ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439

Nonmonotone Filter Method for Nonlinear Optimization

Roger Fletcher, Sven Leyffer, and Chungen Shen

Mathematics and Computer Science Division

Preprint ANL/MCS-P1679-0909

October 14, 2009

This work was supported by the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract DE-AC02-06CH11357.

Contents

Nonmonotone Filter Method for Nonlinear Optimization[*]

Roger Fletcher,[†] Sven Leyffer,[‡] and Chungen Shen[§]

October 14, 2009

Abstract

We propose a new nonmonotone filter method to promote global and fast local convergence for sequential quadratic programming algorithms. Our method uses two filters: a global g-filter for global convergence, and a local nonmonotone l-filter that allows us to establish fast local convergence. We show how to switch between the two filters efficiently, and we prove global and superlinear local convergence. A special feature of the proposed method is that it does not require second-order correction steps. We present preliminary numerical results comparing our implementation with a classical filter SQP method.

Keywords: Nonlinear optimization, nonmonotone filter, global convergence, local convergence.

AMS-MSC2000: 65K05, 90C30.

1 Introduction and Background

We consider the constrained optimization problem

$$\left\{\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & c_i(x) = 0,\ i \in \mathcal{E} \\ & c_i(x) \le 0,\ i \in \mathcal{I}, \end{array}\right. \tag{1.1}$$

where $c(x) = (c_1(x), c_2(x), \cdots, c_m(x))^T$, $\mathcal{E} = \{1, 2, \cdots, m_1\}$, and $\mathcal{I} = \{m_1+1, m_1+2, \cdots, m\}$. The objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraint functions $c_i : \mathbb{R}^n \to \mathbb{R}$ are twice continuously differentiable functions.

The sequential quadratic programming (SQP) method is an iterative method for solving the problem (1.1). Fletcher and Leyffer (2002) proposed the filter technique for SQP methods and used it in the context of a trust-region SQP method for solving nonlinear optimization problems. Their

[*]Preprint ANL/MCS-P1679-0909.

[†]Mathematics Department, University of Dundee, UK, `fletcher@math.dundee.ac.uk`.

[‡]Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, `leyffer@mcs.anl.gov`.

[§]Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, `shenchungen@gmail.com`.

computational results were encouraging. Subsequently, global convergences of the trust-region filter SQP methods were established by Fletcher et al. (2002a) and Fletcher et al. (2002b). Gonzaga et al. (2003) proposed a globally convergent filter method in which each iteration is composed of a feasibility phase and an optimality phase, and Ribeiro et al. (2008) presented an alternative version of that method. Wächter and Biegler (2005b) proposed a line-search filter SQP method and showed its global convergence. Audet and Dennis Jr. (2004), and Karas et al. (2009) applied the filter technique to derivative-free optimization and nonsmooth optimization, respectively.

Unfortunately, filter SQP methods may also encounter the Maratos effect (Conn et al., 2000). To overcome this disadvantage, Ulbrich (2004) presented a trust-region filter method, using the Lagrangian function instead of the objective function as one measure in the entry of the filter. Ulbrich showed local convergence without the use of second-order correction (SOC) steps. Wächter and Biegler (2005a) proposed a line-search filter method and proved fast local convergence with the help of SOC steps. Gould and Toint (2003) introduced a nonmonotone trust-region filter algorithm, which provides a global convergence framework for filter methods. However, they did not show fast convergence proofs. Our nonmonotone filter method differs substantially from the method proposed in (Gould and Toint, 2003), and is easier to implement in our view.

In this paper, we present a new filter method that combines global and fast local convergence. Our method improves on previous results for second-order filter methods. Unlike Ulbrich (2004), we do not use the Lagrangian function in our filter but continue to use the objective in both filters. Thus, we avoid the potential pitfall of converging to a saddle point. In addition, our method does not need to compute second-order correction steps, unlike that of Wächter and Biegler (2005a). This is an advantage because the computation of second-order correction steps can be cumbersome, and complicates the implementation. In Section 5 we show that the omission of SOC steps does not degrade performance.

To obtain global and fast local convergence, our algorithm defines two filters: one is a standard filter (g-filter) for global convergence; the other one is a nonmonotone filter (l-filter) for local convergence. The g-filter forces iterates toward an optimal point, and the l-filter is a local filter that accepts full SQP steps promoting fast local convergence. Without the help of the SOC steps, we prove that, for all sufficiently large iteration numbers, iterates with full SQP steps are accepted by the l-filter and therefore fast local convergence is achieved.

This paper is organized as follows. In Section 2, we provide some definitions of our filters and describe how these filters work in the main algorithm. In Section 3, we prove that the algorithm is well defined. Under the Mangasarian-Fromowitz constraint qualification (MFCQ) condition, we show that at least one of accumulation points is a KKT point. In Section 4, we prove that iterates generated by our filter algorithm converge to a minimizer superlinearly or quadratically under mild conditions. In Section 5, we provide preliminary numerical results showing that the absence of SOC steps does not adversely affect the algorithm.

Notation. We make extensive use of the symbols $o(\cdot)$, $\mathcal{O}(\cdot)$, and $\Theta(\cdot)$. Let η_k and ν_k be two vanishing sequences, where $\eta_k,\ \nu_k \in \mathbb{R}$. If the sequence of ratios $\{\eta_k/\nu_k\}$ approaches zero as $k \to \infty$, then we write $\eta_k = o(\nu_k)$. If there exists a constant $C > 0$ such that $|\eta_k| \le C|\nu_k|$ for all

k sufficiently large, then we write $\eta_k = \mathcal{O}(\nu_k)$. If both $\eta_k = \mathcal{O}(\nu_k)$ and $\nu_k = \mathcal{O}(\eta_k)$, then we write $\eta_k = \Theta(\nu_k)$.

2 Definitions and Algorithm Statement

Our algorithm is an SQP method. It generates iterates by solving a sequence of quadratic programs. At the kth iterate x_k, we compute a trial step by solving the quadratic program

$$\mathrm{QP}(x_k,\rho) \quad \begin{cases} \underset{d}{\text{minimize}} & q(d) = \nabla f(x_k)^T d + \frac{1}{2} d^T B_k d \\ \text{subject to} & \nabla c_i(x_k)^T d + c_i(x_k) = 0,\ i \in \mathcal{E} \\ & \nabla c_i(x_k)^T d + c_i(x_k) \le 0,\ i \in \mathcal{I} \\ & \|d\|_\infty \le \rho, \end{cases}$$

where $\rho > 0$ is the trust-region radius and B_k approximates the Hessian of the Lagrangian

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

at x_k. The solution of QP(x_k,ρ) is denoted by d if QP(x_k,ρ) is feasible. If it is infeasible, our algorithm enters a feasibility restoration phase to find a new point so that the QP subproblem is feasible at this point. After d is computed, we take $\hat{x} := x_k + d$ as the next trial iterate. We define

$$\Delta q(d) = q(0) - q(d) = -\nabla f(x_k)^T d - \frac{1}{2} d^T B_k d \tag{2.3}$$

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

$$\Delta \tilde{f}(d) = \max_{i \in \{0,\cdots,M\}} \tilde{f}(x_{k-i}) - f(\hat{x}) \tag{2.4}$$

as the nonmonotone actual reduction of $f(x)$, where $M \ge 0$ is the level of nonmonotonicity and $M = 0$ corresponds to a monotone algorithm. We also define

$$\Delta f(d) = f(x_k) - f(\hat{x}) \tag{2.5}$$

as the actual reduction of $f(x)$. We define the constraint violation as

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

For convenience, we also define

$$\tilde{h}_k = \max_{i \in \{0,\cdots,M\}} h(x_{k-i}).$$

We use $\bar{d}_k$ to denote the solution of QP(x_k,∞) if it is feasible. We emphasize that we need $\bar{d}_k$ only conceptually and that we do not solve QP(x_k,∞). If the solution of QP(x_k,ρ) satisfies $\rho > \|d\|_\infty$, then we take d as $\bar{d}_k$.

In this paper, we use the filter technique to check the acceptance of a trial point. The following definitions are taken from Chin and Fletcher (2003).

Definition 2.1. *A point $\hat{x}$ (or $(h(\hat{x}), f(\hat{x}))$) is said to be* acceptable to x_l (or $(h(x_l), f(x_l))$) *if one of the following conditions is satisfied:*

$$h(\hat{x}) \leq \beta h(x_l) \tag{2.6a}$$

$$\text{or } f(\hat{x}) - f(x_l) \leq -\gamma h(\hat{x}), \tag{2.6b}$$

where β, $\gamma \in (0, 1)$ are constants. A point $\hat{x}$ (or $(h(\hat{x}), f(\hat{x}))$) is said to be dominated by x_l (or $(h(x_l), f(x_l))$) *if it is not acceptable to x_l (or $(h(x_l), f(x_l))$).*

We use $\mathcal{F}_k^g$ to denote the set of iteration index j ($j \leq k$) such that $(h(x_j), f(x_j))$ is an entry in the current g-filter. Similarly, we use $\mathcal{F}_k^l$ to denote that set of the current l-filter. In fact, the g-filter (see Figure 1, left) is a standard filter (Chin and Fletcher, 2003), which is a list of pairs $\{(h(x_k), f(x_k))\}$ such that any pair in the filter is acceptable to all previous pairs in the filter.

