A Global Convergence Theory of a Filter Line Search Method for Nonlinear Programming

Choong Ming Chin*

**Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, England, UK.*
E-mail: chin@stats.ox.ac.uk

Numerical Optimization Report, July 2002, revised September 2002

Abstract

A framework for proving global convergence for a class of line search filter type methods for nonlinear programming is presented. The underlying method is based on the dominance concept of multiobjective optimization where trial points are accepted provided there is a sufficient decrease in the objective function or constraints violation function. The proposed methods solve a sequence of quadratic programming subproblems for which effective software is readily available, and instead of using trust region strategies, the methods utilize line search techniques to induce global convergence.

The proof technique is presented in a fairly general context, allowing a range of specific algorithm choices associated with choosing the Hessian matrix representation, controlling the step size and feasibility restoration.

Keywords nonlinear programming, global convergence, line search, filter, multiobjective optimization, SQP.

1 Introduction

This paper concerns with the development of an alternative filter algorithm for finding a local solution of the following Nonlinear Programming (NLP) problem

$$P \left\{ \begin{array}{ll} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & c_i(\mathbf{x}) \le 0 \quad i = 1, 2, \ldots, m, \end{array} \right.$$

where we assume $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $\mathbf{c} : \mathbb{R}^n \mapsto \mathbb{R}^m$ are twice differentiable.

Recently Fletcher and Leyffer [11] have proposed solving Problem P using filter method as an alternative to traditional merit functions approach. The underlying concept is fairly simple where trial points generated from solving a sequence of trust region quadratic programming (QP) subproblems are accepted provided there is a sufficient decrease in the objective function or constraints violation function. In addition the computational results reported in Fletcher and Leyffer [11] are also very encouraging and numerical comparisons with LANCELOT and Sℓ_1QP where both algorithms use merit functions show that the filter strategy is competitive. Soon global convergence for trust region SQP methods have been established in Fletcher, Leyffer and Toint [15], Fletcher, Gould, Leyffer and Toint [10], Gould and Toint [18] and Ulbrich and Ulbrich [23]. Furthermore global convergence for trust region SLP methods have also been reported in Fletcher, Leyffer and Toint [14] and also in Chin and Fletcher [8] where methods have been adapted in the latter to allow the possibility of taking EQP steps. Besides trust region methods, there are also various approaches using the filter strategy such as applications to line search approach [4], interior point approach [3, 24], bundle method for non-smooth optimization [12] and derivative-free approach [1].

In view of the latest development in using the filter strategy in nonlinear programming we also dispense with the idea of using merit functions to induce global convergence in algorithms for NLP. Instead of focussing trust region methods in NLP, we on the other hand propose and analyze an alternative filter strategy framework based on line search method. In the paper by Biegler and Wächer [4], the authors use the filter strategy in conjunction with line search method to promote global and local convergence for NLP. However, their work only focus their analysis on barrier interior point method where the objective function of Problem P is replaced by a barrier objective function. In contrast the proposed algorithm in this paper is in the same vein as that of Fletcher and Leyffer [11] with the exception that line search method is investigated instead of trust region method.

It has been known that line search methods incorporating merit functions can converge to singular non-stationary points if the Jacobian matrix of active linearized constraints are linearly dependent at non-stationary points. Examples of this behaviour are discussed in length in the paper by Byrd, Marazzi and Nocedal [5] and Powell [22]. In the context of a line search method, the filter strategy can circumvent this problem and guarantee sufficient progress to the solution. If the trial steps become too small when utilizing backtracking strategy or when the gradients of the active linearized constraints approach near singularity then the filter algorithm will temporarily exit and enter into feasibility restoration phase. The main objective of entering the feasibility restoration phase is to get closer to the feasible region by minimizing the constraints violation function. Therefore the problems associated with line search–merit function approach will not occur.

The recent success of employing EQP strategy with filter methods (see Chin [6, 7] and Chin and Fletcher [8, 9]) shows that steps calculated on the basis of an equality constrained quadratic programming subproblem enables rapid convergence to take place for problems require second order information. Although the active set of an EQP strategy is determined from solving successively a sequence of trust region linear programming subproblems, numerical evidence (see Chin [6, 7] and Chin and Fletcher [9]) confirms that the EQP steps which are unbounded play a pivotal role in promoting rapid convergence.

Therefore in this paper we would like to extend the filter method further by employing QP steps with the use of a line search method. The main reason for choosing this method is that when solving a trust region subproblem there is an undesirable property of encountering inconsistent constraints due to small trust region radius. On the other hand, less frequently we would encounter inconsistent linearized constraints when solving an IQP type subproblem.

We organize this paper in the following. In Section 2 we begin by introducing the filter method and we will also show how it can be adapted in a line search SQP algorithm. In addition we also discuss the "slanting" filter technique which first featured in Chin [6] and Chin and Fletcher [8] so that stronger statements about convergence to a feasible point can be made. Finally in Section 3 the issue of global convergence for the line search SQP filter method is discussed under mild assumptions.

Before presenting the algorithm and the convergence proof we first make a few definitions. We denote the gradient of f by $\nabla f(\mathbf{x})$, the Jacobian of the constraints by $\nabla \mathbf{c}(\mathbf{x})^T$ and the Lagrangian function by $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{c}(\mathbf{x})$ where $\boldsymbol{\lambda}$ are multiplier estimates corresponding to the nonlinear constraints. The Hessian is denoted by $\mathbf{W}$ in which $\mathbf{W}$ is some approximation of the Hessian of the Lagrangian, $\nabla^2 \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$. Subscript $_k$ refers to iteration indices and quantities relating to local solution of Problem P are superscripted with a $^\infty$.

2 The Filter Line Search Algorithm

We repeat again for the purpose of convenience that in this paper we consider an NLP of the form

$$P \left\{ \begin{array}{ll} \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & c_i(\mathbf{x}) \leq 0 \quad i = 1, 2, \ldots, m, \end{array} \right.$$

and we refer $\mathbf{x}^\infty$ as a local solution of Problem P. In this algorithm to obtain second order convergence of the iterates the most attractive choice is to use Sequential Quadratic Programming (SQP) method as the basic iterative method.

At the current iterate $\mathbf{x}_k$, the QP subproblem in our algorithm is defined by

$$QP(\mathbf{x}_k) \left\{ \begin{array}{ll} \underset{\mathbf{d} \in \mathbb{R}^n}{\text{minimize}} & \nabla f(\mathbf{x}_k)^T \mathbf{d} + \frac{1}{2} \mathbf{d}^T \mathbf{W}_k \mathbf{d} \\ \text{subject to} & \nabla c_i(\mathbf{x}_k)^T \mathbf{d} + c_i(\mathbf{x}_k) \leq 0 \ , \quad i = 1, 2, \ldots, m \end{array} \right.$$

and we denote the solution, the search direction as $\mathbf{d}_k$ (if it exists). After $\mathbf{d}_k$ has been computed, a step size $\alpha \in (0, 1]$ is determined in order to obtain a trial iterate

$$\mathbf{x} \ = \ \mathbf{x}_k + \alpha \mathbf{d}_k.$$

The step size α is chosen via a backtracking strategy so that $\mathbf{x}$ satisfies the filter requirements. If $\mathbf{x}$ satisfies filter conditions we then set $\mathbf{x}_{k+1} = \mathbf{x}$ and $\alpha_k = \alpha$.

We now turn our attention to the definition of an NLP filter. In NLP there are two competing aims to satisfy that is

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad f(\mathbf{x})$$

and

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \quad h(\mathbf{c}(\mathbf{x}))$$

where $h(\mathbf{c}(\mathbf{x})) = \sum_{i=1}^{m} \max\{0, c_i(\mathbf{x})\}$. Using the technique as in trust region methods we also denote

$$\Delta f \;\; = \;\; f(\mathbf{x}_k) - f(\mathbf{x}_k + \alpha \mathbf{d}_k)$$

as the actual reduction in $f(\mathbf{x}_k)$ and

$$\Delta l \;\; = \;\; -\alpha \nabla f(\mathbf{x}_k)^T \mathbf{d}_k$$

as the linear reduction in $f(\mathbf{x}_k)$. Our sufficient reduction condition for $f(\mathbf{x}_k)$ then takes the form

$$\Delta f \;\; \geq \;\; \sigma \Delta l$$

where $\sigma \in (0, 1)$ is a pre-assigned parameter. In some ways the sufficient reduction test resembles the use of Armijo line search condition for unconstrained optimization problems.

In a filter we only consider pairs of values $(h(\mathbf{c}(\mathbf{x})), f(\mathbf{x}))$ obtained by evaluating $h(\mathbf{c}(\mathbf{x}))$ and $f(\mathbf{x})$ for various values of $\mathbf{x}$. Following Fletcher and Leyffer [11] we make the following definitions.

Definition 2.1 A pair $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$ obtained on iteration k is said to dominate another pair $(h(\mathbf{c}(\mathbf{x}_l)), f(\mathbf{x}_l))$ if and only if $h(\mathbf{c}(\mathbf{x}_k)) \leq h(\mathbf{c}(\mathbf{x}_l))$ and $f(\mathbf{x}_k) \leq f(\mathbf{x}_l)$.

The above definition means that $\mathbf{x}_k$ is at least as good as $\mathbf{x}_l$ in respect of both measures. With this concept we can now define a filter as a criterion for accepting or rejecting a trial step.

Definition 2.2 A filter is a list of pairs $(h(\mathbf{c}(\mathbf{x}_i)), f(\mathbf{x}_i))$ such that no pair dominates any other. A point $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$ is said to acceptable for inclusion in the filter if it is not dominated by any point in the filter.

We use $\mathcal{F}^{(k)}$ to denote the set of iterations indices j $(j \leq k)$ such that $(h(\mathbf{c}(\mathbf{x}_j)), f(\mathbf{x}_j)$ is an entry in the current filter. A point $\mathbf{x}$ is said to be "acceptable for inclusion in the filter" if

$$\text{either } h(\mathbf{c}(\mathbf{x})) < h(\mathbf{c}(\mathbf{x}_j)) \text{ or } f(\mathbf{x}) < f(\mathbf{x}_j)$$

for all $j \in \mathcal{F}^{(k)}$. We may also "include a point $\mathbf{x}$ in the filter" which means the pair $(h(\mathbf{c}(\mathbf{x})), f(\mathbf{x}))$ is added to the list of pairs in the filter, and any pairs in the filter that are dominated by $(h(\mathbf{c}(\mathbf{x})), f(\mathbf{x}))$ are removed.

In pictorial form the filter can be represented graphically in the (h, f) plane as illustrated in Figure 1.

