size $\alpha_k$, so an $f$ type iterate is accepted. Hence, the inner loop in Step 5 always terminates in a finite number of trial steps, and Algorithm 2.3 is well defined.

**Lemma 3.7.** *Suppose Assumptions A1–A3 hold. Let $\{x_{k_i}\}$ be a subsequence with $\|d_{k_i}\| \ge \epsilon$ for a constant $\epsilon > 0$ independent of $k_i$. Then, there exists an sufficient large integer $K$ such that for all $k_i > K$ the algorithm can generate some trial points either be accepted to the filter or be $f$ type steps.*

*Proof.* By Lemmas 3.1, 3.2, and 3.3, there exist constants $\epsilon_1, \epsilon_2 > 0$ so that

$$V(x_{k_i}) \le \epsilon_1, \qquad \nabla f_{k_i}^T d_{k_i} < -\epsilon_2 \tag{3.26}$$

for all $k_i > K$.

If $V(x_{k_i}) = 0$, the $f$ type switching condition (2.6) is true, there must exist iterates for which are $f$ type iterates. For the remaining iterates with $V(x_{k_i}) > 0$, if

$$V(x_{k_i}) < \min\left\{ \frac{\overline{\alpha}}{c_3}, \frac{c_2}{c_3}, \left[\frac{\tau c_3 \epsilon_2^{e_1}}{\delta}\right]^{1/e_2-1} \right\}, \tag{3.27}$$

with $\overline{\alpha}$ from Lemma 3.4 and $c_2, c_3$ from Lemma 3.6, it implies with $e_2 > 1$

$$\frac{\delta[V(x_{k_i})]^{e_2}}{\epsilon_2^{e_1}} < \tau c_3 V(x_{k_i}), \tag{3.28}$$

as well as

$$c_3 V(x_{k_i}) < \min\{\overline{\alpha}, c_2\}. \tag{3.29}$$

Now choose an arbitrary $k_i \geq K$ with $V(x_{k_i}) > 0$ and define

$$c_4 := c_3 V(x_{k_i}) = \min\{\overline{\alpha}, c_2, c_3 V(x_{k_i})\}. \tag{3.30}$$

Lemmas 3.4 and 3.6 then imply that a trial step size $\alpha_{k_i} \leq c_4$ satisfies both

$$\begin{gathered} f(x_{k_i} + \alpha_{k_i} d_{k_i}) \leq f(x_{k_i}) + \eta \alpha_{k_i} \nabla f_{k_i}^T d_{k_i}, \\ \left(V(x_{k_i} + \alpha_{k_i} d_{k_i}), f(x_{k_i} + \alpha_{k_i} d_{k_i})\right) \in \mathcal{F}_{k_i}. \end{gathered} \tag{3.31}$$

Since $\alpha > \tau\alpha_k > \tau\alpha_k^{\min}$ by the definition of $\alpha_k^{\min}$, the method does not switch to the feasibility restoration phase for those trial step sizes. Then the claim follows. □

Based on the above lemmas, we can give the main global convergence result.

**Theorem 3.8.** *Suppose Assumptions A1–A3 hold, then*

$$\lim_{k\to\infty} V(x_k) = 0, \tag{3.32}$$

$$\lim_{k\to\infty} \|d_k\| = 0, \tag{3.33}$$

*that is, there exits a limit point $\overline{x}$ of $\{x_k\}$ which is a first-order optimal point for* (2.1).

*Proof.* (3.32) follows from Lemmas 3.2 and 3.3.

If the filter is augmented a finite number of times, then the claim (3.33) holds from Lemma 3.5. For either case, there exists a subsequence $\{x_{k_i}\}$ so that $k_i \in J$ for all $i$. Suppose the conclusion (3.33) is not true, there must exist a subsequence $\{x_{k_j}\}$ of $\{x_{k_i}\}$ such that $\|d_{k_j}\| \geq \epsilon$ for some constant $\epsilon > 0$. Hence by Lemmas 3.1 and 3.3, it has $\nabla f_{k_j}^T d_{k_j} < -\epsilon_2$ and $\lim_{i\to\infty} V(x_{k_j}) = 0$ for all $k_j$. Then by Lemma 3.7, when $\alpha < \min\{\overline{\alpha}, c_2, c_3 V(x_{k_j})\}$, the algorithm can generate a $f$ type iterate, that is, the filter is not augmented, this contradicts the choice of $\{x_{k_j}\}$, so that (3.33) holds. □

### 3.2. Local Convergence

In this section, we show the local convergence of Algorithm 2.3.

*Assumption A4.* The iterates $x_k$ converge to a point $\overline{x}$ that satisfies

$$V(\overline{x}) = 0, \quad \nabla_x L\left(\overline{x}, \overline{\lambda}\right) = 0, \quad \nabla_{xx} L\left(\overline{x}, \overline{\lambda}\right) \text{ is positive definite on } \left\{d : \nabla c(\overline{x})^T d = 0\right\}, \\ \nabla c(\overline{x}) \text{ has full-row rank}. \tag{3.34}$$

*Assumption A 5.* There is a neighborhood $N(\overline{x})$ of $\overline{x}$ such that $W_k = \nabla_{xx} L(x_k, \lambda_k)$, for all $x_k \in N(\overline{x})$.