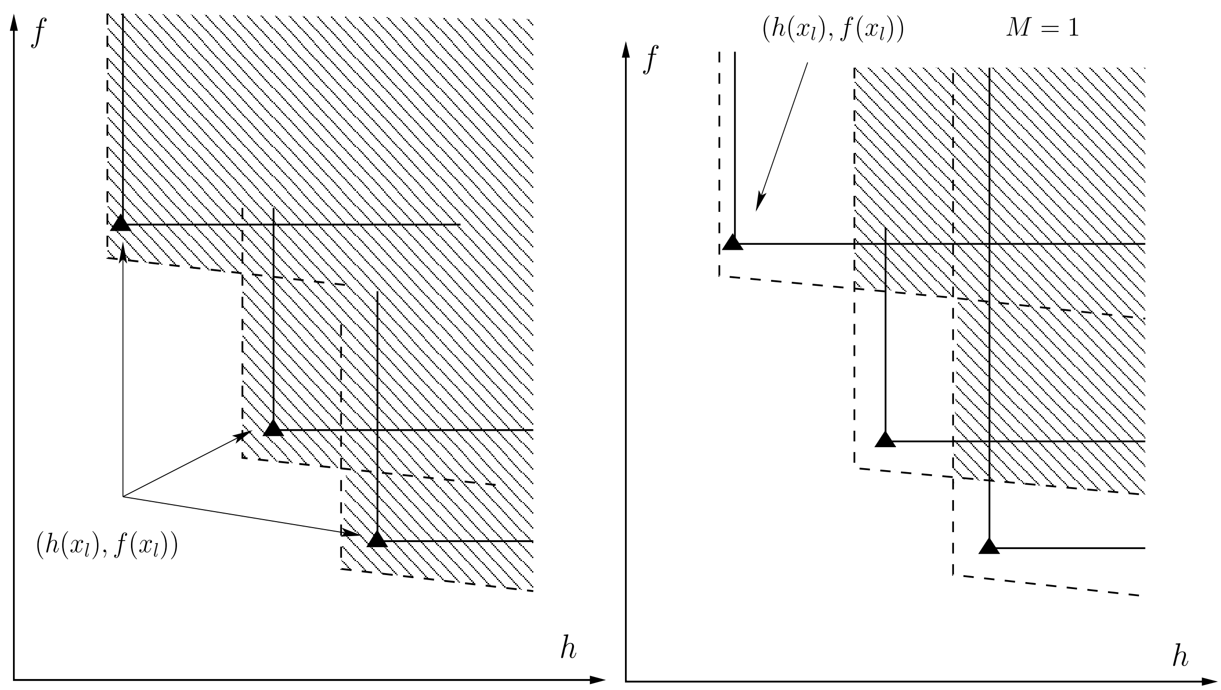

Figure 1: The left figure shows a g-filter. The black triangles correspond to three filter entries, and the shaded area shows the set of points that are dominated by these entries. The right figure shows the corresponding nonmonotone l-filter with $M = 1$.

Definition 2.2. *A point $\hat{x}$ (or $(h(\hat{x}), f(\hat{x}))$) is said to be* acceptable to the g-filter (or $\mathcal{F}_k^g$) *if $\hat{x}$ is acceptable to x_l for all $l \in \mathcal{F}_k^g$.*

Now we define the l-filter, which allows us to accept full SQP steps. Our l-filter is a new nonmonotone filter. The l-filter (see Figure 1, right) is a list of pairs $\{(h(x_k), f(x_k))\}$ such that any pair in the filter is dominated by at most M previous pairs in the filter, where $M \geq 0$ is an integer. If the number of dominated entries is zero (i.e., $M = 0$), then this filter reduces to the standard filter.

Definition 2.3. *Let $M \geq 0$ be an integer. A point $\hat{x}$ (or $(h(\hat{x}), f(\hat{x}))$) is said to be* nonmonotonically acceptable to the l-filter (or $\mathcal{F}_k^l$) *if $(h(\hat{x}), f(\hat{x}))$ is dominated by at most M pairs in $\{(h(x_l), f(x_l)) \mid l \in \mathcal{F}_k^l\}$.*

To control infeasibility of all iterates, we give an upper bound condition for accepting a point, namely

$$h(x) \leq u, \tag{2.7}$$

where u is a positive scalar, which can be implemented in the algorithm by initiating the l/g filters with the pair $(u, -\infty)$.

The two filters interact in a natural way. As long as $\|d\| = \rho$, we measure progress with the g-filter. Once we detect $\|d\| < \rho$, either we start using the l-filter, which we continue to use until we converge, or we compute a step with $\|d\| = \rho$. In the latter case, we flush the l-filter and return to using the g-filter. To prevent cycling between the two filters, we backtrack to the last iterate that was acceptable to the g-filter.

We include a new iterate $(h(x_k), f(x_k))$ in the respective filter if $h(x_k) > 0$, $\hat{x} = x_k + d$ is acceptable to both the filter and x_k and if the sufficient reduction criterion ((2.11) or (2.9)) is satisfied whenever the switching condition ((2.10) or (2.8)) holds. We note that the switching and sufficient reduction conditions differ for the two filters to accommodate fast local convergence. We also note that all entries $(h(x_j), f(x_j))$ in the l-filter have been obtained from a full SQP-step; that is, $x_j = x_j + \bar{d}_j$ for each $j \in \mathcal{F}_k^l$.

Next, we explain how a trial point is accepted in our algorithm. During the global phase, a trial point $\hat{x}$ is required to be acceptable to $\mathcal{F}_k^g \cup \{k\}$. Once we switch to the local filter, a trial point $\hat{x}$ must be nonmonotonically acceptable to $\mathcal{F}_k^l \cup \{k\}$. In addition, if the appropriate switching condition holds, then the trial point must also satisfy an appropriate sufficient reduction condition. The switching condition for the l-filter is

$$\Delta q(\bar{d}_k) > 0 \quad \text{and} \quad \tilde{h}_k \leq \zeta \|\bar{d}_k\|_\infty^\tau. \tag{2.8}$$

If the switching condition holds, then we expect that the objective is reduced over the step. A suitable nonmonotone sufficient reduction condition is

$$\Delta \tilde{f}(\bar{d}_k) \geq \sigma \min\left\{\Delta q(\bar{d}_k), \xi \|\bar{d}_k\|_\infty^2\right\}, \tag{2.9}$$

where $\zeta > 0$, $\tau \in (2, 3]$, $\sigma \in (0, \frac{1}{2})$ and $\xi > 0$. The switching condition and the sufficient reduction criterion for the g-filter are

$$\Delta q(d) > 0 \tag{2.10}$$

and

$$\Delta f(d) \geq \sigma \Delta q(d), \tag{2.11}$$

respectively.

We briefly motivate our choice of the switching condition (2.8) and (2.10) and the sufficient reduction criterion (2.9) or (2.11). For global convergence, we hope that iterates close to the feasible region of problem (1.1) also improve optimality. As in other filter methods, such as that of Fletcher and Leyffer (2002), the switching condition (2.10) and the sufficient reduction criterion (2.11) are used to achieve this goal. We note that the switching condition (2.8) is more stringent than that of Fletcher and Leyffer (2002) because the second condition $\max_{i\in\{0,\cdots,M\}} h(x_{k-i}) \le \zeta\|\bar{d}_k\|_\infty^\tau$ in (2.8) is also required. Therefore our sufficient reduction criterion is easier to satisfy than that of Fletcher and Leyffer (2002). To obtain fast local convergence, we must accept the full SQP step for all sufficiently large k. Thus, we relax the sufficient reduction criterion by strengthening the switching condition. These conditions, along with the nonmonotone acceptance condition for the l-filter, play an important role in obtaining fast local convergence.

Definition 2.4. *A trial point $\hat{x}$ is said to satisfy* the g-filter acceptance conditions *if ($\hat{x}$ is acceptable to the g-filter and x_k ($\mathcal{F}_k^g \cup \{k\}$)) and if the sufficient reduction criterion (2.11) holds whenever the switching condition (2.10) is satisfied.*

Definition 2.5. *A trial point $\hat{x}$ is said to satisfy* the nonmonotone l-filter acceptance conditions *if ($\hat{x}$ is nonmonotonically acceptable to the l-filter and x_k ($\mathcal{F}_k^l \cup \{k\}$)) and if the sufficient reduction criterion (2.9) holds whenever the switching condition (2.8) is satisfied.*

If QP(x_k, ρ) is incompatible, the algorithm switches to the feasibility restoration phase to find a new iterate that is acceptable to the current g-filter by reducing the constraint violation. Any method for solving a nonlinear algebraic system of equalities and inequalities can be used to implement this calculation. Of course, the restoration phase may converge to a nonzero local minimum of $h(x)$. On the other hand, if the iterates generated by the restoration phase are converging to a feasible point, then we can eventually find an acceptable point such that QP is consistent, unless the MFCQ condition fails. In this paper, we do not specify the particular procedure for this feasibility restoration phase.

Algorithm 2.1 Nonmonotone Filter SQP Algorithm

Given $x_0 \in \mathbb{R}^n$.
Choose constants $\sigma \in (0,1)$, $\beta \in (0,1)$, $\gamma \in (0,1)$, $\tau \in (2,3]$, $M \geq 0$, $\zeta > 0$, $\xi > 0$, $u > 0$, $\rho^o > 0$, $\rho_{\max} > 0$.
Initialize $\rho \in (\rho^o, \rho_{\max})$ and the l/g filters with $(u, -\infty)$.
Let $k := 0$, set FLAG=global
while $d \neq 0$ **do**
  repeat
    Solve QP(x_k, ρ) for a step d
    if *infeasible* **then**
      Add $(h(x_k), f(x_k))$ to the g-filter
      Enter feasibility restoration to find x_{k+1} such that QP(x_{k+1}, ρ) feasible for $\rho > \rho^o$
      Set $k := k + 1$
    else
      Set $\hat{x} = x_k + d$
      if $\|d\|_\infty < \rho$ *& FLAG=global* **then**
        Set FLAG=local and save $x_g = x_k$, $\rho_g = \|d\|_\infty$
      if *FLAG=local* **then**
        if $\hat{x}$ *is nonmonotonically acceptable to* $\mathcal{F}_k^l \cup \{k\}$ **then**
          if $\Delta \tilde{f}(d) < \sigma \min\{\Delta q(d), \xi \|d\|_\infty^2\}$, $\Delta q(d) > 0$ *and* $\tilde{h}_k \leq \zeta \|d\|_\infty^\tau$ **then**
            Set FLAG=global, flush $\mathcal{F}_k^l = \emptyset$, and return to $x_k = x_g$, $\rho = \rho_g/2$
          else
            $\hat{x}$ is accepted
        else
          Set FLAG=global, flush $\mathcal{F}_k^l = \emptyset$, and return to $x_k = x_g$, $\rho = \rho_g/2$
      else if *FLAG=global* **then**
        if $\hat{x}$ *is acceptable to* $\mathcal{F}_k^g \cup \{k\}$ **then**
          if $\Delta f(d) < \sigma \Delta q(d)$ *and* $\Delta q(d) > 0$ **then**
            Set $\rho = \rho/2$
          else
            $\hat{x}$ is accepted
        else
          Set $\rho = \rho/2$
  until $\hat{x}$ *is accepted*
  Add $(h(x_k), f(x_k))$ to the l-filter or the g-filter (depends on FLAG) when $h(x_k) > 0$
  Set $\rho_k = \rho$, $d_k = d$, $\Delta q_k = \Delta q(d)$, $x_{k+1} = x_k + d_k$ $\rho = \max(\rho^o, \min(2\rho, \rho_{\max}))$.
  Set $k := k + 1$