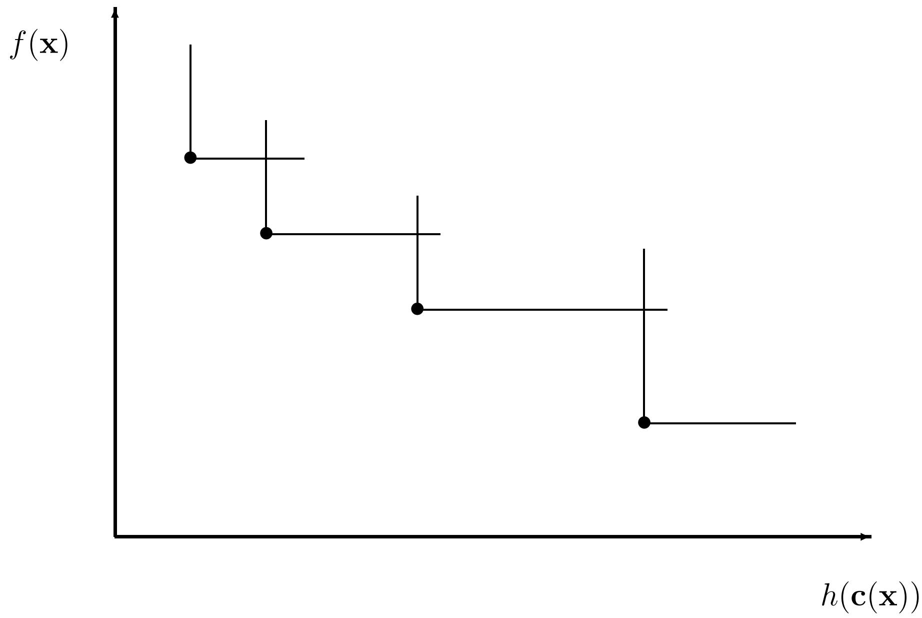

Figure 1: An NLP filter with four pairs of points

Each point in the filter generates a block of unacceptable points and the union of these blocks represents the set of unacceptable points to the filter.

For the purpose of proving convergence, the present definition of a filter is inadequate as it allows points to accumulate in the neighbourhood of filter entries that are not Kuhn-Tucker points. This is readily corrected by defining a small envelope around the current filter entries. In this paper to enable global convergence to be proved, we use a slight modification of the "slanting" envelope described in Chin [6] and Chin and Fletcher [8]. In our test, the trial iterate $\mathbf{x}$ is acceptable to the filter if

$$\text{either } h(\mathbf{c}(\mathbf{x})) \leq (1 - \alpha\eta) h(\mathbf{c}(\mathbf{x}_j)) \text{ or } f(\mathbf{x}) \leq f(\mathbf{x}_j) - \gamma h(\mathbf{c}(\mathbf{x})) \tag{2.1}$$

for all $j \in \mathcal{F}^{(k)}$ where $\eta \in (0, 1)$, $\gamma \in (0, 1)$ are parameters close to zeroes. This idea is illustrated in Figure 2 using values $\alpha = 1$, $\eta = \gamma = 0.1$ although in practice η and γ are close to zeroes.

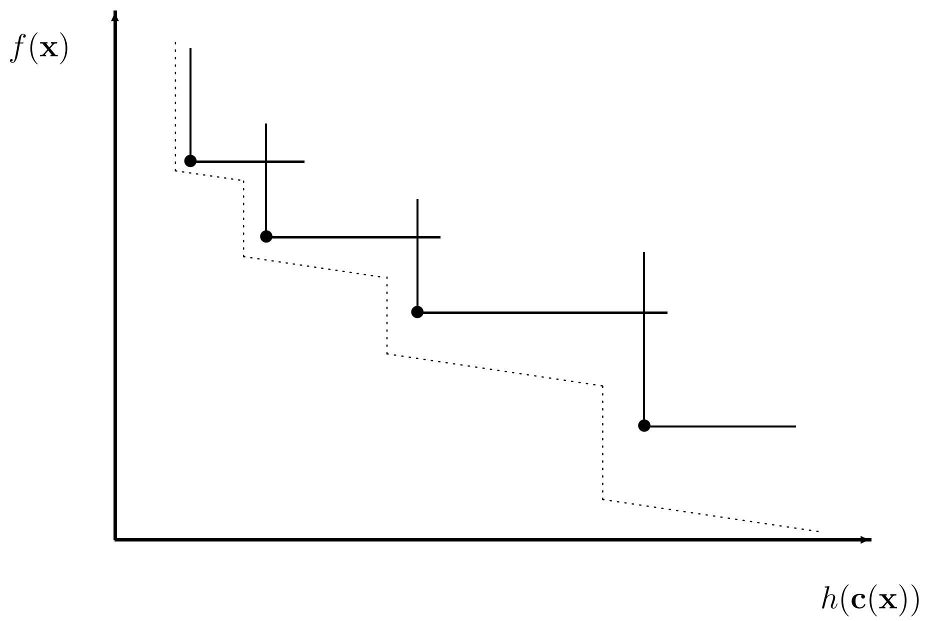

Figure 2: An NLP filter with "slanting" envelope strategy

In addition we also include a necessary condition for accepting a point that is we impose an upper bound

$$h(\mathbf{c}(\mathbf{x})) \leq u$$

($u > 0$) on the constraint infeasibility. We implement this by adding an entry $(u, -\infty)$ in the filter.

We are now in a position to state the filter line search SQP algorithm by means of the following pseudo-code.

Filter Line Search SQP Algorithm

Given initial point $\mathbf{x}_0$, $t \in (0, 1)$, set $k := 0$. If $h(\mathbf{c}(\mathbf{x}_0)) \neq 0$ let $k \in \mathcal{F}^{(k)}$. Set $(u, -\infty)$ in the filter.

Step 1 Solve $QP(\mathbf{x}_k)$ subproblem to obtain $\mathbf{d}_k$. Set $\alpha = 1$ and

$$\alpha_{\min} \begin{cases} = 0 & \text{if } \Delta l > 0, \\ < 1 & \text{otherwise.} \end{cases}$$

Step 2 **If** the $QP(\mathbf{x}_k)$ subproblem is incompatible **Then**

- Goto Feasibility Restoration Phase to find $\mathbf{x}_{k+1}$ so that it is acceptable to the filter and the $QP(\mathbf{x}_{k+1})$ subproblem is compatible.

Else If *convergence criterion* is met **Then**

- STOP.

Endif

Step 3 **If** $\alpha < \alpha_{\min}$ **Then**

- Goto Feasibility Restoration Phase to find $\mathbf{x}_{k+1}$ so that it is acceptable to the filter and the $QP(\mathbf{x}_{k+1})$ subproblem is compatible.

Endif

Step 4 **If** $\mathbf{x}_k + \alpha\mathbf{d}_k$ satisfies the filter test (2.1) and upper bound criteria **Then**

- Goto Step 5.

Else

- Goto Step 6.

Endif

Step 5 **If** $\Delta l > 0$ and $\Delta f < \sigma\Delta l$ **Then**

- Goto Step 6.

Else

- Goto Step 7.

Endif

Step 6 Set $\alpha := \alpha t$ and goto Step 3.

Step 7 Set $\alpha_k = \alpha$, $\Delta f_k = \Delta f$, $\Delta l_k = \Delta l$.

Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{d}_k$.

If $h(\mathbf{c}(\mathbf{x}_{k+1})) > 0$ then set $k+1 \in \mathcal{F}^{(k+1)}$.

Set $k := k+1$.

Goto Step 1.

We begin with an initial guess $\mathbf{x}_0$ of the solution $\mathbf{x}^\infty$ and if $h(\mathbf{c}(\mathbf{x}_k)) > 0$, we would then include $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$ in the filter. At every iteration k there is an inner loop (Step 6) in which backtracking strategy is used, where decreasing values of α are generated. The inner loop terminates when the algorithm satisfies either one of the following scenarios:

(a) the filter acceptability tests are satisfied;

(b) the step size $\alpha < \alpha_{\min}$.

As for proving global convergence, we use the terminology first introduced in Fletcher, Leyffer and Toint [14] and also in Chin and Fletcher [8]. First of all, a step that satisfies $\Delta l > 0$ is referred as an *f-type step* and if $\alpha\mathbf{d}_k$ is accepted by the algorithm to become $\alpha_k\mathbf{d}_k$ then an *f-type iteration* is generated. According to Step 5 of the algorithm, the sufficient reduction test must also be satisfied. Therefore a necessary condition for an f-type iteration to occur is that both

$$\Delta l > 0 \text{ and } \Delta f \geq \sigma \Delta l$$

are satisfied. On the other hand, if $\Delta l \leq 0$ or if the $QP(\mathbf{x}_k)$ subproblem is incompatible then the main purpose of the algorithm is to reduce $h(\mathbf{c}(\mathbf{x}))$ and we refer the resulting iteration as an *h-type iteration*.

Our algorithm also follows the analysis discussed in Chin and Fletcher [8] where all acceptable points with $h(\mathbf{c}(\mathbf{x})) > 0$ are included into the filter inclusive of f-type and h-type iterations. Furthermore we also introduce another terminology to aid our convergence proof that is for a step $\alpha\mathbf{d}_k$ which satisfies $f(\mathbf{x}_k + \alpha\mathbf{d}_k) \leq f(\mathbf{x}_k)$ we denote it as an *f-monotonic step* or an *f-monotonic iteration* if $\alpha\mathbf{d}_k$ is acceptable to the current filter. In addition we denote

$$\tau_k \;\; = \;\; \min_{j \in \mathcal{F}^{(k)}} h(\mathbf{c}(\mathbf{x}_j)) > 0$$

as the minimum of all $h(\mathbf{c}(\mathbf{x}))$ values in the current filter.

Our algorithm also provides an outlet if the current $QP(\mathbf{x}_k)$ subproblem is incompatible or if the backtracking strategy fails to improve either the objective function or the constraints violation function values. We do this by exiting the algorithm temporarily and enter into a feasibility restoration phase where the main purpose is to reduce the constraints infeasibility. The whole process terminates if the restoration phase finds a point that is both acceptable to the filter, and for which the QP subproblem is compatible. In this paper, we do not elaborate how this is done and currently there exists various algorithms to perform this calculation. See Chin [6], Fletcher and Leyffer [11, 13] and Madsen [20].

Note that in the restoration phase, the process of generating iterates that improve the constraints infeasibility could make the resulting objective function $f(\mathbf{x})$ significantly worse than that at the previous point. Hence if the restoration phase does terminate then the point generated would become $\mathbf{x}_{k+1}$ and the resulting step from $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$ is deemed to be an h-type iteration. However, there is always a possibility that the restoration phase might fail to terminate and converge to an infeasible point. An example of this behaviour could happened if there exists a non-zero local minimum of $h(\mathbf{c}(\mathbf{x}))$ which indicates that the original problem P is locally incompatible. On the other hand, if the restoration phase is converging to a feasible point then it is usually likely that the restoration phase will terminate and returns back to the main filter algorithm. This is so because $QP(\mathbf{x})$ is usually compatible if $\mathbf{x}$ is sufficiently close to the feasible region, and also $\tau_k > 0$ allows such a point to be acceptable to the filter. However, this outcome is not guaranteed for any infeasible point $\mathbf{x}$ since it is possible for $QP(\mathbf{x})$ to be incompatible. Thus in this paper we also allow the possibility that the restoration phase may fail to terminate, and regard this scenario as an indication that the constraints of Problem P is locally incompatible.