*Remark 3.9.* Under Assumption A4, the point $\overline{x}$ is a strict local minimum of (2.1).

*Remark 3.10.* Under Assumptions A4 and A5, it is well known that with the choice $x_{k+1} = x_k + d_k$, the sequence $\{x_k\}$ converges $q$-superlinearly to $\overline{x}$ and that the convergence is $q$-quadratic if $\nabla_{xx} f$ and $\nabla_{xx} c_i$ are lipschitz continuous in a neighborhood of $\overline{x}$. That is, for any given $\zeta \in [0,1]$, $x_j \in N(\overline{x})$, $j = k, k+1, \ldots$, and $x_{j+1} = x_j + d_j$, it has

$$\left\|d_{j+1}\right\| \leq \zeta \left\|d_j\right\|, \qquad \left\|x_{j+1} - \overline{x}\right\| \leq \zeta \left\|x_j - \overline{x}\right\|. \tag{3.35}$$

We use the proof techniques of local convergence in [6]. In proof, define $l_\rho(x,\lambda) = L(x,\lambda) + (\rho/2)\|c(x)\|_2^2$ and $\hat{l}_\rho(x,\lambda) = f(x) + (\rho/2)V(x)$ with $\rho$ is a parameter.

**Lemma 3.11.** *Suppose Assumptions A1–A3 hold. Let $\overline{x}$ satisfy the Assumption A4. Then, there exist constants $0 < r < t$, $\rho_0 > 0$ and a neighborhood $N_\sigma(\overline{x}) = \{x : \|x - \overline{x}\| < \sigma\} \subset N(\overline{x})$ such that*

$$\frac{r}{2}\|x - \overline{x}\|^2 + \frac{\rho - \rho_0}{2}\|c(x)\|^2 \leq l_\rho(x,\lambda) - l_\rho\left(\overline{x}, \overline{\lambda}\right) \leq t(1+\rho)\|x - \overline{x}\|^2, \tag{3.36}$$

*for all $\rho \geq \rho_0$ and all $x \in N_\sigma(\overline{x})$.*

*Moreover, possibly after increasing $t > 0$, $\rho_0 > 0$ and reducing $\sigma > 0$, we have that for all $\rho \geq \rho_0$ and all $x \in N_\sigma(\overline{x})$*

$$\frac{r}{2}\|x - \overline{x}\|^2 + \frac{\rho - \rho_0}{2}V(x) \leq \hat{l}_\rho(x,\lambda) - \hat{l}_\rho\left(\overline{x}, \overline{\lambda}\right) \leq t(1+\rho)\|x - \overline{x}\|^2. \tag{3.37}$$

*Proof.* Let $x \in N(\overline{x})$. Using Taylor's Theorem and $\nabla_x l_\rho(\overline{x},\overline{\lambda}) = 0$, $\nabla_\lambda l_\rho(\overline{x},\overline{\lambda}) = 0$, we have with some $(x',\lambda')$ on the line segment between $(\overline{x},\overline{\lambda})$ and $(x,\lambda)$

$$\begin{aligned} l_\rho(x,\lambda) - l_\rho\left(\overline{x},\overline{\lambda}\right) &= \frac{1}{2}(x-\overline{x})^T \nabla_{xx} l_\rho(x',\lambda')(x-\overline{x}) \\ &\quad + (x-\overline{x})^T \nabla_{x\lambda} l_\rho(x',\lambda')\left(\lambda - \overline{\lambda}\right). \end{aligned} \tag{3.38}$$

Obviously, it has

$$\nabla_{xx} l_\rho(x',\lambda') = \nabla_{xx} l_{\rho/2}\left(x',\lambda' + \frac{\rho}{2}c(x')\right) + \frac{\rho}{2}\nabla c(x')^T \nabla c(x'). \tag{3.39}$$

Under Assumption A4, there exists $\overline{\rho} > 0$ such that for all $\rho \geq \overline{\rho}$,

$$d^T \nabla_{xx} l_{\rho/2}\left(\overline{x},\overline{\lambda}\right) d \geq 4r\|d\|^2, \quad \forall d \in \mathbb{R}^n, \tag{3.40}$$

with a constant $r > 0$, see [15, Theorem 17.5].