In Algorithm 2.1, we use FLAG to indicate which filter is considered. FLAG=local indicates that we are using the l-filter, and FLAG=global indicates that we are using the g-filter. When we leave FLAG=local, we empty the l-filter to prevent old entries from interfering with local convergence.

Backtracking to the g-filter is initiated if a new iterate cannot be accepted by the l-filter and we therefore need to reduce the trust region. We use x_g and ρ_g to record information on the latest iterate x_k that was accepted by the g-filter. When we backtrack to the g-filter, we backtrack to the last x_g. We can also stay at some iterate x_{k+l} which is accepted by the l-filter, if x_{k+l} is acceptable to the g-filter (in which case we backtrack to this point). This approach prevents iterates from oscillating between the g-filter and the l-filter.

Algorithm 2.1 has two crucial parts: the l-filter acceptance (lines 15-21) and the g-filter acceptance (lines 23-29). We switch from the g-filter to the l-filter if $\|d\|_\infty < \rho$, indicating that we are potentially generating Newton steps. We switch from the l-filter to the g-filter if we cannot accept a new point and therefore must reduce trust-region radius ρ.

In our convergence proof we use the terminology introduced by Fletcher et al. (2002b). We call d an f-type step if the switching condition (2.8) or (2.10) is satisfied, indicating that the sufficient reduction criterion (2.9) or (2.11) is required. In this case, we refer to iteration as an f-type iteration. Similarly, we call d an h-type step if the switching condition (2.8) or (2.10) is not satisfied; we refer to k as an h-type iteration. If x_k is generated by the restoration phase, we also refer to it as an h-type iteration.

3 Global Convergence Analysis

In this section, we give the global convergence of Algorithm 2.1. Under some mild conditions, we show that the iteration sequence generated by Algorithm 2.1 has at least one accumulation point that is a KKT point. Before presenting the detailed proofs, we give some standard assumptions.

A1 Let $\{x_k\}$ be generated by Algorithm 2.1, and suppose that $\{x_k\}$ are contained in a closed and compact set S of $\mathbb{R}^n$.

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

A3 The matrix B_k is uniformly bounded for all k.

A4 The Mangasarian Fromowitz constraint qualification (MFCQ) condition holds at all feasible accumulation points.

Remark 3.1. *It follows from Assumptions A1 and A2 that there exists a constant $\bar{M} > 0$, independent of k, such that $\|\nabla^2 c_i(x)\| \leq \bar{M}, i \in \mathcal{E} \cup \mathcal{I}$, $\|\nabla^2 f(x)\| \leq \bar{M}$ for all $x \in S$. Assumption A3 is expressed mathematically, without loss of generality, as $y^T B_k y \leq \bar{M}\|y\|^2$ for all $y \in \mathbb{R}^n$.*

Our proof is divided into two steps. First, we show that the iteration sequence has feasible accumulation points. Second, we prove that at least one accumulation point is a KKT point if Assumptions A1-A4 hold.

Lemma 3.1. *Consider an infinite sequence $\{(h(x_k), f(x_k))\}$ in which each pair $(h(x_k), f(x_k))$ is added to the l-filter for satisfying the nonmonotone l-filter acceptance conditions. Assume $\{f(x_k)\}$ is bounded below. Then the sequence $\{h(x_k)\}$ converges to zero.*

Proof. From Algorithm 2.1 and the upper bound condition (2.7), we have $0 < h(x_k) \leq u$ for all k. So the sequence $\{h(x_k)\}$ has at least one accumulation point. Suppose that there exists a subsequence $h(x_{k_i})$ of $\{h(x_k)\}$ such that $h(x_{k_i}) \to \bar{h}$, where $\bar{h} > 0$ is a scalar, and seek a contradiction.

If the sequence $\{f(x_{k_i})\}$ is not bounded above, then we can choose a subsequence so that it is monotonically increasing. Without loss of generality, we assume that $\{f(x_{k_i})\}$ itself has this property. Therefore,

$$f(x_{k_{i+1}}) > f(x_{k_i}) - \gamma h(x_{k_i}) \tag{3.12}$$

for all i. By the nonmonotone l-filter acceptance conditions, $x_{k_{i+1}}$ cannot be dominated by x_{k_i}, $x_{k_{i-1}}, \cdots, x_{k_{i-M}}$ at the same time. This fact, together with (3.12), yields

$$h(x_{k_{i+1}}) \leq \beta \max_{j \in \{0,\cdots,M\}} h(x_{k_{i-j}}).$$

Similarly, we also have

$$h(x_{k_{i+l}}) \leq \beta \max_{j \in \{0,\cdots,M\}} h(x_{k_{i-j}}),$$

where $l \in \{2, \cdots, M+1\}$. Hence,

$$\max_{j \in \{1,\cdots,M+1\}} h(x_{k_{i+j}}) \leq \beta \max_{j \in \{0,\cdots,M\}} h(x_{k_{i-j}}),$$

which implies $h(x_{k_i}) \to 0$. This contradicts the fact that $h(x_{k_i}) \to \bar{h} > 0$. It follows that $h(x_k) \to 0$ in this situation.

If the sequence $\{f(x_{k_i})\}$ is bounded, then there exists a subsequence of $\{(h(x_{k_i}), f(x_{k_i}))\}$ that converges to $(\bar{h}, \bar{f})$, where $\bar{f}$ is a scalar. Without loss of generality, we assume that $(h(x_{k_i}), f(x_{k_i})) \to (\bar{h}, \bar{f})$. We define $r = \frac{\bar{h}}{4}\min(1-\beta, \gamma)$. Then there exists an $i_0 > 0$ such that, for any $i \geq i_0$, $(h(x_{k_i}), f(x_{k_i}))$ lies in the neighborhood $U_{(\bar{h},\bar{f})}(r)$ of $(\bar{h}, \bar{f})$ with radius r, and $h(x_{k_i}) \geq \frac{\bar{h}}{2}$; that is,

$$(h(x_{k_i}), f(x_{k_i})) \in U_{(\bar{h},\bar{f})}(r) =: \{(x,y) \mid (x-\bar{h})^2 + (y-\bar{f})^2 < r^2\}$$

and $h(x_{k_i}) \geq \frac{\bar{h}}{2}$ for all $i \geq i_0$. We choose some $i > i_0$. Then, on the one hand, $(h(k_{i+j}), f(k_{i+j}))$ lies in $U_{(\bar{h},\bar{f})}(r)$ for $j \in \{1, \cdots, M+2\}$. Therefore,

$$|h(x_{k_{i+M+2}}) - h(x_{k_{i+j}})| \leq |h(x_{k_{i+M+2}}) - \bar{h}| + |h(x_{k_{i+j}}) - \bar{h}| < \frac{\bar{h}}{2}\min(1-\beta,\gamma) \leq \frac{\bar{h}}{2}(1-\beta) \tag{3.13}$$

and

$$|f(x_{k_{i+M+2}}) - f(x_{k_{i+j}})| \leq |f(x_{k_{i+M+2}}) - \bar{f}| + |f(x_{k_{i+j}}) - \bar{f}| < \frac{\bar{h}}{2}\min(1-\beta,\gamma) \leq \frac{\bar{h}}{2}\gamma \tag{3.14}$$

for $j \in \{1, \cdots, M+1\}$. It follows that

$$h(x_{k_{i+M+2}}) > h(x_{k_{i+j}}) - \frac{\bar{h}}{2}(1-\beta) \geq h(x_{k_{i+j}}) - h(x_{k_{i+j}})(1-\beta) = h(x_{k_{i+j}})$$

and

$$f(x_{k_{i+M+2}}) > f(x_{k_{i+j}}) - \frac{\bar{h}}{2}\gamma \geq f(x_{k_{i+j}}) - \gamma h(x_{k_{i+M+2}})$$

for $j \in \{1, \cdots, M\}$, which means that $x_{k_{i+M+2}}$ cannot be accepted by $x_{k_{i+j}}$, $j \in \{1, \cdots, M+1\}$. On the other hand, the nonmonotone l-filter acceptance conditions ensure that $x_{k_{i+M+2}}$ must be acceptable to at least one of the points $x_{k_{i+j}}, j \in \{1, \cdots, M+1\}$. This is a contradiction, which implies that the whole sequence $\{h(x_k)\}$ converges to zero. □

The following corollary follows directly from Lemma 3.1, because the g-filter is equivalent to an l-filter with $M = 0$.