3 Global Convergence

In this section we present the global convergence proof of the filter line search SQP algorithm. Before going further we make the following assumptions.

Standard Assumptions

(**A1**) Let $\{\mathbf{x}_k\}$ be generated by the line search filter algorithm and suppose that $\{\mathbf{x}_k\}$ and $\{\mathbf{x}_k + \alpha \mathbf{d}_k\}$ are contained in a compact and convex set $\mathcal{S}$ of $\mathbb{R}^n$.

(**A2**) Assume $f(\mathbf{x})$ and $c_i(\mathbf{x})$, $i = 1, 2, \ldots, m$ are twice continuously differentiable on $\mathcal{S}$.

(**A3**) Assume $\mathbf{W}_k$ is bounded for all k.

Remark A consequence of assumption (**A2**) - (**A3**) is that there exists a constant $M > 0$, independent of $\mathbf{x}$ and k such that for all $\mathbf{x} \in \mathcal{S}$ and for all k, it follows that $\frac{1}{2}\mathbf{s}^T \nabla^2 f(\mathbf{x})\mathbf{s} \leq M$, $\frac{1}{2}\mathbf{s}^T \mathbf{W}_k \mathbf{s} \leq M$, $\frac{1}{2}\mathbf{s}^T \nabla^2 c_i(\mathbf{x})\mathbf{s} \leq M$, $i = 1, 2, \ldots, m$ for all vectors $\mathbf{s}$ such that $\|\mathbf{s}\|_\infty = 1$. Without loss of generality, we may also assume that $\|\nabla f(\mathbf{x})\|_2 \leq M$ and $\|\nabla \mathbf{c}(\mathbf{x})\|_1 \leq M$.

Lemma 1 *Let the standard assumptions hold and consider the QP step $\mathbf{d}_k$ from the solution of the $QP(\mathbf{x}_k)$ subproblem. For any $\mathbf{x}_k \in \mathcal{S}$, $\mathbf{d}_k \in \mathbb{R}^n$, the line segment from $\mathbf{x}_k$ to $\mathbf{x}_k + \alpha \mathbf{d}_k$ is contained in $\mathcal{S}$*

$$h(\mathbf{c}(\mathbf{x}_k + \alpha \mathbf{d}_k)) \leq (1-\alpha) h(\mathbf{c}(\mathbf{x}_k)) + \alpha^2 m M \|\mathbf{d}_k\|_\infty^2$$

and

$$|\Delta f - \Delta l| \leq \alpha^2 M \|\mathbf{d}_k\|_\infty^2.$$

Proof From Taylor's theorem for all $i = 1, 2, \ldots, m$

$$c_i(\mathbf{x}_k + \alpha \mathbf{d}_k) = c_i(\mathbf{x}_k) + \alpha \nabla c_i(\mathbf{x}_k)^T \mathbf{d}_k + \frac{1}{2}\alpha^2 (\mathbf{d}_k)^T \nabla^2 c_i(\mathbf{y}_i) \mathbf{d}_k$$

where $\mathbf{y}_i$ is between $\mathbf{x}_k$ to $\mathbf{x}_k + \alpha \mathbf{d}_k$. From the feasibility of $\mathbf{d}_k$ and a consequence of assumption (**A2**)

$$\begin{aligned} c_i(\mathbf{x}_k + \alpha \mathbf{d}_k) &= (1-\alpha) c_i(\mathbf{x}_k) + \alpha c_i(\mathbf{x}_k) + \alpha \nabla c_i(\mathbf{x}_k)^T \mathbf{d}_k + \frac{1}{2}\alpha^2 (\mathbf{d}_k)^T \nabla^2 c_i(\mathbf{y}_i) \mathbf{d}_k \\ &\leq (1-\alpha) c_i(\mathbf{x}_k) + \alpha^2 M \|\mathbf{d}_k\|_\infty^2 \\ &\leq (1-\alpha) h(c_i(\mathbf{x}_k)) + \alpha^2 M \|\mathbf{d}_k\|_\infty^2 \end{aligned}$$

where $h(c_i(\mathbf{x}_k)) = \max\{0, c_i(\mathbf{x}_k)\}$. Thus

$$\begin{aligned} h(\mathbf{c}(\mathbf{x}_k + \alpha \mathbf{d}_k)) &= \sum_{i=1}^{m} h(c_i(\mathbf{x}_k + \alpha \mathbf{d}_k)) \\ &\leq (1-\alpha) h(\mathbf{c}(\mathbf{x}_k)) + \alpha^2 m M \|\mathbf{d}_k\|_\infty^2. \end{aligned}$$

To obtain the second result, from Taylor's theorem there exists a vector $\mathbf{y}$ between the line segment from $\mathbf{x}_k$ to $\mathbf{x}_k + \alpha \mathbf{d}_k$ such that

$$\begin{aligned} |\Delta f - \Delta l| &= |f(\mathbf{x}_k) - f(\mathbf{x}_k + \alpha \mathbf{d}_k) + \alpha \nabla f(\mathbf{x}_k)^T \mathbf{d}_k| \\ &= |f(\mathbf{x}_k) + \alpha \nabla f(\mathbf{x}_k)^T \mathbf{d}_k - f(\mathbf{x}_k) - \alpha \nabla f(\mathbf{x}_k)^T \mathbf{d}_k \\ &\quad - \frac{1}{2}\alpha^2 (\mathbf{d}_k)^T \nabla^2 f(\mathbf{y}) \mathbf{d}_k| \\ &\leq \alpha^2 M \|\mathbf{d}_k\|_\infty^2. \end{aligned}$$

q.e.d

Lemma 2 *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *where* $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ *and* $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ *then*

$$|h(\mathbf{x}) - h(\mathbf{y})| \leq \|\mathbf{x} - \mathbf{y}\|_1$$

where $h(\mathbf{x}) = \sum_{i=1}^{n} \max\{0, x_i\}$ *and* $h(\mathbf{y}) = \sum_{i=1}^{n} \max\{0, y_i\}$.

Proof We let $u_i = \max\{0, x_i\}$ and $v_i = \max\{0, y_i\}$ for $i = 1, 2, \ldots, n$. Hence

$$\begin{aligned} |h(\mathbf{x}) - h(\mathbf{y})| &= \left| \sum_{i=1}^{n} \max\{0, x_i\} - \sum_{i=1}^{n} \max\{0, y_i\} \right| \\ &= \left| \sum_{i=1}^{n} (u_i - v_i) \right| \\ &\leq \sum_{i=1}^{n} |u_i - v_i|. \end{aligned}$$

To prove the result we consider 4 cases

(a) If $x_i > 0$ and $y_i > 0$ then $u_i = \max\{0, x_i\} = x_i$ and $v_i = \max\{0, y_i\} = y_i$. Thus $|u_i - v_i| = |x_i - y_i|$.

(b) If $x_i > 0$ and $y_i \leq 0$ then $u_i = \max\{0, x_i\} = x_i$ and $v_i = \max\{0, y_i\} = 0$. Thus $|u_i - v_i| = |x_i| \leq |x_i - y_i|$.

(c) If $x_i \leq 0$ and $y_i \leq 0$ then $u_i = \max\{0, x_i\} = 0$ and $v_i = \max\{0, y_i\} = 0$. Thus $|u_i - v_i| = 0 \leq |x_i - y_i|$.

(d) If $x_i \leq 0$ and $y_i > 0$ then $u_i = \max\{0, x_i\} = 0$ and $v_i = \max\{0, y_i\} = y_i$. Thus $|u_i - v_i| = |y_i| \leq |x_i - y_i|$.

Therefore from case (a) to case (d)

$$\begin{aligned}|h(\mathbf{x}) - h(\mathbf{y})| &\leq \sum_{i=1}^{n} |u_i - v_i| \\ &\leq \sum_{i=1}^{n} |x_i - y_i| \\ &= \|\mathbf{x} - \mathbf{y}\|_1.\end{aligned}$$

q.e.d

Lemma 3 *Let the standard assumptions hold. For any* $\mathbf{x}_k \in \mathcal{S}$*,* $\mathbf{d}_k \in \mathbb{R}^n$*, the line segment from* $\mathbf{x}_k$ *to* $\mathbf{x}_k + \alpha\mathbf{d}_k$ *is contained in* $\mathcal{S}$

$$|h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) - h(\mathbf{c}(\mathbf{x}_k))| \leq \alpha\delta_1\|\mathbf{d}_k\|_\infty$$

and

$$|f(\mathbf{x}_k + \alpha\mathbf{d}_k) - f(\mathbf{x}_k)| \leq \alpha\delta_2\|\mathbf{d}_k\|_\infty$$

where $\delta_1, \delta_2 > 0$.

Proof From Lemma 2

$$\begin{aligned}|h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) - h(\mathbf{c}(\mathbf{x}_k))| &\leq \|\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k) - \mathbf{c}(\mathbf{x}_k)\|_1 \\ &= \left\|\mathbf{c}(\mathbf{x}_k) + \alpha\int_0^1 \nabla\mathbf{c}(\mathbf{x}_k + z\alpha\mathbf{d}_k)^T\mathbf{d}_k\ dz - \mathbf{c}(\mathbf{x}_k)\right\|_1 \\ &= \left\|\alpha\int_0^1 \nabla\mathbf{c}(\mathbf{x}_k + z\alpha\mathbf{d}_k)^T\mathbf{d}_k\ dz\right\|_1 \\ &\leq \int_0^1 \alpha\|\nabla\mathbf{c}(\mathbf{x}_k + z\alpha\mathbf{d}_k)^T\mathbf{d}_k\|_1\ dz \\ &\leq \int_0^1 \alpha\|\nabla\mathbf{c}(\mathbf{x}_k + z\alpha\mathbf{d}_k)\|_1\|\mathbf{d}_k\|_1\ dz.\end{aligned}$$

By letting $\|\nabla\mathbf{c}(\mathbf{x}_k + z\alpha\mathbf{d}_k)\|_1 \leq M$ where $M > 0$ is independent of $\mathbf{x}$, k and α we therefore have

$$\begin{aligned}|h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) - h(\mathbf{c}(\mathbf{x}_k))| &\leq \alpha M\|\mathbf{d}_k\|_1 \\ &= \alpha\delta_1\|\mathbf{d}_k\|_\infty\end{aligned}$$

where $\delta_1 = nM$.