Suppose $\lambda(x)$ is Lipschitz continuous and $L_y$ is the Lipschitz constant of $\lambda$, and $\overline{\rho} > 0$ is a constant. Let $\rho_0 := \max\{\overline{\rho}, 4L_y^2/r\}$ for all $x$ with $\|x - \overline{x}\| \leq \sigma$, it has by continuity

$$d^T \nabla_{xx} l_{\rho_0/2}\left(x,\lambda + \frac{\rho_0}{2}c(x)\right) d \geq 2r\|d\|^2, \quad \forall d \in \mathbb{R}^n. \tag{3.41}$$

Thus, we obtain for all $x \in N_\sigma(\overline{x})$ by (3.38), (3.39), and (3.41)

$$l_{\rho_0}(x,\lambda) - l_{\rho_0}\left(\overline{x},\overline{\lambda}\right) \geq r\|x-\overline{x}\|^2 + \frac{\rho_0}{4}\left\|\nabla c(x')(x-\overline{x})\right\|^2 + \left(\lambda - \overline{\lambda}\right)^T \nabla c(x')(x-\overline{x}). \tag{3.42}$$

It is obvious for all $s > 0$ that

$$\begin{aligned} 2\left(\lambda-\overline{\lambda}\right)^T \nabla c(x')(x-\overline{x}) &\geq -\frac{1}{s}\left\|\nabla c(x')(x-\overline{x})\right\|^2 - s\left\|\lambda - \overline{\lambda}\right\|^2 \\ &\geq -\frac{1}{s}\left\|\nabla c(x')(x-\overline{x})\right\|^2 - sL_y^2\left\|\lambda - \overline{\lambda}\right\|^2. \end{aligned} \tag{3.43}$$

If $s = r/L_y^2$, $1/s = L_y^2/r \leq \rho^0/4$, then it has

$$l_{\rho_0}(x,\lambda) - l_{\rho_0}\left(\overline{x},\overline{\lambda}\right) \geq \frac{r}{2}\|x-\overline{x}\|^2 + \frac{\rho_0}{8}\left\|\nabla c(x')(x-\overline{x})\right\|^2. \tag{3.44}$$

Since $c(\overline{x}) = 0$, $l_\rho(\overline{x},\overline{\lambda}) = l_{\rho_0}(\overline{x},\overline{\lambda}) = f(\overline{x})$, and by (3.44) it has

$$\begin{aligned} l_\rho(x,\lambda) - l_\rho\left(\overline{x},\overline{\lambda}\right) &= l_{\rho_0}(x,\lambda) - l_{\rho_0}\left(\overline{x},\overline{\lambda}\right) + \frac{\rho-\rho_0}{2}\|c(x)\|^2 \\ &\geq \frac{r}{2}\|x-\overline{x}\|^2 + \frac{\rho-\rho_0}{2}\|c(x)\|^2, \end{aligned} \tag{3.45}$$

that is the left inequality in (3.36). For the right inequality in (3.36), it is obvious from (3.38) that for all $x \in N(\overline{x})$,

$$l_\rho(x,\lambda) - l_\rho\left(\overline{x},\overline{\lambda}\right) \le t(1+\rho)\|x-\overline{x}\|^2, \tag{3.46}$$

with a constant $t > 0$. This proves the inequality (3.36).

Choose $K$ large enough such that for $j \ge k \ge K$, $c(x_j) \to 0$. We can assume $\|c(x)\| < 1$. Then, if $\rho_0$ is large enough, it has from (3.44) and $\hat{l}_\rho(\overline{x},\overline{\lambda}) = l_{\rho_0}(\overline{x},\overline{\lambda}) = f(\overline{x})$ that

$$\hat{l}_{\rho_0}(x,\lambda) - \hat{l}_{\rho_0}\left(\overline{x},\overline{\lambda}\right) \ge l_{\rho_0}(x,\lambda) - l_{\rho_0}\left(\overline{x},\overline{\lambda}\right) \ge \frac{r}{2}\|x-\overline{x}\|^2. \tag{3.47}$$

By an analogue of (3.45) holds for $\hat{l}_\rho$, this proves the left inequality in (3.37).

On the other hand, it has

$$\hat{l}_\rho(x,\lambda) - \hat{l}_\rho\left(\overline{x},\overline{\lambda}\right) = f(x) + \frac{\rho}{2}V(x) - f(\overline{x}). \tag{3.48}$$

Since $f(x)$ and $c(x)$ are twice continuously differentiable on closed set $S$, we have $f(x) - f(\overline{x}) = O(\|x-\overline{x}\|^2)$ and $V(x) = O(\|x-\overline{x}\|^2)$. This shows the right inequality in (3.37) possibly after increasing $t$. □

**Lemma 3.12.** *Let $\overline{x}$ satisfy Assumptions A4 and A5. Then for any $\zeta \in [0,1]$ and $M \ge 1$, there is an index $K$ such that for all $k \ge K$, with*

$$\|x_k - \overline{x}\| \le M \min_{(V_l,f_l)\in\mathcal{F}_k} \|x_l - \overline{x}\|, \tag{3.49}$$