Corollary 3.1. *Consider an infinite sequence $\{(h(x_k), f(x_k))\}$ in which each pair $(h(x_k), f(x_k))$ is added to the g-filter for satisfying the g-filter acceptance conditions. Assume $\{f(x_k)\}$ is bounded below. Then the sequence $\{h(x_k)\}$ converges to zero.*

From Algorithm 2.1, it follows that either $h(x_k) = 0$ or $(h(x_k), f(x_k))$ is included in the g-filter or the l-filter for all sufficiently large k. Combining Lemma 3.1 and Corollary 3.1, we obtain that the whole sequence converges to zero.

Before we show that Algorithm 2.1 is well defined and converges globally, we state some preliminary results.

Lemma 3.2. *Let Assumptions A1-A4 hold. If d is a feasible point of the subproblem QP(x_k, ρ), then it follows that*

$$\Delta f \geq \Delta q - n\rho^2 \bar{M} \tag{3.15}$$

and

$$h(x_k + d) \leq \frac{1}{2}\rho^2 mn\bar{M}. \tag{3.16}$$

Proof. By the definition of $h(x)$ and Fletcher et al. (2002b, Lemma 3), the conclusion follows.

Next, we show that in a neighborhood of a feasible but not optimal point, QP(x, ρ) has a positive predicted reduction.

Lemma 3.3. *Let Assumptions A1-A4 hold, and let $x^* \in X$ be a feasible point of problem (1.1) at which MFCQ holds but which is not a KKT point. Then there exist a neighborhood N of x^* and positive constants ϵ, ν, and $\overline{\kappa}$ such that for all $x \in X \cap$ N and all ρ for which*

$$\nu h(x) \leq \rho \leq \overline{\kappa}, \tag{3.17}$$

it follows that QP(x, ρ) has a feasible solution d. Moreover, the predicted reduction satisfies

$$\triangle q \geq \frac{1}{3}\rho\varepsilon, \tag{3.18}$$

the sufficient reduction criterion (2.11) holds, and the actual reduction satisfies

$$\Delta f(d) \geq \gamma h(x + d). \tag{3.19}$$

Proof. The conclusion follows from Fletcher et al. (2002b, Lemma 5) with slight modifications. □

Now, we prove that Algorithm 2.1 is well defined, that is, that the inner iteration terminates finitely.

Lemma 3.4. *Let Assumptions A1-A4 hold. Then the inner iteration terminates finitely.*

Proof. The conclusion follows from Fletcher et al. (2002b, Lemma 6) with slight modifications. □

We are now able to prove our global convergence result.

Theorem 3.1. *Let Assumptions A1-A4 hold, and assume that* $\mathrm{QP}(x_k, \rho)$ *is solved to global optimality. Then one of the following three cases occurs.*

(i) *The restoration phase fails to terminate and converges to a stationary point of the constraint violation.*

(ii) *A KKT point of problem (1.1) is found (*$d = 0$ *is generated for some* k*).*

(iii) *There exists at least one accumulation point* x^* *of* $\{x_k\}$ *generated from Algorithm 2.1 such that it is a KKT point.*

Proof. If the restoration phase fails to terminate or $d = 0$ for some k, cases (i) and (ii) follow trivially. Since the inner loop terminates finitely, we need only to consider that the outer iteration sequence is infinite. We distinguish two cases depending on whether there are a finite number of h-type iterations or not.

First, we consider the case that there exist an infinite number of h-type iterations contained in the main iteration sequence. If there exist an infinite number of h-type iterates added to the g-filter, then it follows from Assumption A1 and Lemma 3.1 that there exists a subsequence of this h-type sequence that converges to x^*, which is feasible for problem (1.1). Let $\mathcal{G}$ denote the index set of this subsequence. By Lemma 3.3 and Assumption A4, the feasibility of x^* implies that the subproblem QP is consistent, $f(x_k+d)-f(x_k) \geq \gamma h(x_k+d)$, and the switching condition (2.10) and the sufficient reduction condition (2.11) hold for sufficiently large k if ρ satisfies condition (3.17). This together with Algorithm 2.1 yields that $x_k + d$ is acceptable to the filter and x_k if $\rho^2 \leq \frac{2\beta h(x_k)}{mn\bar{M}}$ for sufficiently large k. Therefore, an f-type iteration is generated if

$$\nu h(x_k) < \rho \leq \min\left\{\overline{\kappa}, \sqrt{\frac{2\beta h(x_k)}{mn\bar{M}}}\right\} \tag{3.20}$$

holds. Now we show that (3.20) can be satisfied for sufficiently large k. We note that the upper bound in (3.20) is more than twice the lower bound, as $h(x_k) \to 0$. From Algorithm 2.1, a value $\rho \geq \rho^o$ is chosen at the beginning of each iteration. Then it will be greater than the upper bound in (3.20) for sufficiently large k. Hence, by successively halving ρ in the inner loop, we will eventually locate ρ in the range of (3.20) or to the right of this interval. Since d is a global optimizer of $\mathrm{QP}(x_k, \rho)$, the predicted reduction $\Delta q(d)$ decreases monotonically as ρ decreases. As a result, no h-type iterations are generated for ρ larger than the upper bound in (3.20). Therefore, for sufficiently large $k \in \mathcal{G}$, an f-type iteration is generated that contradicts the definition of $\mathcal{G}$. Therefore, x^* must be a KKT point of problem (1.1).

Next, we consider the case where an infinite number of h-type iterates is added to the l-filter while only a finite number of h-type iterates are added to the g-filter. Let $\mathcal{L}$ denote the index set such that each $k \in \mathcal{L}$ is an h-type iterate added to the l-filter. Assumption A1 ensures that the sequence $\{x_k\}$ has at least one accumulation point. If there exists an infinite subset $\mathcal{K}$ such that $d = \bar{d}_k,\ k \in \mathcal{K}$ and $\{\bar{d}_k\}_{\mathcal{K}}$ converges to a zero vector, then $\{x_k\}$ must have an accumulation point that is a KKT point, which completes the proof. Now we assume that $\|\bar{d}_k\|_\infty \geq \bar{\epsilon}$ for some scalar $\bar{\epsilon} > 0$. It follows that from Lemma 3.1 and Corollary 3.1 that the second inequality of (2.8) can be satisfied by choosing k large enough. Similar to the earlier proof, for sufficiently large k, if ρ satisfies (3.20), then k is an f-type iteration. Even if ρ lies in the right of interval (3.20), the condition (2.8) is satisfied. Then, any $k \in \mathcal{L}$ sufficiently large could not be an h-type iteration, which contradicts the definition of $\mathcal{L}$. Therefore, x^* is a KKT point of problem (1.1).

Now, we consider the case that only a finite number of h-type iterations are generated. Then there exists an integer $K > 0$ such that for all $k \geq K$, k is an f-type iteration. We consider two subcases in the following. One is that there exists an integer $K_1 \geq K$ such that $d = \bar{d}_k$ for all $k \geq K_1$. By (2.9), we have that for any $k \geq K_1$,

$$\max_{j \in \{0,\cdots,M\}} f(x_{k+l-j-1}) - f(x_{k+l}) \geq \sigma \min\left\{\Delta q(\bar{d}_{k+l-1}), \xi\|\bar{d}_{k+l-1}\|_\infty^2\right\} \geq 0 \tag{3.21}$$

for $l \in \{1, \cdots, M+1\}$. Then the sequence $\left\{\max_{j \in \{0,\cdots,M\}} f(x_{k-j})\right\}$ decreases monotonically. This together with (3.21) gives

$$\max_{j \in \{0,\cdots,M\}} f(x_{k-j}) - \max_{j \in \{1,\cdots,M+1\}} f(x_{k+j}) \geq \sigma \min_{j \in \{1,\cdots,M+1\}} \left\{\Delta q(\bar{d}_{k+j-1}), \xi\|\bar{d}_{k+j-1}\|_\infty^2\right\}$$

for all $k \geq K_1$. Since Assumptions A1-A2 imply boundedness of f, it follows that

$$\min_{j \in \{1,\cdots,M+1\}} \left\{\Delta q(\bar{d}_{k+j-1}), \xi\|\bar{d}_{k+j-1}\|_\infty^2\right\} \to 0 \tag{3.22}$$

as $k \to \infty$. We define

$$\overline{\mathcal{K}} = \left\{ l \mid \min\{\Delta q(\bar{d}_l), \xi\|\bar{d}_l\|^2\} = \min_{j \in \{1,\cdots,M+1\}} \{\Delta q(\bar{d}_{k+j-1}), \xi\|\bar{d}_{k+j-1}\|_\infty^2\},\ k \geq K_1 \right\}.$$

Without loss of generality, we assume that $\{x_k\}_{k \in \overline{\mathcal{K}}}$ converges to x^*, which is a feasible point for problem (1.1) from Lemma 3.1. By Lemma 3.3, if

$$\nu h(x_k) < \rho \leq \bar{\kappa},\ k \in \overline{\mathcal{K}}, \tag{3.23}$$

then $\Delta q(d) \geq \frac{1}{3}\rho\epsilon$. Since $h(x_k) \to 0$, the radius ρ must lie in the interval (3.23) or the right of this interval. The global optimality of d ensures that

$$\Delta q(\bar{d}_k) \geq \frac{1}{3}\rho\epsilon > \frac{\epsilon}{3}\|\bar{d}_k\|_\infty.$$

This together with (3.22) and the definition of $\overline{\mathcal{K}}$ implies that $\|\bar{d}_k\|_\infty \to 0,\ k \in \overline{\mathcal{K}}$. Therefore, x^* is a KKT point.

Now, we consider the other subcase that all $d \neq \bar{d}_k$ for all $k \geq K$; in other words, all sufficiently large iterations are added to the g-filter, which is similar to situation discussed by Chin and Fletcher (2003). It then follows that $\{f(x_k)\}$ is monotone for all $k \geq K$. Since f is bounded below, the sufficient reduction criterion (2.11) gives $\Delta q(d) \to 0$ as $k \to \infty$. Let x^* be an accumulation point of $\{x_k\}$.