From Taylor's theorem, there exists a vector $\mathbf{z}$ between $\mathbf{x}_k$ and $\mathbf{x}_k + \alpha\mathbf{d}_k$ such that

$$\begin{aligned}|f(\mathbf{x}_k+\alpha\mathbf{d}_k)-f(\mathbf{x}_k)| &= |f(\mathbf{x}_k)+\alpha\nabla f(\mathbf{z})^T\mathbf{d}_k-f(\mathbf{x}_k)| \\ &= \alpha|\nabla f(\mathbf{z})^T\mathbf{d}_k| \\ &\leq \alpha\|\nabla f(\mathbf{z})\|_2\|\mathbf{d}_k\|_2.\end{aligned}$$

Let $\|\nabla f(\mathbf{z})\|_2 \leq M$ where $M > 0$ and we therefore have

$$|f(\mathbf{x}_k+\alpha\mathbf{d}_k)-f(\mathbf{x}_k)| \leq \alpha\delta_2\|\mathbf{d}_k\|_\infty$$

where $\delta_2 = \sqrt{n}M$.

q.e.d

Our global convergence proof concerns with Karush-Kuhn-Tucker (KKT) necessary conditions under a strict Mangasarian-Fromowitz constraint qualification (SMFCQ) (see Kyparisis [19]). Let $\mathbf{x}^\infty$ be a feasible point of Problem P, then $\mathbf{x}^\infty$ satisfies SMFCQ if and only if

(1) $\nabla c_i(\mathbf{x}^\infty)$, $i \in \mathcal{J}(\mathbf{x}^\infty)$ are linearly independent; and

(2) there exists a vector $\mathbf{s} \in \mathbb{R}^n$ such that

$$\begin{aligned}\nabla c_i(\mathbf{x}^\infty)^T\mathbf{s} &= 0, \quad i \in \mathcal{J}(\mathbf{x}^\infty) \\ \nabla c_i(\mathbf{x}^\infty)^T\mathbf{s} &< 0, \quad i \in \mathcal{J}^\perp(\mathbf{x}^\infty)\end{aligned}$$

where $\mathcal{J}(\mathbf{x}^\infty) = \{i \in \mathcal{E}(\mathbf{x}^\infty) : \lambda_i^\infty > 0\}$, $\mathcal{J}^\perp(\mathbf{x}^\infty) = \{i \in \mathcal{E}(\mathbf{x}^\infty) : \lambda_i^\infty = 0\}$ and $\mathcal{E}(\mathbf{x}^\infty) = \{i : c_i(\mathbf{x}^\infty) = 0\}$

We now turn our attention by stating the KKT necessary conditions. Let $\mathbf{x}^\infty$ solves Problem P and define the sets

$$\begin{aligned}\mathcal{G} &= \left\{\mathbf{s} : \nabla f(\mathbf{x}^\infty)^T\mathbf{s} < 0\right\} \\ \mathcal{C}_{\mathcal{J}} &= \left\{\mathbf{s} : \nabla c_i(\mathbf{x}^\infty)^T\mathbf{s} = 0, i \in \mathcal{J}(\mathbf{x}^\infty)\right\} \\ \mathcal{C}_{\mathcal{J}^\perp} &= \left\{\mathbf{s} : \nabla c_i(\mathbf{x}^\infty)^T\mathbf{s} < 0, i \in \mathcal{J}^\perp(\mathbf{x}^\infty)\right\}.\end{aligned}$$

Thus the KKT necessary conditions for $\mathbf{x}^\infty$ to solve Problem P are

(i) $\mathbf{x}^\infty$ is a feasible point; and

(ii) the set of directions $\mathcal{G} \cap \mathcal{C}_{\mathcal{J}} \cap \mathcal{C}_{\mathcal{J}^\perp} = \emptyset$.

When both conditions (i) - (ii) are true, then we shall refer $\mathbf{x}^\infty$ as a KKT point. Furthermore the consequence of conditions (i) - (ii) implies the existence of unique multipliers (see Kyparisis [19]), and it can be shown (see Gauvin [16]) that the multiplier vector $\boldsymbol{\lambda}$ is bounded. Take note that if the strict complementary slackness conditions holds at $\mathbf{x}^\infty$, i.e. $\mathcal{J}^{\perp}(\mathbf{x}^{\infty}) = \emptyset$, then the SMFCQ is equivalent to the Cottle constraint qualification (see Bazaraa, Sherali and Shetty [2] and Mangasarian [21]).

What is needed for the global convergence proof is that if $\mathbf{x}^{\infty}$ is a feasible point but not a KKT point then there exists an $\varepsilon > 0$ and a vector $\mathbf{s}$ such that $\|\mathbf{s}\|_{\infty} = 1$ for which

$$\nabla f(\mathbf{x}_k)^T\mathbf{s} \leq -\varepsilon \tag{3.1}$$
$$\nabla c_i(\mathbf{x}_k)^T\mathbf{s} = 0 \quad i \in \mathcal{J}(\mathbf{x}^{\infty}) \tag{3.2}$$
$$\nabla c_i(\mathbf{x}_k)^T\mathbf{s} \leq -\varepsilon \quad i \in \mathcal{J}^{\perp}(\mathbf{x}^{\infty}) \tag{3.3}$$

for all $\mathbf{x}_k$ in some neighbourhood $\mathcal{N}^{\infty}$ of $\mathbf{x}^{\infty}$. The results (3.1) – (3.3) are a direct consequence of condition (ii) and continuity of vectors $\nabla f(\mathbf{x})$ and $\nabla c_i(\mathbf{x})$, $i \in \mathcal{E}(\mathbf{x}^{\infty})$. With the results (3.1) - (3.2), we are now in a position to prove the global convergence for our algorithm. First we need to show that the inner iteration must terminate as α is reduced via backtracking strategy.

Theorem 1 *Let the standard assumptions hold, then the inner loop terminates in a finite number of iterations.*

Proof If $\mathbf{x}_k$ is a KKT point of Problem P then $\mathbf{d}_k = \mathbf{0}$ solves the $QP(\mathbf{x}_k)$ subproblem and the algorithm terminates. Otherwise if the inner iteration does not terminate finitely we need to consider two cases depending on whether $QP(\mathbf{x}_k)$ is incompatible or compatible.

case 1 $QP(\mathbf{x}_k)$ subproblem is incompatible
Let the set $\mathcal{V}_k$ be defined as

$$\mathcal{V}_k = \{i : c_i(\mathbf{x}_k) > 0\}.$$

For all $\mathbf{d}_k \in \mathbb{R}^n$ and for a particular $i \in \mathcal{V}_k$, it follows that

$$\begin{aligned} c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d}_k &\geq c_i(\mathbf{x}_k) - \|\nabla c_i(\mathbf{x}_k)\|_2\|\mathbf{d}_k\|_2 \\ &\geq c_i(\mathbf{x}_k) - \sqrt{n}\|\nabla c_i(\mathbf{x}_k)\|_2\|\mathbf{d}_k\|_{\infty} \\ &\geq c_i(\mathbf{x}_k) - \|\nabla c_i(\mathbf{x}_k)\|_1\|\mathbf{d}_k\|_{\infty} \\ &> 0 \end{aligned}$$

if either $\|\nabla c_i(\mathbf{x}_k)\|_1 = 0$ or $\|\mathbf{d}_k\|_{\infty} < c_i(\mathbf{x}_k)/\|\nabla c_i(\mathbf{x}_k)\|_1$. Thus constraint $i \in \mathcal{V}_k$ cannot be satisfied and the $QP(\mathbf{x}_k)$ subproblem is incompatible. Hence the inner iteration terminates finitely for this case.

case 2 $QP(\mathbf{x}_k)$ subproblem is compatible
Suppose the current $QP(\mathbf{x}_k)$ subproblem is feasible and $\mathbf{d}_k$ is the optimal point. To show that the inner loop terminates finitely, we subdivide this case into two smaller subcases depending on whether $\Delta l \leq 0$ or $\Delta l > 0$.

subcase (a) $\Delta l \leq 0$

As $\mathbf{d}_k$ is the optimal solution of the current $QP(\mathbf{x}_k)$ subproblem, in this case we can assume $h(\mathbf{c}(\mathbf{x}_k)) > 0$. From Lemma 1, for some values of $\alpha > 0$

$$h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) \leq (1-\alpha)h(\mathbf{c}(\mathbf{x}_k)) + \alpha^2 mM\|\mathbf{d}_k\|_\infty^2$$

and if

$$\alpha \leq \frac{(1-\eta)h(\mathbf{c}(\mathbf{x}_k))}{mM\|\mathbf{d}_k\|_\infty^2} \tag{3.4}$$

then $h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) \leq (1-\alpha\eta)h(\mathbf{c}(\mathbf{x}_k))$.

In addition, we let the filter space be divided into four orthants relative to $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$. In this case we focus on $(h^{(n)}, f^{(n)})$ and $(h^{(s)}, f^{(s)})$ where $h^{(n)}$ and $h^{(s)}$ are the maximum and minimum $h(\mathbf{c}(\mathbf{x}))$ values in the NW and SE orthants of $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$ respectively. Since $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$ is acceptable to the filter, therefore

$$f(\mathbf{x}_k) \leq f^{(n)} - \gamma h(\mathbf{c}(\mathbf{x}_k))$$

and

$$h(\mathbf{c}(\mathbf{x}_k)) < h^{(s)}.$$

From Lemma 1 and Lemma 3,

$$\begin{aligned} f(\mathbf{x}_k + \alpha\mathbf{d}_k) - f^{(n)} + \gamma h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) &\leq f(\mathbf{x}_k) + \alpha\delta_2\|\mathbf{d}_k\|_\infty + \\ &\quad \gamma[(1-\alpha)h(\mathbf{c}(\mathbf{x}_k)) + \alpha^2 mM\|\mathbf{d}_k\|_\infty^2] \\ &\leq f(\mathbf{x}_k) - f^{(n)} + \gamma h(\mathbf{c}(\mathbf{x}_k)) + \\ &\quad \alpha\delta_2\|\mathbf{d}_k\|_\infty - \alpha\gamma h(\mathbf{c}(\mathbf{x}_k)) + \alpha^2\gamma mM\|\mathbf{d}_k\|_\infty^2 \\ &\leq \alpha\left[\alpha\gamma mM\|\mathbf{d}_k\|_\infty^2 - (\gamma h(\mathbf{c}(\mathbf{x}_k)) - \delta_2\|\mathbf{d}_k\|_\infty)\right] \end{aligned}$$

and if

$$\alpha \leq \frac{\gamma h(\mathbf{c}(\mathbf{x}_k)) - \delta_2\|\mathbf{d}_k\|_\infty}{\gamma mM\|\mathbf{d}_k\|_\infty^2}$$

then $f(\mathbf{x}_k + \alpha\mathbf{d}_k) \leq f^{(n)} - \gamma h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k))$.