*the points $x_{j+1}$, $j = k, k+1, \ldots$, with $\alpha_k = 1$ are acceptable to*

$$\mathcal{F}_j := \mathcal{F}_k \cup (V_k, f_k) \cup (V_{k+1}, f_{k+1}) \cup \cdots \cup (V_j, f_j). \tag{3.50}$$

*Proof.* Let $N(\overline{x})$ as in Assumption A5, $\zeta = 1/2$, $\rho_0$ and $N_\sigma(\overline{x}) \subset N(\overline{x})$ be given by Lemma 3.11. For all $k \ge K_1$, $K_1$ is a sufficient large integer, and choose $\rho \ge \rho_0$ so large that

$$\beta\left(\frac{\rho}{2} - \gamma\right) \ge \beta\frac{\rho}{4}, \qquad \frac{\rho-\rho_0}{2} \ge \left(1 - \frac{\beta}{4}\right)\frac{\rho}{2}. \tag{3.51}$$

Then, it has from (3.37) in Lemma 3.11

$$\frac{r}{2}\|x-\overline{x}\|^2 + \left(1 - \frac{\beta}{4}\right)\frac{\rho}{2}V(x) = \hat{l}_\rho(x,\lambda) - \hat{l}_\rho\left(\overline{x},\overline{\lambda}\right) \le t(1+\rho)\|x-\overline{x}\|^2. \tag{3.52}$$

Let

$$V' := \min_{(V_j,f_j)\in\mathcal{F}_{K_1}} V_j. \tag{3.53}$$

By $V' > 0$ and continuity there exists $0 < \sigma_1 < \sigma$ such that $V(x) \leq \beta V'$ for all $x \in N_{\sigma_1}(\overline{x})$, so the point $x$ is acceptable to $\mathcal{F}_{K_1}$. Since $x_k \rightarrow \overline{x}$, $x_k \in N_{\sigma_1}(\overline{x})$ for all $k \geq K_2 > K_1$, $K_2$ is an integer. By (3.35), we can choose $\sigma_1$ so small that for all $k \geq K_2$, the sequence $\{x_j\}_{j\geq k}$ with $\alpha_k = 1$ converges linearly with a contraction factor of at least

$$\|x_{j+1} - \overline{x}\| \leq \sqrt{\frac{1}{2}\left(1 - \frac{1-\beta/2}{1-\beta/4}\right)\frac{r}{2M^2 t(1+\rho)}}\|x_j - \overline{x}\| \leq \frac{1}{2}\|x_j - \overline{x}\|. \tag{3.54}$$

Suppose an arbitrary $k_2 \geq K_2$ such that (3.49) holds, and set $\sigma_2 := \|x_k - \overline{x}\|$. By (3.54), it has $x_{k+1} \in N_{\sigma_2}(\overline{x})$ and $x_{k+1}$ is acceptable to $\mathcal{F}_{K_1}$. Next, it is a need to show that $x_k$ is acceptable with respect to $(V_l, f_l) \in \mathcal{F}_k \cup (V_k, f_k)$ for $K_1 \leq l \leq k$. By (3.49), it has $x_l \in N_{\sigma}(\overline{x}) \setminus N_{\sigma_2/M}(\overline{x})$, so by (3.52)

$$\hat{l}_\rho(x_l, \lambda_l) - \hat{l}_\rho\left(\overline{x}, \overline{\lambda}\right) \geq \frac{r}{2M^2}\sigma_2^2 + \left(1 - \frac{\beta}{4}\right)\frac{\rho}{2}V_l. \tag{3.55}$$

And, by (3.52) and (3.54)

$$\hat{l}_\rho(x_{k+1}, \lambda_{k+1}) - \hat{l}_\rho\left(\overline{x}, \overline{\lambda}\right) \leq t(1+\rho)\|x_{k+1} - \overline{x}\|^2 \leq \frac{1}{2}\left(1 - \frac{1-\beta/2}{1-\beta/4}\right)\frac{r}{2M^2}\sigma_2^2. \tag{3.56}$$

Next, suppose $(x, \lambda)$ with $x \in N_{\sigma_2}(\overline{x})$ is not acceptable to $(V_l, f_l)$ then

$$V(x) > \beta V_l, \qquad f(x) + \gamma V(x) > f_l. \tag{3.57}$$

Thus, it has with (3.51) and (3.52)

$$\begin{aligned}
\hat{l}_\rho(x, \lambda) &= f(x) + \frac{\rho}{2}V(x) > f_l + \left(\frac{\rho}{2} - \gamma\right)V(x) \\
&> f_l + \beta\left(\frac{\rho}{2} - \gamma\right)V_l \geq f_l + \beta\left(\frac{\rho}{4}\right)V_l \\
&= \hat{l}_\rho(x_l, \lambda_l) - \left(1 - \frac{\beta}{2}\right)\frac{\rho}{2}V_l \\
&\geq \hat{l}_\rho(x_l, \lambda_l) - \frac{1-\beta/2}{1-\beta/4}\left(\hat{l}_\rho(x_l, \lambda_l) - \hat{l}_\rho\left(\overline{x}, \overline{\lambda}\right)\right).
\end{aligned} \tag{3.58}$$