We define

$$\tau_K = \min_{j \in \mathcal{F}_K^g, h(x_j) > h(x_k)} h(x_j). \tag{3.24}$$

From Lemmas 3.2 and 3.3, if

$$\nu h(x_k) < \rho \leq \min\left\{ \sqrt{\frac{\beta \tau_K}{mn\bar{M}}}, \bar{\kappa} \right\},\ k \geq K, \tag{3.25}$$

then (3.18), (3.19), and (2.11) are satisfied. Thus, $x_k + d$ is acceptable to x_k and all x_j with $h(x_j) > h(x_K)$, $j \in \mathcal{F}_K^g$. For all j with $h(x_j) \leq h(x_K)$, $j \in \mathcal{F}_K^g$, we must have $f(x_j) > f(x_K)$; otherwise $(h(x_j), f(x_j))$ must have been deleted. It follows from the monotonicity of $\{f(x_k)\}$ for all $k \geq K$ that $f(x_j) > f(x_k)$ for all $j \in \{K, \cdots, k-1\}$ and all j with $h(x_j) \leq h(x_K)$, $j \in \mathcal{F}_K^g$. This together with (3.19) yields that $x_k + d$ is acceptable to all x_j for all $j \in \{K, K+1, \cdots, k-1\}$ and all j with $h(x_j) \leq h(x_K)$, $j \in \mathcal{F}_K^g$. Similar to the earlier proof, an f-*type* iteration is generated when (3.25) is satisfied. The right-hand side of (3.25) is a constant, independent of k. Since the upper bound of (3.25) 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, $\rho \geq \min\{\frac{1}{2}\bar{\kappa}, \rho^o\}$. The global optimality of d_k ensures $\Delta q(d) \geq \frac{1}{3}\epsilon \min\{\frac{1}{2}\bar{\kappa}, \rho^o\}$ holds even if ρ is greater than the right-hand side of (3.25). This contradicts the fact $\Delta q(d) \to 0$ as $k \to \infty$. Thus, x^* is a KKT point. $\square$

We are aware that requiring global solutions of the QP subproblems in our global convergence analysis is undesirable. The same assumption was used by Fletcher et al. (2002b). Later, Fletcher et al. (2002a) proposed a trust-region SQP-filter algorithm that uses a decomposition of the step in its normal and tangential components. Under some mild conditions, they obtained global convergence without requiring the global solutions of the QP subproblems. As a matter of fact, we can remove the global optimality assumption by using this decomposition technique and weaker assumptions (Fletcher et al., 2002a, (Equations (2.12) and (2.15)). These assumptions can be guaranteed by implementation of algorithm if the generalized Cauchy step is generated by solving one additional linear program subproblem. Similar to Fletcher et al. (2002a, Lemmas 3.5-3.7), we can obtain that

$$\Delta q(d) \geq \kappa \rho \epsilon,$$

if $\chi_k \geq \epsilon$ and $0 < \rho < \delta_m$, where $\kappa > 0$, $\delta_m > 0$, and $\epsilon > 0$ are scalars and χ_k is the measure of first-order criticality (Fletcher et al., 2002a, Equation (2.13)). Applying this conclusion to Theorem 3.1, we obtain global convergence without requiring the global optimality condition.

4 Local Convergence Analysis

In this section, we prove the local convergence properties of Algorithm 2.1. As we mentioned earlier, the l-filter promotes fast local convergence. We will prove that when the iterates approach a local optimal point, the nonmonotone l-filter conditions are satisfied for all Newton steps and that all iterates with $h(x) > 0$ are added to the l-filter. Therefore, fast local convergence is achieved.

Let x^* be an accumulation point of $\{x_k\}$ generated by Algorithm 2.1, which is a KKT point of the problem (1.1). The corresponding multiplier is denoted by $\lambda^* = (\lambda_1^*, \cdots, \lambda_m^*)$. Before stating the main results, we need some additional assumptions.

A5 Let f and $c_i,\ i \in \mathcal{E} \cup \mathcal{I}$ be twice continuously differentiable with Lipschitz continuous Hessian. The point x^* associated with its multiplier λ^* satisfies the linear independence constraint qualification (LICQ), the strict complementarity condition (SCC), and the second-order sufficient conditions (SOSC). That is,

1. $\nabla c_{\mathcal{E}\cup\mathcal{I}^*}(x^*)$ has full column-rank, where $\mathcal{I}^* := \{i \mid c_i(x^*) = 0, i \in \mathcal{I}\}$;
2. $\lambda_i^* > 0$, $i \in \mathcal{I}^*$; $\lambda_i^* = 0$, $i \in \mathcal{I}\backslash\mathcal{I}^*$; and
3. $y^T\nabla^2 L(x^*, \lambda^*)y \geq \kappa\|y\|^2$ holds for all y satisfying $\nabla c_i(x^*)^T y = 0, i \in \mathcal{E} \cup \mathcal{I}^*$, where $\kappa > 0$ is a scalar.

From the previous section, we know that $\bar{d}_k$ is the solution of QP(x_k, ∞) for all k. Then $\bar{d}_k$ satisfies the KKT conditions of QP(x_k, ∞), namely,

$$\begin{cases} \nabla f(x_k) + \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_{k,i}\nabla c_i(x_k) + B_k\bar{d}_k = 0, \\ \nabla c_i(x_k)^T\bar{d}_k + c_i(x_k) = 0,\ i \in \mathcal{E}, \\ \big(\nabla c_i(x_k)^T\bar{d}_k + c_i(x_k)\big)\lambda_{k,i} = 0,\ i \in \mathcal{I}, \\ \lambda_{k,i} \geq 0, \nabla c_i(x_k)^T\bar{d}_k + c_i(x_k) \leq 0,\ i \in \mathcal{I}, \end{cases} \tag{4.26}$$

where $B_k = \nabla^2 f(x_k) + \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_{k-1,i}\nabla^2 c_i(x_k)$ is the Hessian of the Lagrangian, and $\lambda_k = (\lambda_{k,1}, \cdots, \lambda_{k,m})^T \in \mathbb{R}^m$.

To obtain fast convergence, we need to prove two results. One is that the Newton step $\bar{d}_k$ is computed for all sufficiently large k. The other is that the Newton step $\bar{d}_k$ is accepted for all sufficiently large k. As we discussed in Section 2, the Newton step $\bar{d}_k$ is not computed explicitly at any iteration. However, if the solution d of QP(x_k, ρ) satisfies $\|d\|_\infty < \rho$, then d is the Newton step $\bar{d}_k$ provided Assumption A5 holds. From the mechanism of Algorithm 2.1, the first trial trust-region radius ρ is always greater than or equal to the constant ρ^o. Therefore, we need only to prove that $\bar{d}_k \to 0$ as $k \to +\infty$, and then all d solving QP(x_k, ρ) with $\rho \geq \rho^o$ are Newton steps, which implies that $\bar{d}_k$ is computed for all sufficiently large k. In Lemmas 4.1 and 4.2 and Proposition 4.1 we show that $\bar{d}_k \to 0$ as $k \to +\infty$.

Lemma 4.1. *Let Assumption A5 hold. If $(x_k, \lambda_k) \to (x^*, \lambda^*)$ as $k \to \infty$ and $k \in \mathcal{K}$, where $\mathcal{K}$ is an infinite index set, then $\|\bar{d}_k\| \to 0$ as $k \to \infty$ and $k \in \mathcal{K}$.*

Proof. Since x^* is a local minimizer of the problem (NLP), it follows with Assumption A5 that there exist no strictly feasible descent directions, that is,

$$D' \cap F' = \{0\}, \tag{4.27}$$

where $D' = \{d \mid \nabla f(x^*)^T d < 0\}$ and $F' = \{d \mid \nabla c_i(x^*)^T d = 0, i \in \mathcal{E}; \nabla c_i(x^*)^T d \leq 0, i \in \mathcal{I}^*\}$. We distinguish two cases, depending on whether the sequence $\{\bar{d}_k\}$ is bounded or not.

If the sequence $\{\bar{d}_k\}$ is bounded, then it must have a convergent subsequence. Suppose that there exists an infinite set $\mathcal{K}' \subseteq \mathcal{K}$ such that $\{d_k\}_{\mathcal{K}'} \to \bar{d} \neq 0$. In view of KKT conditions (4.26), we obtain the following systems:

$$\nabla f(x_k)^T \bar{d}_k = -\sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_{k,i} c_i(x_k) - \bar{d}_k^T B_k \bar{d}_k, \tag{4.28}$$

$$\nabla c_i(x_k)^T \bar{d}_k + c_i(x_k) = 0,\ i \in \mathcal{E}, \tag{4.29}$$

$$\nabla c_i(x_k)^T \bar{d}_k + c_i(x_k) \leq 0,\ i \in \mathcal{I}. \tag{4.30}$$

Letting k tend to infinity, we obtain

$$\nabla f(x^*)^T \bar{d} = -\bar{d}^T \nabla^2 L(x^*, \lambda^*) \bar{d} < 0 \tag{4.31}$$

and $\bar{d} \in F'$, where the last inequality of (4.31) follows from Assumption A5. However, $0 \neq \bar{d} \in D' \cap F'$, which contradicts (4.27). Therefore, $\{\bar{d}_k\}_{\mathcal{K}} \to 0$ in this situation.