Furthermore since

$$h(\mathbf{c}(\mathbf{x}_k)) < h^{(s)}$$

and if (3.4) holds then

$$\begin{aligned} h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) &\leq (1-\alpha\eta)h(\mathbf{c}(\mathbf{x}_k)) \\ &< (1-\alpha\eta)h^{(s)}. \end{aligned}$$

Hence provided

$$\alpha_{\min} \leq \alpha \leq \min\left\{\frac{(1-\eta)h(\mathbf{c}(\mathbf{x}_k))}{mM\|\mathbf{d}_k\|_\infty^2}, \frac{\gamma h(\mathbf{c}(\mathbf{x}_k)) - \delta_2\|\mathbf{d}_k\|_\infty}{\gamma mM\|\mathbf{d}_k\|_\infty^2}\right\}$$

then the trial pair $(h(\mathbf{c}(\mathbf{x}_k+\alpha\mathbf{d}_k)), f(\mathbf{x}_k+\alpha\mathbf{d}_k))$ is acceptable to $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$ and it is also acceptable to the current filter. Therefore the inner iteration also terminates finitely for this case also.

subcase (b) $\Delta l > 0$
Because $\mathbf{x}_k$ is not a KKT point and since $\Delta l > 0$ we can then denote the current optimal QP solution $\mathbf{d}_k = \|\mathbf{d}_k\|_\infty \bar{\mathbf{s}}$ where $\|\bar{\mathbf{s}}\|_\infty = 1$ such that

$$\nabla f(\mathbf{x}_k)^T\bar{\mathbf{s}} \quad \leq \quad -\epsilon$$

where $\epsilon > 0$. Thus

$$\begin{aligned}\Delta l &= -\alpha\nabla f(\mathbf{x}_k)^T\mathbf{d}_k \\ &\geq \alpha\epsilon\|\mathbf{d}_k\|_\infty.\end{aligned}$$

Hence from Lemma 1

$$\begin{aligned}\left|\frac{\Delta f - \Delta l}{\Delta l}\right| &\leq \frac{\alpha^2 M\|\mathbf{d}_k\|_\infty^2}{\alpha\epsilon\|\mathbf{d}_k\|_\infty} \\ &= \frac{\alpha M\|\mathbf{d}_k\|_\infty}{\epsilon}\end{aligned}$$

and if $\alpha \leq \dfrac{(1-\sigma)\epsilon}{M\|\mathbf{d}_k\|_\infty}$ then $\Delta f \geq \sigma\Delta l$.

For the special case of $h(\mathbf{c}(\mathbf{x}_k)) = 0$, from Lemma 1

$$\begin{aligned}h(\mathbf{c}(\mathbf{x}_k+\alpha\mathbf{d}_k)) &\leq \alpha^2 mM\|\mathbf{d}_k\|_\infty^2 \\ &\leq \alpha mM\|\mathbf{d}_k\|_\infty^2\end{aligned}$$

and we focus on $\tau_k = \min\limits_{j\in\mathcal{F}^{(k)}} h(\mathbf{c}(\mathbf{x}_j))$. If

$$\alpha \quad \leq \quad \frac{\tau_k}{\eta\tau_k + mM\|\mathbf{d}_k\|_\infty^2}$$

then $h(\mathbf{c}(\mathbf{x}_k+\alpha\mathbf{d}_k)) \leq (1-\alpha\eta)\tau_k$. It follows that for sufficiently small α in the range

$$\alpha \quad \leq \quad \min\left\{\frac{(1-\sigma)\epsilon}{M\|\mathbf{d}_k\|_\infty}, \frac{\tau_k}{\eta\tau_k + mM\|\mathbf{d}_k\|_\infty^2}\right\}$$

then $\mathbf{x}_k + \alpha\mathbf{d}_k$ is acceptable to the filter and the trial step also satisfies the sufficient reduction test. Therefore the inner loop terminates finitely in this special case.

On the other hand if $h(\mathbf{c}(\mathbf{x}_k)) > 0$ then from Lemma 1, if

$$\alpha \leq \frac{(1-\eta)h(\mathbf{c}(\mathbf{x}_k))}{mM\|\mathbf{d}_k\|_\infty^2}$$

then $h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) \leq (1-\alpha\eta)h(\mathbf{c}(\mathbf{x}_k))$. It follows that for sufficiently small α in the range

$$\alpha \leq \min\left\{\frac{(1-\sigma)\epsilon}{M\|\mathbf{d}_k\|_\infty}, \frac{(1-\eta)h(\mathbf{c}(\mathbf{x}_k))}{mM\|\mathbf{d}_k\|_\infty^2}\right\} \tag{3.5}$$

then the trial pair $(h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)), f(\mathbf{x}_k + \alpha\mathbf{d}_k))$ is acceptable to $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$ and it also satisfies the sufficient reduction test.

To show that $(h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)), f(\mathbf{x}_k + \alpha\mathbf{d}_k))$ is acceptable to the current filter, we let the filter space be divided into four orthants relative to $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$. Using similar techniques as described in subcase (a), we concentrate on the filter pairs $(h^{(n)}, f^{(n)})$ and $(h^{(s)}, f^{(s)})$ where $h^{(n)}$ and $h^{(s)}$ are the maximum and minimum $h(\mathbf{c}(\mathbf{x}))$ values in the NW and SE orthants respectively of $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$. Since $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$ is acceptable to the filter we then have

$$f(\mathbf{x}_k) \leq f^{(n)} - \gamma h(\mathbf{c}(\mathbf{x}_k))$$

and

$$h(\mathbf{c}(\mathbf{x}_k)) < h^{(s)}$$

If (3.5) holds then

$$\begin{aligned} f(\mathbf{x}_k + \alpha\mathbf{d}_k) - f^{(n)} + \gamma h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) &\leq f(\mathbf{x}_k) - \sigma\Delta l - f^{(n)} + \gamma(1-\alpha\eta)h(\mathbf{c}(\mathbf{x}_k)) \\ &\leq f(\mathbf{x}_k) - f^{(n)} + \gamma h(\mathbf{c}(\mathbf{x}_k)) - \\ &\quad \alpha\sigma\epsilon\|\mathbf{d}_k\|_\infty - \alpha\eta\gamma h(\mathbf{c}(\mathbf{x}_k)) \\ &\leq 0 \end{aligned}$$

and

$$\begin{aligned} h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) &\leq (1-\alpha\eta)h(\mathbf{c}(\mathbf{x}_k)) \\ &< (1-\alpha\eta)h^{(s)}. \end{aligned}$$

Thus if (3.5) holds then $(h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)), f(\mathbf{x}_k + \alpha\mathbf{d}_k))$ is also acceptable to all the filter entries in the current filter. Therefore the iteration terminates finitely in this case as well.

q.e.d

In the next theorem we will show that the filter line search algorithm does provide a global convergence proof for the class of problems that we are going to solve. Using similar strategy as in Chin and Fletcher (2001) that when the algorithm is applied one of four different possible outcomes ($\mathbb{A}$ - $\mathbb{D}$) can occur.

($\mathbb{A}$) The feasibility restoration phase converges infinitely and fails to find a point $\mathbf{x}$ which is acceptable by the filter and for which the $QP(\mathbf{x})$ subproblem is compatible.

($\mathbb{B}$) A KKT point is found.

($\mathbb{C}$) There exists an infinite subsequence of consecutive f-monotonic iterations for k sufficiently large.

($\mathbb{D}$) There exists an infinite subsequence of iterations that are not f-monotonic.

A consequence of Theorem 1 is that if the filter line search algorithm does not terminate in either type ($\mathbb{A}$) or ($\mathbb{B}$) then an infinite subsequence of type ($\mathbb{C}$) or ($\mathbb{D}$) will occur. The next theorem completes the global convergence proof by showing that subsequences from type ($\mathbb{C}$) or ($\mathbb{D}$) converge to a KKT point.

Theorem 2 *Let the standard assumptions hold, then for the filter line search SQP algorithm either iterates of type* ($\mathbb{A}$) *or* ($\mathbb{B}$) *will occur, or there exists an infinite subsequence of type* ($\mathbb{C}$) *or* ($\mathbb{D}$)*. If we denote* $\mathbf{x}^\infty$ *as an accumulation point for this subsequence then* $\mathbf{x}^\infty$ *satisfies Karush-Kuhn-Tucker conditions.*

Proof We only need to consider the case when neither iterates of type ($\mathbb{A}$) nor type ($\mathbb{B}$) occurs. The existence of an infinite subsequence of type ($\mathbb{C}$) or ($\mathbb{D}$) follows from Theorem 1. Let $\mathbf{x}^\infty$ be an accumulation point of this subsequence and $\{\mathbf{x}_k\}_{k\in\mathcal{K}}$ be any thinner subsequences converging to $\mathbf{x}^\infty$. We first show that $\mathbf{x}^\infty$ is a feasible point.

Assume that $h(\mathbf{c}(\mathbf{x}_k)) \to h(\mathbf{c}(\mathbf{x}^\infty)) > 0$ for $k \in \mathcal{K}$. Let i and j be any two adjacent indices in $\mathcal{K}$ where $i < j$. If $h(\mathbf{c}(\mathbf{x}^\infty)) > 0$, then there exists $k^{'} \in \mathcal{K}$ such that for all $i \geq k^{'}$ and because $\mathbf{x}_j$ is acceptable to the filter, we have

$$h(\mathbf{c}(\mathbf{x}_j) \quad > \quad h(\mathbf{c}(\mathbf{x}_i))$$

and

$$f(\mathbf{x}_j) \quad \leq \quad f(\mathbf{x}_i) - \gamma h(\mathbf{c}(\mathbf{x}_j)).$$

Since $\{f(\mathbf{x}_k)\}_{k\in\mathcal{K}}$ is a monotonically decreasing subsequence for $k \geq k^{'}$ and is bounded below therefore for $i, j \in \mathcal{K}$, $i, j \geq k^{'}$ and $i < j$,

$$\sum_{i,j\in\mathcal{K}} \Delta f_{i,j} = \sum_{i,j\in\mathcal{K}} \{f(\mathbf{x}_i) - f(\mathbf{x}_j)\}$$

is bounded above. However, since $f(\mathbf{x}_j) \leq f(\mathbf{x}_i) - \gamma h(\mathbf{c}(\mathbf{x}_j))$, therefore by summing over all indices $i, j \in \mathcal{K}$, $i, j \geq k^{'}$ and $i < j$,

$$\sum_{i,j \in \mathcal{K}} \Delta f_{i,j} \geq \gamma \sum_{j \in \mathcal{K}} h(\mathbf{c}(\mathbf{x}_j)) \rightarrow +\infty$$

which contradicts the fact $\sum_{i,j \in \mathcal{K}} \Delta f_{i,j}$ is bounded above. Thus $h(\mathbf{c}(\mathbf{x}^{\infty})) = 0$ and hence $\mathbf{x}^{\infty}$ is a feasible point.