This shows with (3.55) that

$$\begin{aligned}
\hat{l}_\rho(x, \lambda) - \hat{l}_\rho\left(\overline{x}, \overline{\lambda}\right) &> \left(1 - \frac{1-\beta/2}{1-\beta/4}\right)\left(\hat{l}_\rho(x_l, \lambda_l) - \hat{l}_\rho\left(\overline{x}, \overline{\lambda}\right)\right) \\
&\geq \left(1 - \frac{1-\beta/2}{1-\beta/4}\right)\frac{r}{2M^2}\sigma_2^2.
\end{aligned} \tag{3.59}$$

This produces a contradiction to (3.56), so $x_{k+1}$ is acceptable to $\mathcal{F}_k \cup (V_k, f_k)$. Then, the acceptability of $(V_{j+1}, f_{j+1})$ for $\mathcal{F}_l$ follows by induction. □

Next, we show that the sequence $\{x_j\}_{j\geq k}$ with $\alpha_k = 1$ can make the sufficient decreasing condition (2.7) hold.

**Lemma 3.13.** *Suppose Assumptions A1–A3 hold. Let $\overline{x}$ satisfy Assumptions A4 and A5 and let $K$ be as in Lemma 3.11. Then for all $k > K$ the sequence $\{x_j\}_{j\geq K}$ with $\alpha_j = 1$ satisfies*

$$f(x_{j+1}) \leq f(x_j) + \eta \alpha_j \nabla f_j^T d_j. \tag{3.60}$$

*Proof.* Suppose $\alpha_j \nabla f_j^T d_j < 0$ and $\alpha_j [\nabla f_j^T d_j]^{e_1} < -\delta [V(x_j)]^{e_2}$ hold. By $\alpha_j = 1$, thus

$$\nabla f_j^T d_j < -(\delta)^{1/e_1} [V(x_j)]^{e_2/e_1}. \tag{3.61}$$

On the other hand, with $\alpha_j = 1$ the assertion $f(x_{j+1}) \leq f(x_j) + \eta \nabla f_j^T d_j$ yields

$$\eta \nabla f_j^T d_j \geq \nabla f_j^T d_j + \frac{1}{2} d_j^T \nabla_{xx}^2 f(x') d_j. \tag{3.62}$$

Thus,

$$\nabla f_j^T d_j \leq -\frac{1}{2(1-\eta)} d_j^T \nabla_{xx}^2 f(x') d_j, \tag{3.63}$$

where $x'$ on the line segment between $x_j$ and $x_{j+1}$.

Obviously, we can prove the conclusion if with $K > 0$ large enough and for all $j \geq k \geq K$ the following holds

$$-(\delta)^{1/e_1} [V(x_j)]^{e_2/e_1} \leq -\frac{1}{2(1-\eta)} d_j^T \nabla_{xx}^2 f(x') d_j. \tag{3.64}$$

Since $c_j + \nabla c_j^T d_j = 0$, $d_j = -\nabla c_j^T (\nabla c_j \nabla c_j^T)^{-1} c_j$. By Assumption A4, $\nabla c_j$ has full-row rank, there exists $c_d > 0$ such that

$$\|d_j\| \leq c_d V(x_j), \tag{3.65}$$

thus,

$$\|d_j\|^2 \leq c_d^2 [V(x_j)]^2. \tag{3.66}$$

Choose $K$ large enough such that $V(x_j) \leq 1$ for all $j \geq k \geq K$. By $e_1 > 2e_2$ and (3.66) it has

$$\|d_j\|^2 \leq c_d^2 c_5 [V(x_j)]^{e_2/e_1}, \tag{3.67}$$

**Table 1:** Description on headers.

| Header | Description |
|---|---|
| Problem | The name of the CUTE problem being solved |
| $n$ | Number of variables of the problem |
| $m$ | The total number of constraints |
| $m_{\mathrm{nl}}$ | Number of nonlinear constraints |
| NIT1 | Number of iterations of Algorithm 2.3 |
| NIT2 | Number of iterations of algorithm Tri-filter |
| NIT3 | Number of iterations of algorithm SNOPT |
| NF1 | Number of $f$ evaluations of Algorithm 2.3 |
| NF2 | Number of $f$ evaluations of algorithm Tri-filter |
| NF3 | Number of $f$ evaluations of algorithm SNOPT |

for a constant $c_5 := 2\delta^{1/e_1}(1-\eta)/c_f c_d^2$, we can choose suitable parameters such that the last inequality holds. Thus,

$$\delta^{1/e_1}\left[V\left(x_j\right)\right]^{e_2/e_1} \geq \frac{1}{2(1-\eta)} c_f \left\|d_j\right\|^2 \geq \frac{1}{2(1-\eta)} d_j^T \nabla_{xx}^2 f\left(x'\right) d_j. \tag{3.68}$$