If the sequence $\{\bar{d}_k\}$ is unbounded, then its normalized sequence $\{\bar{d}_k / \|\bar{d}_k\|\}$ must be bounded. Suppose that there is a $\mathcal{K}'$ such that $\bar{d}_k / \|\bar{d}_k\| \to \bar{d} \neq 0$ and $\|\bar{d}_k\| \to \infty$ as $k \in \mathcal{K}'$ and $k \to \infty$. Dividing (4.28) by $\|\bar{d}_k\|^2$ and dividing (4.29)-(4.30) by $\|\bar{d}_k\|$, we obtain

$$\nabla f(x_k)^T \bar{d}_k / (\|\bar{d}_k\|^2) = -\sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_{k,i} c_i(x_k) / (\|\bar{d}_k\|^2) - \bar{d}_k^T B_k \bar{d}_k / (\|\bar{d}_k\|^2), \tag{4.32}$$

$$\nabla c_i(x_k)^T \bar{d}_k / \|\bar{d}_k\| + c_i(x_k) / \|\bar{d}_k\| = 0,\ i \in \mathcal{E}, \tag{4.33}$$

$$\nabla c_i(x_k)^T \bar{d}_k / \|\bar{d}_k\| + c_i(x_k) / \|\bar{d}_k\| \leq 0,\ i \in \mathcal{I}. \tag{4.34}$$

Taking the limit as $k \to \infty$, we obtain

$$0 = -\bar{d}^T \nabla^2 L(x^*, \lambda^*) \bar{d} < 0 \tag{4.35}$$

and $\bar{d} \in F'$, where the last inequality of (4.35) follows from Assumption A5, which is a contradiction. Therefore, $\{\bar{d}_k\}_{\mathcal{K}} \to 0$. □

Proposition 4.1. *Assume $w^* \in \mathbb{R}^t$ is an isolated accumulation point of a sequence $\{w_k\} \subseteq \mathbb{R}^t$ such that for every subsequence $\{w_k\}_{\mathcal{K}}$ converges to w^*. Assume, moreover, that there exists an infinite subset $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{\|w_{k+1} - w_k\|_{\bar{\mathcal{K}}}\} \to 0$. Then the whole sequence $\{w_k\}$ converges to w^*.*

Proof. See Moré and Sorensen (1983, Lemma 4.10) or Qi and Qi (2000, Proposition 5.4). □

Lemma 4.2. *Let Assumption A5 hold. Then the whole sequence $\{(x_k, \lambda_k)\}$ converges to (x^*, λ^*).*

Proof. Assumption A5 implies that x^* is an isolated solution of the problem (1.1); see (Robinson, 1980, Theorems 2.4, 4.2). Let $\{x_k\}_{\mathcal{K}}$ be a subsequence of $\{x_k\}$ converging to x^*. By Lemma 4.1, there exists an infinite set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{\bar{d}_k\}_{\bar{\mathcal{K}}} \to 0$. The mechanism of Algorithm 2.1 guarantees that

$$\|x_{k+1} - x_k\| = \|d_k\| \leq \|\bar{d}_k\|,$$

where d_k is from Algorithm 2.1, that is, an accepted step. Hence,

$$\{\|x_{k+1} - x_k\|\}_{\bar{\mathcal{K}}} \to 0,$$

which together with Proposition 4.1 yields $x_k \to x^*$ as $k \to \infty$. Since Assumption (A5) implies the uniqueness of multipliers associated with x^*, the sequence $\{\lambda_k\}$ converges to λ^*. □

It follows from Lemma 4.1 and Lemma 4.2 that $\bar{d}_k \to 0$ as $k \to \infty$. Next, we show that the Newton step provides superlinear convergence.

Lemma 4.3. *Let Assumption A5 hold. Then it follows that*

$$\|x_k + \bar{d}_k - x^*\| = o(\|x_k - x^*\|) \tag{4.36}$$

and

$$\left\| \begin{array}{c} x_k + \bar{d}_k - x^* \\ \lambda_k - \lambda^* \end{array} \right\| = \mathcal{O}\left(\left\| \begin{array}{c} x_k - x^* \\ \lambda_{k-1} - \lambda^* \end{array} \right\|^2 \right). \tag{4.37}$$

Moreover,

$$\|\bar{d}_k\| = \Theta(\|x_k - x^*\|). \tag{4.38}$$

Proof. Equations (4.36) and (4.37) follow from Facchinei and Lucidi (1995, Theorem 4.1). Using (4.36), we have

$$\frac{\|x_k + \bar{d}_k - x^*\|}{\|x_k - x^*\|} \geq \left| \frac{\|\bar{d}_k\|}{\|x_k - x^*\|} - 1 \right| \to 0, \text{ as } k \to \infty.$$

Therefore,

$$\frac{\|\bar{d}_k\|}{\|x_k - x^*\|} \to 1, \text{ as } k \to +\infty,$$

which implies (4.38). □

From Lemma 4.3, it follows that if Newton steps are accepted for all sufficiently large k, then Algorithm 2.1 has a superlinear rate of convergence for the primal variable x and a quadratic rate of convergence for the primal-dual pair (x, λ). Next, we establish some preliminary results for proving l-filter acceptance of $\bar{d}_k$ for sufficiently large k.

Lemma 4.4. *Let Assumption A5 hold. Then*

$$c_i(x_k + \bar{d}_k) = \mathcal{O}(\|\bar{d}_k\|^2), \; i \in \mathcal{E} \cup \mathcal{I}^* \tag{4.39}$$

holds for all sufficiently large k.

Proof. Lemmas 4.1 and 4.2 and Assumption A5 ensure that QP(x_k, ∞) is equivalent to

$$EQP(x_k) \left\{ \begin{array}{ll} \underset{d}{\text{minimize}} & q(d) = \nabla f(x_k)^T + \frac{1}{2} d^T B_k d \\ \text{subject to} & \nabla c_i(x_k)^T d + c_i(x_k) = 0, i \in \mathcal{E} \cup \mathcal{I}^*, \end{array} \right. \tag{4.40}$$

when x_k is sufficiently close to x^*. Thus, it follows that

$$\nabla c_i(x_k)^T \bar{d}_k + c_i(x_k) = 0, i \in \mathcal{E} \cup \mathcal{I}^*,$$

for all sufficiently large k. The conclusion follows with Taylor expansion and Assumption A5. □

To prove the local convergence of Algorithm 2.1, we introduce the exact penalty function

$$\Phi_\psi(x) = f(x) + \psi h(x), \tag{4.41}$$

where $\psi > \|\lambda^*\|_\infty$ is the penalty parameter. We emphasize that we use the penalty function only as a proof technique. The following result is based on the penalty function, which plays a key role in proving acceptance of the Newton step $\bar{d}_k$.

Lemma 4.5. *Let Assumption A5 hold, let $x_{k+i-1} = x_{k+i-2} + \bar{d}_{k+i-2},\ i \in \{0, 1, 2\}$, and let $\psi > \|\lambda^*\|_\infty$. Then there exists an integer $K_1 > 0$ such that for all $k \geq K_1$*

$$\Phi_\psi(x_{k+i-2}) - \Phi_\psi(x_{k+1}) \geq \big(\gamma + (\frac{1}{\beta} - 1)\psi\big)h(x_{k+1}),\ i \in \{0, 1\}, \tag{4.42}$$

holds.

Proof. From a Taylor expansion of the Lagrangian and the KKT conditions of problem (1.1), we have that

$$\begin{aligned} f(x_{k+1}) + \sum_{i \in \mathcal{E} \cup \mathcal{I}^*} \lambda_i^* c_i(x_{k+1}) - f(x^*) &= L(x_{k+1}, \lambda^*) - L(x^*, \lambda^*) \\ &= \nabla_x L(x^*, \lambda^*)(x_{k+1} - x^*) + \mathcal{O}(\|x_{k+1} - x^*\|^2) \\ &= \mathcal{O}(\|x_{k+1} - x^*\|^2). \end{aligned}$$

Rearranging this equation gives

$$f(x_{k+1}) = f(x^*) - \sum_{i \in \mathcal{E} \cup \mathcal{I}^*} \lambda_i^* c_i(x_{k+1}) + \mathcal{O}(\|x_{k+1} - x^*\|^2). \tag{4.43}$$

It follows from (4.43) and Lemma 4.4 that

$$\begin{aligned} & \Phi_\psi(x_{k+1}) + (\gamma + (\frac{1}{\beta} - 1)\psi)h(x_{k+1}) \\ = \ & f(x_{k+1}) + (\gamma + \frac{\psi}{\beta})h(x_{k+1}) \\ = \ & f(x^*) - \sum_{i \in \mathcal{E} \cup \mathcal{I}^*} \lambda_i^* c_i(x_{k+1}) + \big(\gamma + \frac{\psi}{\beta}\big)h(x_{k+1}) + \mathcal{O}(\|x_{k+1} - x^*\|^2) \\ = \ & f(x^*) + \mathcal{O}(\|x_{k+1} - x^*\|^2) + \mathcal{O}(\|\bar{d}_k\|^2). \end{aligned}$$

Substituting (4.38) and (4.36) into this equation, we have

$$\Phi_\psi(x_{k+1}) + \big(\gamma + (\frac{1}{\beta} - 1)\psi\big)h(x_{k+1}) = f(x^*) + o(\|x_{k+i-2} - x^*\|^2),\ i \in \{0, 1\}. \tag{4.44}$$

On the other hand, from Chamberlain et al. (1982, Lemma 1) and Assumption A5, we obtain that there exists a scalar $\bar{c} > 0$ such that when x is sufficiently close to x^*,

$$\Phi_\psi(x) \geq f(x^*) + \bar{c}\|x - x^*\|^2. \tag{4.45}$$

Combining this equation with (4.44) gives

$$\begin{aligned}\Phi_\psi(x_{k+i-2}) &\geq f(x^*) + \bar{c}\|x_{k+i-2} - x^*\|^2 \\ &\geq \Phi_\psi(x_{k+1}) + \big(\gamma + (\frac{1}{\beta} - 1)\psi\big)h(x_{k+1}),\ i \in \{0, 1\}\end{aligned}$$

for all $k \geq K_1$, where $K_1 > 0$ is an integer. □

The following lemma shows that the sufficient reduction criterion (2.9) holds if the switching condition (2.8) is satisfied for all sufficiently large k. Therefore, for all sufficiently large k, the Newton step $\bar{d}_k$ will not be rejected by the sufficient reduction criterion (2.9).