Next we need to show $\mathbf{x}^{\infty}$ is a KKT point and we prove the theorem by contradiction. Suppose $\mathbf{x}^{\infty}$ is not a KKT point. Then for $k \in \mathcal{K}$, k sufficiently large, $\mathbf{x}_k$ is in some neighbourhood $\mathcal{N}^{\infty}$ of $\mathbf{x}^{\infty}$, then there exists a vector $\mathbf{s}^{\circ}$ such that $\|\mathbf{s}^{\circ}\|_{\infty} = 1$ for which

$$\begin{array}{rcl} \nabla f(\mathbf{x}_k)^T \mathbf{s}^{\circ} & < & 0 \\ \nabla c_i(\mathbf{x}_k)^T \mathbf{s}^{\circ} & = & 0, \quad i \in \mathcal{J}(\mathbf{x}^{\infty}) \\ \nabla c_i(\mathbf{x}_k)^T \mathbf{s}^{\circ} & < & 0, \quad i \in \mathcal{J}^{\perp}(\mathbf{x}^{\infty}). \end{array}$$

Using similar techniques as described in Fletcher, Leyffer and Toint [15], let the matrix $\mathbf{A}_k$ consists of column vectors $\nabla c_i(\mathbf{x}_k)$, $i \in \mathcal{J}(\mathbf{x}^{\infty})$. By linear independence and continuity there exists a neighbourhood $\mathbf{x}_k \in \mathcal{N}^{\infty}$ in which $\mathbf{A}_k$ has full column rank. Hence for the linear system

$$\mathbf{A}_k^T \mathbf{d} + \mathbf{b}_k = \mathbf{0}$$

where $\mathbf{cb}_k$ is the corresponding column vector, we can write the range space vector $\mathbf{p}$ as

$$\mathbf{p} = -(\mathbf{A}_k^T)^{+} \mathbf{b}_k$$

such that $(\mathbf{A}_k^T)^{+}$ is the pseudo-inverse of $\mathbf{A}_k^T$.

In addition since $\mathbf{A}_k$ has full column rank, the null space of $\mathbf{A}_k^T$ defines the tangent space to the active inequality constraints, $i \in \mathcal{J}(\mathbf{x}^{\infty})$ at $\mathbf{x}_k$. We can then write the projection onto this tangent space as

$$\mathbf{P}_k = \mathbf{I} - \mathbf{A}_k \left[\mathbf{A}_k^T \mathbf{A}_k\right]^{-1} \mathbf{A}_k^T$$

and we let

$$\mathbf{s} = \frac{\mathbf{P}_k \mathbf{s}^{\circ}}{\|\mathbf{P}_k \mathbf{s}^{\circ}\|_2}$$

which is the closest unit vector to $\mathbf{s}^{\circ}$ in the null space of $\mathbf{A}_k^T$. Hence by continuity there exists a (smaller) neighbourhood $\mathcal{N}^{\infty}$ and a constant $\varepsilon > 0$ such that

$$\begin{array}{rcl} \nabla f(\mathbf{x}_k)^T \mathbf{s} & \leq & -\varepsilon \\ \nabla c_i(\mathbf{x}_k)^T \mathbf{s} & = & 0, \quad i \in \mathcal{J}(\mathbf{x}^{\infty}) \\ \nabla c_i(\mathbf{x}_k)^T \mathbf{s} & \leq & -\varepsilon, \quad i \in \mathcal{J}^{\perp}(\mathbf{x}^{\infty}) \end{array}$$

for any $\mathbf{x}_k \in \mathcal{N}^\infty$. By definition the range space vector $\mathbf{p}$ satisfies

$$\begin{aligned}\|\mathbf{p}\|_2 &\leq \|(\mathbf{A}_k^T)^+\|_2\|\mathbf{b}_k\|_2\\ &\leq M\|\mathbf{c}_k\|_2\end{aligned}$$

where $M > 0$ and $\|\mathbf{c}_k\|_2 = \sqrt{\sum_{i\in\mathcal{E}(\mathbf{x}^\infty)}\{c_i(\mathbf{x}_k)\}^2}$. Furthermore if we denote $\mathbf{d}_k$ as the optimal point for the current $QP(\mathbf{x}_k)$ subproblem and if

$$\|\mathbf{d}_k\|_\infty \geq M\|\mathbf{c}_k\|_2 \tag{3.6}$$

then $\|\mathbf{p}\|_\infty \leq \|\mathbf{p}\|_2 \leq \|\mathbf{d}_k\|_\infty$.

We now consider the solution of $QP(\mathbf{x}_k)$ subproblem and if (3.6) holds, we let the line segment be

$$\mathbf{d} = \mathbf{p} + (\|\mathbf{d}_k\|_\infty - \|\mathbf{p}\|_2)\mathbf{s}$$

where $\|\mathbf{d}_k\|_\infty \geq \|\mathbf{p}\|_2$. Since the vectors $\mathbf{p}$ and $\mathbf{s}$ are orthogonal, it follows that $\|\mathbf{d}\|_\infty \leq \|\mathbf{d}\|_2 \leq \|\mathbf{d}_k\|_\infty$.

For active constraints at $\mathbf{x}^\infty$, $i \in \mathcal{J}(\mathbf{x}^\infty)$

$$\begin{aligned}c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d} &= c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{p} + (\|\mathbf{d}_k\|_\infty - \|\mathbf{p}\|_2)\nabla c_i(\mathbf{x}_k)^T\mathbf{s}\\ &= 0.\end{aligned}$$

On the other hand for active constraints at $\mathbf{x}^\infty$, $i \in \mathcal{J}^\perp(\mathbf{x}^\infty)$

$$\begin{aligned}c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d} &= c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{p} + (\|\mathbf{d}_k\|_\infty - \|\mathbf{p}\|_2)\nabla c_i(\mathbf{x}_k)^T\mathbf{s}\\ &\leq \|\mathbf{c}_k\|_2 + M\|\mathbf{p}\|_2 - (\|\mathbf{d}_k\|_\infty - \|\mathbf{p}\|_2)\varepsilon\\ &\leq [1 + M(\varepsilon + M)]\,\|\mathbf{c}_k\|_2 - \varepsilon\|\mathbf{d}_k\|_\infty \end{aligned}\tag{3.7}$$

and if

$$\|\mathbf{d}_k\|_\infty \geq \frac{[1 + M(\varepsilon + M)]\,\|\mathbf{c}_k\|_2}{\varepsilon}$$

then the set of constraints (3.7) are consistent. For inactive constraints $i \notin \mathcal{E}(\mathbf{x}^\infty)$ and for k sufficiently large, $k \in \mathcal{K}$, $c_i(\mathbf{x}_k) \leq -c$ and $\nabla c_i(\mathbf{x}_k)^T\mathbf{d} \leq a\|\mathbf{d}_k\|_\infty$ where $a > 0$ and $c > 0$ are independent of k. It follows that

$$c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d} \leq -c + a\|\mathbf{d}_k\|_\infty \tag{3.8}$$

and if $\|\mathbf{d}_k\|_\infty \leq c/a$ then the set of constraints (3.8) are consistent. Hence it follows that for k sufficiently large and if

$$\max\left\{M\|\mathbf{c}_k\|_2, \frac{[1 + M(\varepsilon + M)]\,\|\mathbf{c}_k\|_2}{\varepsilon}\right\} \leq \|\mathbf{d}_k\|_\infty \leq \frac{c}{a}$$

then $\mathbf{d} = \mathbf{p} + (\|\mathbf{d}_k\|_\infty - \|\mathbf{p}_2)\mathbf{s}$ is a feasible point and the current $QP(\mathbf{x}_k)$ subproblem is compatible.

To show that $\mathbf{d}$ is a descent direction, we can write

$$\begin{aligned}\nabla f(\mathbf{x}_k)^T\mathbf{d} &= \nabla f(\mathbf{x}_k)^T\mathbf{p} + (\|\mathbf{d}_k\|_\infty - \|\mathbf{p}\|_2)\nabla f(\mathbf{x}_k)^T\mathbf{s} \\ &\leq M\|\mathbf{p}\|_2 + \varepsilon\|\mathbf{p}\|_2 - \varepsilon\|\mathbf{d}_k\|_\infty \\ &\leq M^2\|\mathbf{c}_k\|_2 + \varepsilon M\|\mathbf{c}_k\|_2 - \varepsilon\|\mathbf{d}_k\|_\infty\end{aligned}$$

and if

$$\|\mathbf{d}_k\|_\infty \geq \frac{2M(M+\varepsilon)\|\mathbf{c}_k\|_2}{\varepsilon}$$

then $\nabla f(\mathbf{x}_k)^T\mathbf{d} \leq -\varepsilon/2$. Hence if we let

$$\max\left\{M\|\mathbf{c}_k\|_2, \frac{[1+M(\varepsilon+M)]\,\|\mathbf{c}_k\|_2}{\varepsilon}, \frac{2M(M+\varepsilon)\|\mathbf{c}_k\|_2}{\varepsilon}\right\} = \beta\|\mathbf{c}_k\|_2$$

where $\beta > 0$ and if

$$\beta\|\mathbf{c}_k\|_2 \leq \|\mathbf{d}_k\|_\infty \leq c/a \tag{3.9}$$

then $\mathbf{d} = \mathbf{p} + (\|\mathbf{d}_k\|_\infty - \|\mathbf{p}\|_2)\mathbf{s}$ is both a feasible and a descent direction in the current $QP(\mathbf{x}_k)$ subproblem. For the purpose of showing $\mathbf{x}^\infty$ is a KKT point, as $\mathbf{x}_k \to \mathbf{x}^\infty$ for $k \in \mathcal{K}$ we can assume that the subsequence $\{\mathbf{d}_k\}_{k\in\mathcal{K}}$ converges to $\mathbf{d}^\infty$ where $\|\mathbf{d}^\infty\|_\infty = \bar{d} > 0$. This is due to the fact that $\mathbf{d}^\infty = 0$ if and only if $\mathbf{x}^\infty$ is a feasible point satisfying KKT conditions.