This completes the proof. □

**Theorem 3.14.** *Suppose Assumptions A1–A5 hold. Then, there exists $K > 0$ such that Algorithm 2.3 takes steps with $\alpha_k = 1$ for all $k \geq K$, that is,*

$$x_{k+1} = x_k + d_k. \tag{3.69}$$

*In particular, $x_k$ converges q-superlinearly to $\overline{x}$. If $\nabla_{xx} f$ and $\nabla_{xx} c_i$ are Lipschitz continuous in a neighborhood of $\overline{x}$ then $x_k$ converges q-quadratically.*

*Proof.* Since Assumptions A4 and A5 hold, $x_k \to \overline{x}$ with $\overline{x}$ satisfying A4. By Lemmas 3.12 and 3.13, the iterate $(x_{k+1}, \lambda_{k+1}) = (x_k + d_k, \lambda(x_k + d_k))$ is acceptable to the filter $\mathcal{F}_k \cup (V_k, f_k)$ and satisfies the sufficient decreasing condition (2.7). Thus, the trial iterate $(x_k + d_k, \lambda(x_k + d_k))$ is accepted by the algorithm and it has

$$x_{k+1} = x_k + d_k, \qquad \lambda_{k+1} = \lambda(x_k + d_k). \tag{3.70}$$

That is in both cases the algorithm takes the steps with $\alpha_k = 1$. And according to Remark 3.10, $\{x_k\}$ converges $q$-superlinearly to $\overline{x}$. □

## 4. Numerical Experience

In this section, we give some numerical results of Algorithm 2.3. We take some CUTE problems [17], which are available freely on NEOS, to test our algorithm. The test codes are edited in MATLAB. The details about the implementation are described as follows.

**Table 2:** Numerical results of small-scale problems.

| Problem | $n$ | $m$ | $m_{\mathrm{nl}}$ | NIT1/NIT2 | NF1/NF2 |
|---|---|---|---|---|---|
| hs017 | 2 | 5 | 3 | 11/14 | 14/20 |
| hs019 | 2 | 6 | 4 | 2/5 | 5/6 |
| hs024 | 2 | 4 | 2 | 1/4 | 2/5 |
| hs037 | 3 | 7 | 6 | 5/7 | 6/10 |
| hs042 | 3 | 5 | 4 | 6/7 | 8/10 |
| hs043 | 4 | 3 | 0 | 11/14 | 12/16 |
| hs046 | 5 | 4 | 2 | 19/30 | 41/102 |
| hs047 | 5 | 6 | 3 | 9/19 | 28/70 |
| hs049 | 5 | 4 | 2 | 17/23 | 27/26 |
| hs056 | 7 | 15 | 11 | 8/18 | 16/41 |
| hs059 | 2 | 7 | 4 | 13/16 | 25/20 |
| hs063 | 3 | 7 | 5 | 13/15 | 30/44 |
| hs071 | 4 | 11 | 9 | 6/28 | 12/103 |
| hs076 | 4 | 7 | 4 | 5/6 | 6/7 |
| hs077 | 5 | 4 | 2 | 15/26 | 25/36 |
| hs078 | 5 | 6 | 3 | 8/6 | 18/29 |
| hs079 | 5 | 6 | 3 | 6/9 | 10/11 |
| hs092 | 6 | 1 | 1 | 25/18 | 35/52 |
| hs098 | 6 | 16 | 12 | 6/6 | 6/7 |
| hs099 | 19 | 42 | 28 | 15/19 | 25/78 |
| hs104 | 8 | 22 | 16 | 16/23 | 33/60 |
| hs106 | 8 | 22 | 21 | 32/36 | 12/37 |
| hs108 | 9 | 14 | 10 | 15/12 | 13/13 |
| hs111 | 10 | 26 | 23 | 42/48 | 50/52 |
| hs112 | 10 | 16 | 16 | 25/52 | 25/30 |
| hs113 | 10 | 8 | 1 | 9/13 | 15/19 |
| hs114 | 10 | 34 | 23 | 25/28 | 23/29 |
| hs116 | 13 | 41 | 35 | 37/44 | 30/54 |
| hs117 | 15 | 20 | 15 | 15/16 | 15/23 |
| hs118 | 15 | 59 | 59 | 12/18 | 22/19 |
| hs119 | 16 | 48 | 40 | 13/18 | 16/19 |

**Table 3:** Numerical results of mid-scale problems.

| Problem | $n$ | $m$ | $m_{\mathrm{nl}}$ | NIT1/NIT3 | NF1/NF3 |
|---|---|---|---|---|---|
| Ngone | 97 | 1371 | 1369 | 16/72 | 21/20 |
| Grouping | 100 | 450 | 350 | 1/0 | 4/2 |
| Eigen a2 | 110 | 110 | 55 | 8/11 | 5/6 |
| Eigen a | 110 | 110 | 110 | 52/226 | 158/208 |

(a) The parameters are set to $\beta = 0.99$, $\gamma = 0.0001$, $\eta = 0.4$, $\delta = 1$, $\tau = 0.6$, $e_1 = 2.3$, $e_2 = 1.1$, the termination tolerance $\epsilon = 1E - 6$.