Lemma 4.6. *Let Assumption A5 hold. Then there exists an integer $K_2 \geq K_1$ (K_1 is given by Lemma 4.5) such that if (2.8) holds for $k \geq K_2$, then (2.9) holds for $x_{k+i} = x_{k+i-1} + \bar{d}_{k+i-1},\ i \in \{0, 1\}$.*

Proof. We need only to prove that

$$f(x_k + \bar{d}_k) + \sigma\xi\|\bar{d}_k\|^2 \leq f(x_{k-1}) \tag{4.46}$$

holds for all sufficiently large k. Condition (2.8) implies that

$$h(x_{k-1}) = \mathcal{O}(\|\bar{d}_k\|^\tau),$$

where $\tau \in (2, 3]$. This together with (4.43) yields

$$\begin{aligned}&f(x_k + \bar{d}_k) + \sigma\xi\|\bar{d}_k\|^2 + \psi h(x_{k-1}) \\ =\ & f(x^*) - \sum_{i \in \mathcal{E}\cup\mathcal{I}^*} \lambda_i^* c_i(x_k + \bar{d}_k) + \sigma\xi\|\bar{d}_k\|^2 + \psi h(x_{k-1}) + \mathcal{O}(\|x_k + \bar{d}_k - x^*\|^2) \\ =\ & f(x^*) + \mathcal{O}(\|\bar{d}_k\|^2) + \mathcal{O}(\|x_k + \bar{d}_k - x^*\|^2) \\ =\ & f(x^*) + o(\|x_{k-1} - x^*\|^2),\end{aligned}$$

where the second equality holds because of (4.39) and the third equality holds because of (4.36) and (4.38). Using (4.45), we obtain

$$\Phi_\psi(x_{k-1}) \geq f(x_k + \bar{d}_k) + \sigma\xi\|\bar{d}_k\|^2 + \psi h(x_{k-1})$$

for all $k \geq K_2$. This together with the definition of $\Phi_\psi(x)$ yields (4.46). □

We illustrate our proof in Figure 2. The next lemma shows that any pair $(h(\hat{x}), f(\hat{x}))$ on the line

$$l_1 :\ f = -\psi h + f(\hat{x}) + \psi h(\hat{x})$$

is acceptable to any pair $(h(x_l), f(x_l))$ on and above the line

$$l_2 :\ f = -\psi h + f(x_l) + \psi h(x_l)$$

so long as the intercept on the f-axis of the line l_1 is $\big(\gamma + (\frac{1}{\beta} - 1)\psi\big)h(\hat{x})$ less than that of the line l_2. In fact, $(h(\hat{x}), f(\hat{x}))$ is acceptable to A, B, and C since they are all above the line l_2.

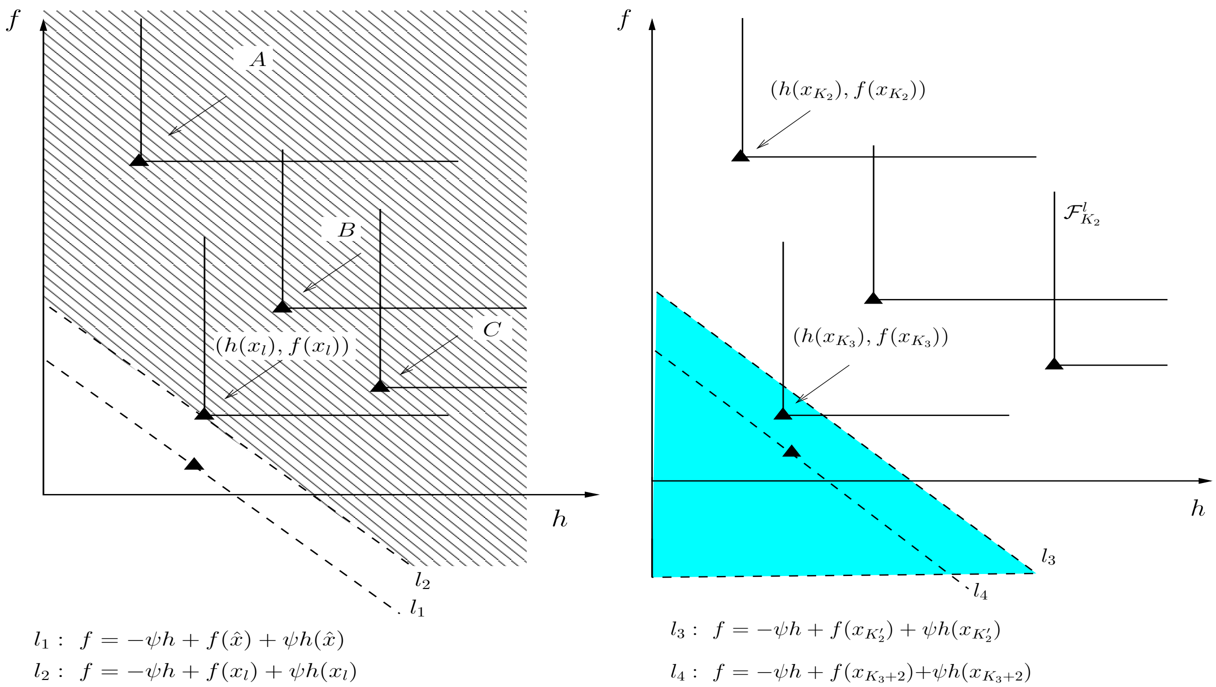

Figure 2: The left figure shows that the pair corresponding to the black triangle on the line l_1 is acceptable to any pair corresponding to the triangle in the shaded area. The right figure shows the pair $(h(x_{K_3}), f(x_{K_3}))$ is the first entry in the l-filter entering into the area $\mathcal{D}_{K_2'}(\psi)$.

Lemma 4.7. *Let $\hat{x}$ be a trial point. For any point x_l, if*

$$\Phi_\psi(x_l) - \Phi_\psi(\hat{x}) \geq \big(\gamma + (\frac{1}{\beta} - 1)\psi\big)h(\hat{x}), \tag{4.47}$$

then $\hat{x}$ is acceptable to x_l.

Proof. If $h(\hat{x}) \leq \beta h(x_l)$, then $\hat{x}$ is acceptable to x_l. Otherwise, $h(\hat{x}) > \beta h(x_l)$. Since Equation (4.47) can be rewritten as

$$f(x_l) - f(\hat{x}) \geq \psi\big(\frac{1}{\beta}h(\hat{x}) - h(x_l)\big) + \gamma h(\hat{x}),$$

it follows that $f(x_l) - f(\hat{x}) > \gamma h(\hat{x})$, which also implies that $\hat{x}$ is acceptable to x_l. Therefore, the conclusion follows in both cases. $\square$

In what follows, we consider an infinite sequence of iterations contained in the main iteration sequence. Figure 2 (right) gives the (h, f) half-plane with the l-filter. We define

$$\mathcal{D}_k(\psi) = \big\{(h, f) \mid f \leq -\psi h + f(x_k) + \psi h(x_k) \text{ and } h \geq 0\big\}.$$

Since all the entries entered into the l-filter have $h(x) > 0$, there exist an integer $K_2' > K_2$ and $\psi > \|\lambda^*\|_\infty$ such that $\forall\ (h, f) \in \mathcal{D}_{K_2'}(\psi) \Rightarrow (h, f)$ is acceptable to $\mathcal{F}^l_{K_2}$, where K_2 is from Lemma 4.6. Without loss of generality, we assume that K_3 is the first iteration $K_3 > K_2$ in the l-filter such that $(h(x_{K_3}), f(x_{K_3})) \in \mathcal{D}_{K_2'}(\psi)$.

Next, we prove that the Newton step $\bar{d}_k$ is accepted by the l-filter for all sufficiently large k. The following lemma enables us to achieve our main results.

Lemma 4.8. *Let Assumption A5 hold. Then there exists an integer $K_3 \geq K_2$ (K_2 is given by Lemma 4.6) such that the trial point $x_k + \bar{d}_k$ is accepted by the l-filter for all $k \geq K_3$.*

Proof. Taking K_3, we have $x_{K_3+1} = x_{K_3} + \bar{d}_{K_3}$ from the property of the l-filter. First, we prove that $x_{K_3+2} = x_{K_3+1} + \bar{d}_{K_3+1}$ is again the Newton step. Since K_3 is the first iteration in which $(h(x_{K_3}), f(x_{K_3})) \in \mathcal{D}_{K_2'}(\psi)$, it follows that

$$\Phi_\psi(x_{K_3}) \leq \Phi_\psi(x_l)$$

holds for all $l \in \mathcal{F}_{K_3} \cup \{K_3\}$. It then follows with Lemma 4.5 that

$$\Phi_\psi(x_l) - \Phi_\psi(x_{K_3+1} + \bar{d}_{K_3+1}) \geq (\gamma + (\frac{1}{\beta} - 1)\psi) h(x_{K_3+1} + \bar{d}_{K_3+1}) \tag{4.48}$$

holds for all $l \in \mathcal{F}_{K_3} \cup \{K_3\}$. In view of Lemma 4.7, $x_{K_3+1} + \bar{d}_{K_3+1}$ is acceptable to x_{K_3} and the filter $\mathcal{F}_{K_3}$. Thus, $x_{K_3+1} + \bar{d}_{K_3+1}$ is acceptable to the filter $\mathcal{F}_{K_3+1}$. Whether $x_{K_3+1} + \bar{d}_{K_3+1}$ is acceptable to x_{K_3+1} or not, the nonmonotone l-filter acceptance conditions are satisfied. If the condition (2.8) is also satisfied, then it follows with Lemma 4.6 that an f-type iteration is generated. Otherwise, an h-type iteration is generated. Therefore, $x_{K_3+2} = x_{K_3+1} + \bar{d}_{K_3+1}$.