By denoting $\mathbf{d}_k$ as the minimizer of the current $QP(\mathbf{x}_k)$ subproblem, it follows by optimality of $\mathbf{d}_k$

$$\nabla f(\mathbf{x}_k)^T\mathbf{d}_k + \frac{1}{2}\mathbf{d}_k^T\mathbf{W}_k\mathbf{d}_k \;\leq\; \nabla f(\mathbf{x}_k)^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T\mathbf{W}_k\mathbf{d}.$$

If (3.9) holds such that $\nabla f(\mathbf{x}_k)^T\mathbf{d} \leq -\varepsilon/2$ and using the boundedness assumption, we then have

$$\begin{aligned}\nabla f(\mathbf{x}_k)^T\mathbf{d}_k &\leq \nabla f(\mathbf{x}_k)^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T\mathbf{W}_k\mathbf{d} - \frac{1}{2}\mathbf{d}_k^T\mathbf{W}_k\mathbf{d}_k \\ &\leq -\frac{\varepsilon}{2}\|\mathbf{d}_k\|_\infty + 2M\|\mathbf{d}_k\|_\infty^2\end{aligned}$$

and if $\|\mathbf{d}_k\|_\infty \leq \varepsilon/8M$ then $\nabla f(\mathbf{x}_k)^T\mathbf{d}_k \leq -\frac{\varepsilon}{4}\|\mathbf{d}_k\|_\infty$. We then have

$$\begin{aligned}\Delta l &= -\alpha\nabla f(\mathbf{x}_k)^T\mathbf{d}_k \\ &\geq \frac{1}{4}\alpha\varepsilon\|\mathbf{d}_k\|_\infty.\end{aligned}$$

Thus if

$$\beta\|\mathbf{c}_k\|_2 \le \|\mathbf{d}_k\|_\infty \le \min\left\{\frac{\varepsilon}{8M}, \frac{c}{a}\right\} \tag{3.10}$$

then the $QP(\mathbf{x}_k)$ subproblem is consistent and $\nabla f(\mathbf{x}_k)^T\mathbf{d}_k \le -\frac{\varepsilon}{4}\|\mathbf{d}_k\|_\infty$.

Hence from Lemma 1

$$\begin{aligned}\left|\frac{\Delta f - \Delta l}{\Delta l}\right| &\le \frac{\alpha^2 M\|\mathbf{d}_k\|_\infty^2}{\alpha\frac{\varepsilon}{4}\|\mathbf{d}_k\|_\infty} \\ &= \frac{4\alpha M\|\mathbf{d}_k\|_\infty}{\varepsilon}\end{aligned}$$

and if

$$\alpha \le \frac{(1-\sigma)\varepsilon}{4M\|\mathbf{d}_k\|_\infty} \tag{3.11}$$

then $\Delta f \ge \sigma\Delta l$.

We now consider the“slanting” envelope test and if (3.11) holds true then from Lemma 3 we have

$$\begin{aligned} f(\mathbf{x}_k) - f(\mathbf{x}_k+\alpha\mathbf{d}_k) - \gamma h(\mathbf{c}(\mathbf{x}_k+\alpha\mathbf{d}_k)) &= \Delta f - \gamma h(\mathbf{c}(\mathbf{x}_k+\alpha\mathbf{d}_k)) \\ &\ge \sigma\Delta l - \gamma h(\mathbf{c}(\mathbf{x}_k)) - \alpha\gamma\delta_1\|\mathbf{d}_k\|_\infty \\ &= \alpha\left(\frac{\sigma\varepsilon}{4}\|\mathbf{d}_k\|_\infty - \gamma\delta_1\|\mathbf{d}_k\|_\infty\right) - \gamma h(\mathbf{c}(\mathbf{x}_k)).\end{aligned}$$

So if

$$\alpha \ge \frac{4\gamma h(\mathbf{c}(\mathbf{x}_k))}{(\sigma\varepsilon - 4\gamma\delta_1)\|\mathbf{d}_k\|_\infty} \tag{3.12}$$

then $f(\mathbf{x}_k+\alpha\mathbf{d}_k) \le f(\mathbf{x}_k) - \gamma h(\mathbf{c}(\mathbf{x}_k+\alpha\mathbf{d}_k))$ and it follows that $(h(\mathbf{c}(\mathbf{x}_k+\alpha\mathbf{d}_k)), f(\mathbf{x}_k+\alpha\mathbf{d}_k))$ is acceptable to the filter.

We are now in the position to examine subsequences of type ($\mathbb{C}$) and ($\mathbb{D}$) iterations. First of all, let us consider the case of a subsequence arising from type ($\mathbb{C}$) iterations where there exists an iteration $\bar{k}$ so that for all $k \ge \bar{k}$, $k \in \mathcal{K}$ all iterations are f-monotonic. We let $\mathbf{x}^\infty$ be any accumulation point of the main sequence and assume that $\mathbf{x}^\infty$ is not a KKT point. Without loss of generality, we assume that $\mathbf{x}_k \in \mathcal{N}^\infty$ and $k \ge \bar{k}$ for all $k \in \mathcal{K}$.

We shall choose $\bar{k}$ such that $h(\mathbf{c}(\mathbf{x}_{\bar{k}})) > 0$ and we focus on the filter entry $(h(\mathbf{c}(\mathbf{x}_{\bar{k}}))$, $f(\mathbf{x}_{\bar{k}}))$ and regard the filter space as being divided into four orthants relative to the $(h(\mathbf{c}(\mathbf{x}_{\bar{k}}))$, $f(\mathbf{x}_{\bar{k}}))$ entry. On iteration $\bar{k}$, there are no filter entries in the SW orthant or the NE orthant. Let us consider what happens on some iterations when $k > \bar{k}$.

Because all iterations j such that $\bar{k} \le j < k$ are f-monotonic, their function values are monotonically decreasing. Assume (3.10) is true so that the current $QP(\mathbf{x}_k)$ subproblem is compatible. Hence using inequalities (3.11) and (3.12), we see that if

$$\frac{4\gamma h(\mathbf{c}(\mathbf{x}_k))}{(\sigma\varepsilon - 4\gamma\delta_1)\|\mathbf{d}_k\|_\infty} \le \alpha \le \frac{(1-\sigma)\varepsilon}{24M\|\mathbf{d}_k\|_\infty} \tag{3.13}$$

then $f(\mathbf{x}_k + \alpha\mathbf{d}_k) \le f(\mathbf{x}_k) - \gamma h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k))$, $\Delta l > 0$ and $\Delta f \ge \sigma\Delta l$. Thus the trial pair $(h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)), f(\mathbf{x}_k + \alpha\mathbf{d}_k))$ is acceptable to $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$.

Since $\{f(\mathbf{x}_k)\}_{k\ge\bar{k},k\in\mathcal{K}}$ is monotonically decreasing, therefore

$$f(\mathbf{x}_k) \le f(\mathbf{x}_j)$$

for all $\bar{k} \le j < k$. If (3.13) holds, therefore

$$\begin{aligned} f(\mathbf{x}_k + \alpha\mathbf{d}_k) &\le f(\mathbf{x}_k) - \gamma h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) \\ &\le f(\mathbf{x}_j) - \gamma h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) \end{aligned}$$

for all $\bar{k} \le j < k$. Therefore $(h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)), f(\mathbf{x}_k + \alpha\mathbf{d}_k))$ is also acceptable to all entries in $\mathcal{F}^{(k)}$ with indices $\bar{k} \le j < k$.

For $i < \bar{k}$ in the NW orthant of $(h(\mathbf{c}(\mathbf{x}_{\bar{k}})), f(\mathbf{x}_{\bar{k}}))$, we have

$$f(\mathbf{x}_{\bar{k}}) \le f(\mathbf{x}_i).$$

Thus if (3.11) is true then

$$\begin{aligned} f(\mathbf{x}_k + \alpha\mathbf{d}_k) &\le f(\mathbf{x}_{\bar{k}}) - \gamma h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) \\ &\le f(\mathbf{x}_i) - \gamma h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) \end{aligned}$$

for all $i < \bar{k}$ in the NW orthant of $(h(\mathbf{c}(\mathbf{x}_{\bar{k}})), f(\mathbf{x}_{\bar{k}}))$. Thus $(h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)), f(\mathbf{x}_k + \alpha\mathbf{d}_k))$ is also acceptable to all entries in the NW orthant of $(h(\mathbf{c}(\mathbf{x}_{\bar{k}})), f(\mathbf{x}_{\bar{k}}))$ with indices $i < \bar{k}$, $i \in \mathcal{F}^{(k)}$.

In order for the trial pair $(h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)), f(\mathbf{x}_k + \alpha\mathbf{d}_k))$ to be acceptable to all the filter entries in $\mathcal{F}^{(k)}$, we now focus our attention on the current filter entry $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$. In this case let the filter space be divided into four orthants relative to $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$ and we shall focus on the SE orthant of $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$. We can then find

$$\tau_k^{(+)} = \min_{j<k} \quad h(\mathbf{c}(\mathbf{x}_j))$$

for all filter entries which lie in the SE orthant of $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$ such that

$$h(\mathbf{c}(\mathbf{x}_k)) < \tau_k^{(+)}.$$

Since for $k \geq \bar{k}$, $k \in \mathcal{K}$, all iterations are f-monotonic therefore

$$h(\mathbf{c}(\mathbf{x}_{\bar{k}})) < \tau_k^{(+)}.$$

In a special case if all iterations have $h(\mathbf{c}(\mathbf{x}_k)) = 0$ then we can set $\bar{k} = 0$ and let $\tau_k^{(+)} = u$, the upper bound criteria for $h(\mathbf{c}(\mathbf{x}_k))$. Hence we can deduce that $\tau_k^{(+)}$ lies in the interval

$$h(\mathbf{c}(\mathbf{x}_{\bar{k}})) < \tau_k^{(+)} \leq u.$$

From Lemma 1,

$$\begin{aligned} h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) &\leq (1-\alpha)h(\mathbf{c}(\mathbf{x}_k)) + \alpha^2 mM\|\mathbf{d}_k\|_\infty^2 \\ &< (1-\alpha)\tau_k^{(+)} + \alpha^2 mM\|\mathbf{d}_k\|_\infty^2 \end{aligned}$$

and if

$$\alpha \leq \frac{(1-\eta)\tau_k^{(+)}}{mM\|\mathbf{d}_k\|_\infty^2} \tag{3.14}$$

then $h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) < (1-\alpha\eta)\tau_k^{(+)}$. Hence if (3.10) is true so that the $QP(\mathbf{x}_k)$ subproblem is compatible then by combining (3.13) and (3.14), we see that if

$$\frac{4\gamma h(\mathbf{c}(\mathbf{x}_k))}{(\sigma\varepsilon - 4\gamma\delta_1)\|\mathbf{d}_k\|_\infty} \leq \alpha \leq \min\left\{\frac{(1-\eta)\tau_k^{(+)}}{mM\|\mathbf{d}_k\|_\infty^2}, \frac{(1-\sigma)\varepsilon}{4M\|\mathbf{d}_k\|_\infty}\right\} \tag{3.15}$$

then $(h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)), f(\mathbf{x}_k + \alpha\mathbf{d}_k))$ is acceptable to all the filter entries in $\mathcal{F}^{(k)}$. Thus as $h(\mathbf{c}(\mathbf{x}_k)) \to 0$ for $k \in \mathcal{K}$, $\|\mathbf{d}_k\|_\infty \to \bar{d}$ the left-hand-sides of (3.10) and (3.15) go to zero while the right-hand-side of (3.15) approaches to a constant number, say $\bar{\alpha}$ where $\bar{\alpha} > 0$.