(b) The optimal residual is defined as

$$\text{res} = \max\{\|\nabla f(x_k) + \nabla c(x_k)\lambda\|, V(x_k)\}. \tag{4.1}$$

That is, the algorithm terminates when res $< \epsilon$.

(c) $W_k$ is updated by damped BFGS formula [15].

The detailed results of the numerical test on small-scale problems are summarized in Table 2. For comparison purposes, we also give the numerical results of tridimensional line-search filter solver (Tri-filter) in Shen et al. [12] in Table 2. The row headers in Tables 2 and 3 are presented in Table 1.

The results in Table 2 indicate that Algorithm 2.3 has a good effect.

In addition, we also test some mid-scale problems. And we compare the numerical results which are summarized in Table 3 in Algorithm 2.3 and SNOPT solver in Gill et al. [18] since no mid-scale problems are given in trifilter solver.

From Table 3, we find the efficiency of Algorithm 2.3 is also improved significantly. From both Tables 2 and 3, in general, the behavior of the proposed algorithm is rather stable. Finally, we may conclude that, as far as our limited computational experience is concerned, the proposed algorithm is well comparable to trifilter solver and SNOPT solver.

## Acknowledgment

This research is partially supported by the National Natural Science Foundation of China (Grant nos. 10771162 and 91024021).

## References

[1] R. Fletcher and S. Leyffer, “Nonlinear programming without a penalty function,” *Mathematical Programming*, vol. 91, no. 2, pp. 239–269, 2002.

[2] R. Fletcher, N. I. M. Gould, S. Leyffer, P. L. Toint, and A. Wächter, “Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming,” *SIAM Journal on Optimization*, vol. 13, no. 3, pp. 635–659, 2002.

[3] R. Fletcher, S. Leyffer, and P. L. Toint, “On the global convergence of a filter-SQP algorithm,” *SIAM Journal on Optimization*, vol. 13, no. 1, pp. 44–59, 2002.

[4] C. M. Chin and R. Fletcher, “On the global convergence of an SLP-filter algorithm that takes EQP steps,” *Mathematical Programming*, vol. 96, no. 1, Ser. A, pp. 161–177, 2003.

[5] A. A. Ribeiro, E. W. Karas, and C. C. Gonzaga, “Global convergence of filter methods for nonlinear programming,” *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1231–1249, 2008.

[6] S. Ulbrich, “On the superlinear local convergence of a filter-SQP method,” *Mathematical Programming*, vol. 100, no. 1, pp. 217–245, 2004.

[7] M. Ulbrich, S. Ulbrich, and L. N. Vicente, “A globally convergent primal-dual interior-point filter method for nonlinear programming,” *Mathematical Programming*, vol. 100, no. 2, pp. 379–410, 2003.

[8] A. Wächter and L. T. Biegler, “On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming,” *Mathematical Programming*, vol. 106, no. 1, pp. 25–57, 2006.

[9] A. Wächter and L. T. Biegler, “Line search filter methods for nonlinear programming: motivation and global convergence,” *SIAM Journal on Optimization*, vol. 16, no. 1, pp. 1–31, 2005.

[10] A. Wächter and L. T. Biegler, “Line search filter methods for nonlinear programming: local convergence,” *SIAM Journal on Optimization*, vol. 16, no. 1, pp. 32–48, 2005.

[11] N. I. M. Gould, C. Sainvitu, and P. L. Toint, “A filter-trust-region method for unconstrained optimization,” *SIAM Journal on Optimization*, vol. 16, no. 2, pp. 341–357, 2006.

[12] C. G. Shen, W. J. Xue, and D. G. Pu, “Global convergence of a tri-dimensional filter SQP algorithm based on the line search method,” *Applied Numerical Mathematics*, vol. 59, no. 2, pp. 235–250, 2009.

[13] K. Su and D. G. Pu, “A nonmonotone filter trust region method for nonlinear constrained optimization,” *Journal of Computational and Applied Mathematics*, vol. 223, no. 1, pp. 230–239, 2009.