In the following, we prove that $x_k = x_{k-1} + \bar{d}_{k-1}$ is accepted as a new iterate for all $k > K_3 + 2$ by induction. Denote $i := k - K_3$. For $p = 2$, the above proof has shown that $x_{K_3+p} = x_{K_3+p-1} + \bar{d}_{K_3+p-1}$ is accepted as a new iterate. Assume that $x_{K_3+p} = x_{K_3+p-1} + \bar{d}_{K_3+p-1}$ holds for any $p < i$. We need to prove that $x_{K_3+p} = x_{K_3+p-1} + \bar{d}_{K_3+p-1}$ holds for $p = i$. From the induction hypothesis and Lemma 4.5, we obtain that

$$\Phi_\psi(x_{K_3+j} + \bar{d}_{K_3+j}) \leq \Phi_\psi(x_{K_3+j-2}) - \big(\gamma + (\frac{1}{\beta} - 1)\psi\big) h(x_{K_3+j} + \bar{d}_{K_3+j})$$

and

$$\Phi_\psi(x_{K_3+j} + \bar{d}_{K_3+j}) \leq \Phi_\psi(x_{K_3+j-1}) - \big(\gamma + (\frac{1}{\beta} - 1)\psi\big) h(x_{K_3+j} + \bar{d}_{K_3+j})$$

for $j \in \{2, \cdots, i-1\}$. It then follows that

$$\Phi_\psi(x_{K_3+i-1} + \bar{d}_{K_3+i-1}) \leq \Phi_\psi(x_{K_3+j}) - \big(\gamma + (\frac{1}{\beta} - 1)\psi\big) h(x_{K_3+i-1} + \bar{d}_{K_3+i-1}) \tag{4.49}$$

for $j \in \{0, \cdots, i-2\}$. Since K_3 is the first iteration $K_3 > K_2$ in the l-filter such that $(h(x_{K_3}), f(x_{K_3})) \in \mathcal{D}_{K_2'}(\psi)$, it follows that

$$\Phi_\psi(x_{K_3}) \leq \Phi_\psi(x_j)$$

for all $j \in \mathcal{F}^l_{K_3}$. These together with Lemma 4.7 yield that $x_{K_3+i-1} + \bar{d}_{K_3+i-1}$ is acceptable to x_j for $j \in \{K_3, \cdots, K_3 + i - 2\} \cup \mathcal{F}_{K_3}$. Therefore the nonmonotone l-filter acceptance conditions are satisfied. Similar to the earlier proof, for $p = i$, we also have $x_{K_3+p} = x_{K_3+p-1} + \bar{d}_{K_3+p-1}$. Therefore, by induction, the claim of this theorem is true. □

Lemmas 4.3 and 4.8 imply the main result of this section stated in the following.

Theorem 4.1. *Let Assumption A5 hold. The sequence $\{x_k\}$ generated by Algorithm 2.1 converges to x^* q-superlinearly, and the sequence $\{(x_k, \lambda_k)\}$ converges to (x^*, λ^*) q-quadratically.*

5 Numerical Experience

We summarize our experience with a preliminary version of the second-order filter method described in Algorithm 2.1. Our goal is to demonstrate that the approach is viable and comparable to our previous implementation. Detailed computational tests and comparisons with other solvers are left for later.

We choose all 411 CUTEr (Bongartz et al., 1995) problems with less than 100 variables or constraints that are available in AMPL (Fourer et al., 2003) from Bob Vanderbei's collection (Benson and Vanderbei, 1998). We compare the established filterSQP solver (Fletcher and Leyffer, 1998) to our new implementation, called FASTr (for filter active-set trust-region solver). Both solvers use BQPD (Fletcher, 2000) to solve the QP subproblems, and we use the number of QPs solved as our performance measure, which is roughly proportional to CPU time. Our implementation of Algorithm 2.1 uses a nonmonotone g- and l-filter with $M = 2$, though we have also experimented with other values of M. FASTr does not use second-order correction steps. A second difference from filterSQP is that FASTr uses the main loop both for feasibility restoration and optimality, making the code shorter and easier to maintain.

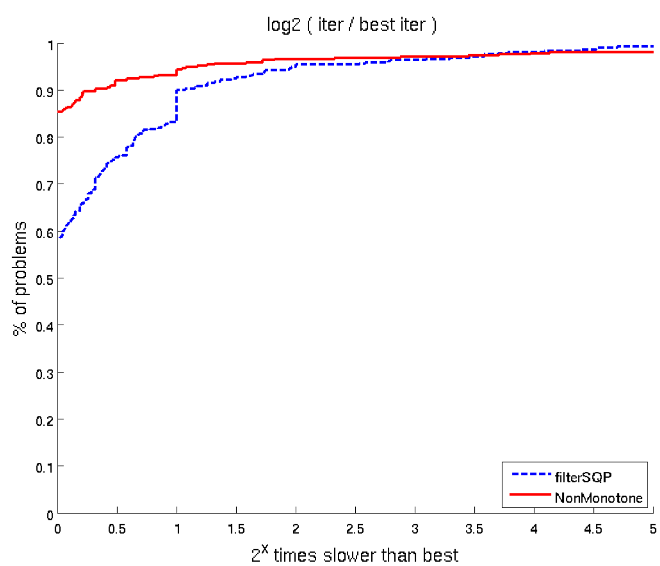

Figure 3: Performance profile comparing the number of QP solves for filterSQP and FASTr.

Figure 3 shows a performance profile (Dolan and Moré, 2002) that compares filterSQP and FASTr. We observe that, in general, the new implementation outperforms filterSQP. We believe that some of this improvement can be attributed to the fact that FASTr does not invoke SOC steps far from the solution. Instead, the nonmonotonicity allows us to accept more steps, even far from the solution, resulting in larger trust-region radii.

6 Conclusion and Discussion

We have presented a nonmonotone filter method for nonlinear optimization and have shown its global and fast local convergence under mild conditions. We introduce two filters: the g-filter and the l-filter. The g-filter guarantees global convergence, while the l-filter is a nonmonotone filter that promotes fast local convergence. The l-filter includes only the full SQP steps, which are important to local convergence analysis. The proposed algorithm improves on the algorithm in Wächter and Biegler (2005a), since it achieves fast local convergence without the use of SOC steps. Moreover, the proposed algorithm uses the objective function in the filter, instead of the Lagrangian function (Ulbrich, 2004), thereby avoiding the potential issue of converging to a saddle point.

Acknowledgments

This work was supported by the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract DE-AC02-06CH11357. This work was also supported by the U.S. Department of Energy through the grant DE-FG02-05ER25694.

References

Audet, C. and Dennis Jr., J. (2004). A pattern search filter method for nonlinear programming without derivatives. *SIAM Journal on Optimization*, 14(4):980–1010.

Benson, H. and Vanderbei, R. (1998). Cute models in AMPL. http://orfe.princeton.edu/ rvdb/ampl/nlmodels/cute/.

Bongartz, I., Conn, A. R., Gould, N. I. M., and Toint, P. L. (1995). CUTE: Constrained and Unconstrained Testing Enviroment. *ACM Transactions on Mathematical Software*, (21):123–160.

Chamberlain, R. M., Powell, M. J. D., Lemarechal, C., and Petersen, H. C. (1982). The watchdog technique for forcing convergence in algorithms for constrained optimization. *Mathematical Programming Study*, 16:1–17.

Chin, C. M. and Fletcher, R. (2003). On the global convergence of an SLP-filter algorithm that takes EQP steps. *Mathematical Programming*, 96(1):161–177.

Conn, A. R., Gould, N. I. M., and Toint, P. L. (2000). *Trust-Region Methods*. MPS-SIAM Series on Optimization. SIAM, Philadelphia.

Dolan, E. D. and Moré, J. (2002). Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213.

Facchinei, F. and Lucidi, S. (1995). Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minimization problems. *Journal of Optimization Theory and Applications*, 85:265–289.

Fletcher, R. (2000). Stable reduced Hessian updates for indefinite quadratic programming. *Mathematical Programming*, 87(2):251–264.

Fletcher, R., Gould, N. I. M., Leyffer, S., Toint, P. L., and Wächter, A. (2002a). Global convergence of trust-region SQP-filter algorithms for general nonlinear programming. *SIAM Journal on Optimization*, 13(3):635–659.

Fletcher, R. and Leyffer, S. (1998). User manual for filterSQP. Numerical Analysis Report NA/181, University of Dundee.

Fletcher, R. and Leyffer, S. (2002). Nonlinear programming without a penalty function. *Mathematical Programming*, 91:239–270.

Fletcher, R., Leyffer, S., and Toint, P. L. (2002b). On the global convergence of a filter-SQP algorithm. *SIAM Journal on Optimization*, 13(1):44–59.

Fourer, R., Gay, D. M., and Kernighan, B. W. (2003). *AMPL: A Modelling Language for Mathematical Programming*. Books/Cole—Thomson Learning, 2nd edition.

Gonzaga, C. C., Karas, E. W., and Vanti, M. (2003). A globally convergent filter method for nonlinear programming. *SIAM Journal on Optimization*, 14(3):646–669.

Gould, N. I. M. and Toint, P. L. (2003). Global convergence of a non-monotone trust-region SQP-filter algorithm for nonlinear programming. Numerical Analysis Report RAL-TR-2003-003, Rutherford Appleton Laboratory, UK. Available online at www.numerical.rl.ac.uk/reports/reports.shtml.

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

Moré, J. J. and Sorensen, D. C. (1983). Computing a trust region step. *SIAM Journal on Scientific & Statistical Computing*, 4:553–572.

Qi, H. D. and Qi, L. Q. (2000). A new QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization. *SIAM Journal on Optimization*, 11(1):113–132.

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

Robinson, S. M. (1980). Strongly regular generalized equations. *Mathematics of Operations Research*, 5:43–62.

Ulbrich, S. (2004). On the superlinear local convergence of a filter-SQP method. *Mathematical Programming*, 100(1):217–245.

Wächter, A. and Biegler, L. (2005a). Line search filter methods for nonlinear programming: Local convergence. *SIAM Journal on Optimization*, 16(1):32–48.

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