Therefore by initializing $\alpha = 1$ at the start of the backtracking strategy and if $\bar{\alpha} > 1$ then $\alpha = 1$ will be accepted by the algorithm. Otherwise, as α is reduced in the inner iteration, eventually it must fall within this interval for it to be acceptable by the algorithm. Hence for k sufficiently large, $k \in \mathcal{K}$ we can guarantee

$$\alpha \geq \bar{\alpha}t$$

to be chosen so that an f-type iteration will occur. Hence

$$\begin{aligned} \Delta l_k &\geq \frac{1}{4}\alpha_k\varepsilon\|\mathbf{d}_k\|_\infty \\ &\geq \frac{1}{4}\bar{\alpha}t\varepsilon\bar{d} \end{aligned}$$

and because $\Delta f_k \geq \sigma\Delta l_k$, Δf_k is thus uniformly bounded away from zero which contradicts the fact $\sum_{k\in\mathcal{K}} \Delta f_k$ is convergent. Thus in this case of type ($\mathbb{C}$) iterations, any limit point is a KKT point.

We now come to the stage of examining type ($\mathbb{D}$) iterations and we consider the case such that there exists an infinite subsequence of iterations that are not f-monotonic. We denote $\mathcal{K}$ as any subsequence of iterations that are not f-monotonic and for which $\mathbf{x}_k \to \mathbf{x}^\infty$ and $\mathbf{x}_k \in \mathcal{N}^\infty$ for k sufficiently large, $k \in \mathcal{K}$.

Using the result of Lemma 1, if

$$\alpha \ \leq \ \frac{(1-\eta)h(\mathbf{c}(\mathbf{x}_k))}{mM\|\mathbf{d}_k\|_\infty^2} \tag{3.16}$$

then $h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)) \leq (1-\alpha\eta)h(\mathbf{c}(\mathbf{x}_k))$. Furthermore, assuming (3.10) holds so that the $QP(\mathbf{x}_k)$ subproblem is compatible and by combining (3.11) and (3.16), we see that if

$$\alpha \leq \min\left\{\frac{(1-\eta)h(\mathbf{c}(\mathbf{x}_k))}{mM\|\mathbf{d}_k\|_\infty^2},\ \frac{(1-\sigma)\varepsilon}{4M\|\mathbf{d}_k\|_\infty}\right\} \tag{3.17}$$

then $(h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)), f(\mathbf{x}_k + \alpha\mathbf{d}_k))$ is acceptable to $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$ and it satisfies the sufficient reduction test. In addition, using similar arguments as in Theorem 1, if (3.17) holds then $(h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)), f(\mathbf{x}_k + \alpha\mathbf{d}_k))$ is also acceptable to all filter entries in the NW and SE orthants of $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$. Hence if (3.17) holds the trial pair $(h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)), f(\mathbf{x}_k + \alpha\mathbf{d}_k))$ satisfies the conditions for an f-type step and also the condition for an f-monotonic step.

As $h(\mathbf{c}(\mathbf{x}_k)) \to 0$ and $\|\mathbf{d}_k\|_\infty \to \bar{d}$ for k sufficiently large, $k \in \mathcal{K}$, we can then locate a value for α within the interval (3.17) via backtracking strategy so that $(h(\mathbf{c}(\mathbf{x}_k + \alpha\mathbf{d}_k)), f(\mathbf{x}_k + \alpha\mathbf{d}_k))$ is acceptable to all the filter entries in $\mathcal{F}^{(k)}$ and also satisfies the condition for an f-monotonic step. It is not possible for a large value of α to produce a step such that $\Delta f < 0$ since $\Delta l \geq \frac{1}{4}\alpha\varepsilon\|\mathbf{d}_k\|_\infty$ increases monotonically as α increases. Thus if $k \in \mathcal{K}$ is sufficiently large, an f-monotonic step will be generated. But this contradicts the fact the case being considered is formed by a subsequence of steps that are not f-monotonic.

Hence there are no descent directions at $\mathbf{x}^\infty$ for both type ($\mathbb{C}$) and type ($\mathbb{D}$) iterations and it follows that $\mathbf{x}^\infty$ is a KKT point.

q.e.d

4 Conclusion

A prototypical algorithm of applying filter strategy in line search SQP methods has been described and global convergence has been shown, demonstrating the fact convergence for NLP can be achieved without the need to maintain sufficient descent in a traditional penalty type merit function approach. Of course the algorithm is incomplete in many areas and can only be served as a guide to what might be successfully implemented in practice. For instance, there is a need to specify a suitable algorithm in the restoration phase that can guarantee global convergence if the generated iterates do not return back to the main filter algorithm. The issue of utilizing efficient backtracking strategy in controlling the step size α and the choice of a suitable Hessian matrix $\mathbf{W}_k$ also need to be looked into. One possibility is to use the Hessian of the Lagrangian calculated from second derivatives of f and $\mathbf{c}$, and also using estimates of Lagrange multiplier. The disadvantage of such a procedure is that the matrix $\mathbf{W}_k = \nabla^2\mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k)$ could be indefinite and hence the task of finding the global minimizer of $QP(\mathbf{x}_k)$ could be problematic. Another possibility is to use quasi-Newton methods to update $\mathbf{W}_k$ at every iteration. This alternative strategy could ensure the positive-definiteness of $\mathbf{W}_k$ be maintained so that any KKT point of the QP subproblem is a global solution. In a future paper we hope to present some numerical evidence concerning a quasi-Newton approach of the filter line search SQP algorithm.

Finally, the question of providing a local convergence proof remains an open one. Although local convergence results have been established by Biegler and Wächer [4], the analysis is done for the barrier approach only. The main difficulty in our approach is to ensure Maratos effect does not occur near the limit so that second order convergence property of the SQP iteration is not compromised. The algorithm also permits the use of second order correction (SOC) steps where in traditional line search methods, SOC steps are usually employed to circumvent the Maratos effect and hence improve the convergence properties of the iterates toward the solution. However, in using the filter strategy, it is by no means certain that acceptability to the filter is compatible with accepting the SOC step with step size one in the limit. All these questions are still the subject of ongoing research.

References

[1] C. Audet and J.E. Dennis Jr, *A pattern search filter method for nonlinear programming without derivatives*, Technical Report, Department of Computational and Applied Mathematics, Rice University, Houston, TX, (2000).

[2] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, Nonlinear Programming Theory and Algorithms, 2nd Edition, John Wiley and Sons, New York, (1993).

[3] H.Y. Benson, D.F. Shanno and R.J. Vanderbei, *Interior-point methods for nonconvex nonlinear programming: filter methods and merit functions*, Technical Report ORFE-00-06, Operations Research and Financial Engineering, Princeton University, (2001).

[4] L.T. Biegler and A. Wächter, *Global and local convergence of line search filter methods for nonlinear programming*, CAPD Technical Report B-01-09, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA, (2001).

[5] R.H. Byrd, M. Marazzi and J. Nocedal, *On the convergence of Newton iterations to non-stationary points*, Report OTC 2001/01, Optimization Technology Center, Northwestern University, Evanston, IL, (2001).

[6] C.M. Chin, A New Trust Region based SLP Filter Algorithm which uses EQP Active Set Strategy, PhD Thesis, Department of Mathematics, University of Dundee, Scotland, (2001).

[7] C.M. Chin, *Numerical results of* SLPSQP, filterSQP *and* LANCELOT *on selected CUTE test problems*, Numerical Analysis Report NA/203, Deparment of Mathematics, University of Dundee, Scotland, (2001).

[8] C.M. Chin and R. Fletcher, *On the global convergence of an SLP–filter algorithm that takes EQP steps*, Numerical Analysis Report NA/199, Department of Mathematics, University of Dundee, Scotland, (2001), to appear in Mathematical Programming.

[9] C.M. Chin and R. Fletcher, *Numerical performance of an SLP–filter algorithm that takes EQP steps*, Numerical Analysis Report NA/202, Department of Mathematics, University of Dundee, Scotland, (2001).

[10] R. Fletcher, N.I.M. Gould, S. Leyffer and Ph.L. Toint, *Global convergence of trust-region SQP-filter algorithms for nonlinear programming*, Technical Report 99/03, Department of Mathematics, University of Namur, Belgium. (2001).

[11] R. Fletcher and S. Leyffer, *Nonlinear programming without a penalty function*, Technical Report NA/171, Department of Mathematics, University of Dundee, Scotland, to appear in Mathematical Programming, (1997).

[12] R. Fletcher and S. Leyffer, *A bundle filter method for nonsmooth nonlinear optimization*, Numerical Analysis Report NA/195, Department of Mathematics, University of Dundee, Scotland, (1999).

[13] R. Fletcher and S. Leyffer, *Filter type algorithms for solving systems of algebraic equations and inequalities*, Numerical Analysis Report NA/204, Department of Mathematics, University of Dundee, Scotland, (2001).

[14] R. Fletcher, S. Leyffer and Ph.L. Toint, *On the global convergence of an SLP-filter algorithm*, Numerical Analysis Report NA/183, Department of Mathematics, University of Dundee, Scotland, (1998).

[15] R. Fletcher, S. Leyffer and Ph.L. Toint, *On the global convergence of a filter-SQP algorithm*, Numerical Analysis Report NA/197, Department of Mathematics, University of Dundee, (2000).

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

[17] C.C. Gonzaga, E. Karas and M. Vanti, *A globally convergent filter method for nonlinear programming*, Technical Report, Department of Mathematics, Federal University of Santa Catarina, Brazil, (2001).

[18] N.I.M. Gould and Ph.L. Toint, *Global convergence of a hybrid trust–region SQP–filter algorithm for general nonlinear programming*, Technical Report 01/07, Department of Mathematics, University of Namur, Belgium, (2001).

[19] J. Kyparisis, *On uniqueness of Kuhn-Tucker multipliers in nonlinear programming*, Mathematical Programming, **32**, (1985), pp. 242 – 246.

[20] K. Madsen, *An algorithm for minimax solution of overdetermined systems of nonlinear equations*, Journal of the Institute of Mathematics and its Applications, **16**(3), (1975), pp. 321 – 328.

[21] O.L. Mangasarian, Nonlinear Programming, McGraw–Hill Book Company, New York, (1969).

[22] M.J.D. Powell, *A hybrid method for nonlinear equations*, In P. Rabinowitz (editor), Numerical Methods for Nonlinear Algebraic Equations, Gordon and Breach, London, (1970), pp. 87 – 114.

[23] M. Ulbrich and S. Ulbrich, *Non-monotone trust region methods for nonlinear equality constrained optimization without a penalty function*, to appear in Mathematical Programming, (2002).

[24] M. Ulbrich, S. Ulbrich and L.N. Vicente, *A globally convergent primal-dual interior-point filter method for nonlinear programming*, Technical Report TR00-12, Deparment of Computational and Applied Mathematics, Rice University, Houston, TX, (2002).