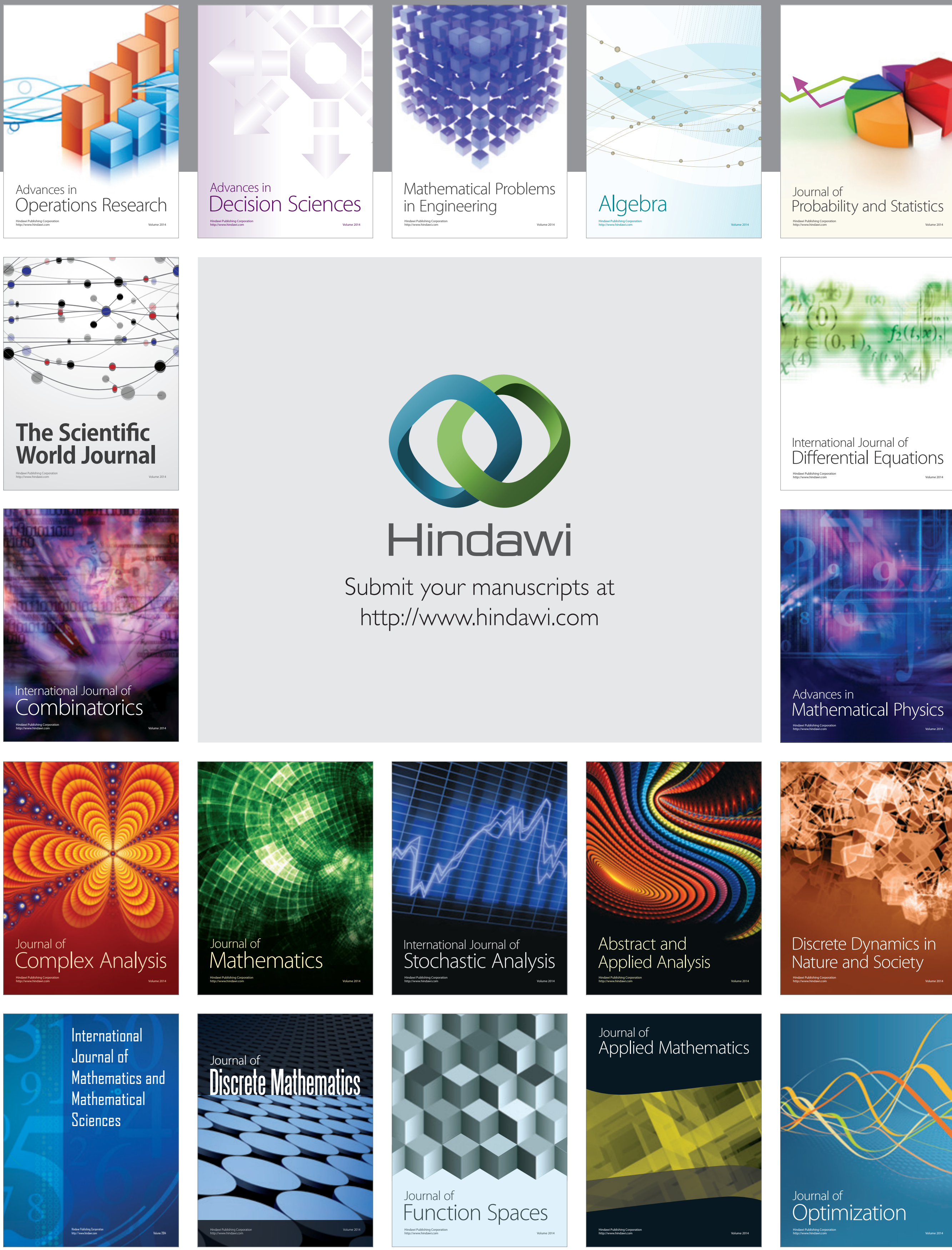
[14] P. Y. Nie, “A filter method for solving nonlinear complementarity problems,” *Applied Mathematics and Computation*, vol. 167, no. 1, pp. 677–694, 2005.

[15] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Series in Operations Research, Springer, New York, NY, USA, 1999.

[16] D. G. Pu, S. H. Gui, and W. W. Tian, “A class of revised Broyden algorithms without exact line search,” *Journal of Computational Mathematics*, vol. 22, no. 1, pp. 11–20, 2004.

[17] I. Bongartz, A. R. Conn, N. Gould, and P. L. Toint, “CUTE: constrained and unconstrained testing environment,” *ACM Transactions on Mathematical Software*, vol. 21, no. 1, pp. 123–160, 1995.

[18] P. E. Gill, W. Murray, and M. A. Saunders, “SNOPT: an SQP algorithm for large-scale constrained optimization,” *SIAM Review*, vol. 47, no. 1, pp. 99–131, 2005.


Advances in
Operations Research
Advances in
Decision Sciences
Mathematical Problems
in Engineering
Algebra
Journal of
Probability and Statistics
The Scientific
World Journal
International Journal of
Differential Equations
Hindawi
Submit your manuscripts at
http://www.hindawi.com
International Journal of
Combinatorics
Advances in
Mathematical Physics
Journal of
Complex Analysis
Journal of
Mathematics
International Journal of
Stochastic Analysis
Abstract and
Applied Analysis
Discrete Dynamics in
Nature and Society
International
Journal of
Mathematics and
Mathematical
Sciences
Journal of
Discrete Mathematics
Journal of
Function Spaces
Journal of
Applied Mathematics
Journal of
Optimization