A Local 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, United Kingdom.*
E-mail: `chin@stats.ox.ac.uk`

Numerical Optimization Report, December 2002, revised January 2003

Abstract

In this paper the theory of local convergence for a class of line search filter type methods for nonlinear programming is presented. The algorithm presented here is globally convergent (see Chin [4]) and the rate of convergence is two-step superlinear. The proposed algorithm solves a sequence of quadratic progrmming subproblems to obtain search directions and instead of using penalty functions to determine the required step size, a filter technique is used to induce convergence. In addition to avoid the Maratos effect, the algorithm also employs second order correction (SOC) steps so that fast local convergence to the solution can be achieved.

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

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

1 Introduction

This paper concerns with developing a local convergence theory for a class of line search filter method solving the following Nonlinear Programming (NLP) problem

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

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

The recent paper by Chin [4] shows that trial points generated from solving a sequence of quadratic programming (QP) subproblems together with a concept of a filter and a line search technique to determine the required step sizes is globally convergent.

Thus in this paper we further extend the filter method so that local convergence can be proved. It has been mentioned by Fletcher and Leyffer [9] that the filter technique like ℓ_1 penalty function methods could suffer from Maratos effect when iterates are near to local solutions. To date many strategies have been suggested to overcome this problem and chiefly among them are the *watchdog technique* by Chamberlain, Lemarechal, Pedersen and Powell [2], the *non-monotone line search scheme* by Grippo, Lampariello and Lucidi [15] and the *second order correction steps* technique (see Coleman and Conn [7], Gabay [14], Fletcher [8] and, Mayne and Polak [19]). In essence, the first two techniques try to preserve the global as well as local convergence properties by allowing some iterations to violate the standard line search criterion based on the ℓ_1 penalty function while the aim of second order correction steps is to "bend" the search directions closer to the feasible region. In order to avoid truncation of the step size of one when near a solution, the most popular choice in NLP is to employ second order correction (SOC) steps to maintain fast local convergence properties. In fact the usage of SOC steps in filter methods are not entirely new as Chin [3], Chin and Fletcher [6], and Fletcher and Leyffer [9] have used it in the context of trust region SQP filter methods.

In the paper by Biegler and Wächer [1], the authors have established local convergence for a class of line search filter methods using second order correction steps and the analysis is done for the barrier approach only. Recently Ulbrich [24] has provided a superlinear local convergence proof for a modified version of the trust region filter SQP algorithm without the usage of second order correction steps. However the approach taken is different in spirit as compared with the original filter method [9] where Ulbrich [24] uses a Lagrangian function instead of the objective function in the filter. The same argument can also be applied to the work of Biegler and Wächer [1] in which they used a barrier function in replace of the objective function. In contrast the proposed algorithm in this paper is in the same vein as that of Fletcher and Leyffer [9] with the exception that line search method is utilized instead of trust region method. In order to established local convergence of the iterates, our algorithm would also employ SOC steps to circumvent the Maratos effect together with the concept of a filter. Apart from preserving fast local convergence, another advantage of incorporating SOC steps in our filter line search SQP algorithm is that the global convergence proof presented in the paper by Chin [4] will not be compromised. Therefore the global convergence results will still hold true even after extending the filter algorithm to accept SOC steps.

The organization of this paper is as follows. In Section 2 we will introduce the filter method and show how SOC steps are incorporated in a filter line search SQP algorithm. In addition we also discuss the "slanting" filter technique which first featured in Chin [3], Chin and Fletcher [5], and Fletcher, Leyffer and Toint [12] so that global convergence to a Karush-Kuhn-Tucker (KKT) point can be preserved. Finally in Section 3, the local convergence properties of the extended filter line search SQP algorithm will be discussed. In this section also we will show by using the filter strategy, acceptability to the filter is compatible with accepting the full SOC step with step size one in the limit.

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 $^\infty$. In addition we will denote $\|\cdot\|$ to represent the Euclidean norm for any vector or matrix representation. Finally we denote $\|\mathbf{d}_k\| = O(\|\mathbf{u}_k\|)$ for a sequence of $\{\|\mathbf{d}_k\|\}$ satisfying $\|\mathbf{d}_k\| \leq M\|\mathbf{u}_k\|$ for some constant $M > 0$, independent of $\mathbf{x}$ and k, and $\|\mathbf{d}_k\| = o(\|\mathbf{v}_k\|)$ for a sequence of $\{\|\mathbf{d}_k\|\}$ satisfying $\lim_{k\to\infty} \|\mathbf{d}_k\|/\|\mathbf{v}_k\| = 0$.

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

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

and we refer $\mathbf{x}^\infty$ as a local solution of Problem P. In this algorithm to obtain rapid 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

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

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

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

If $\mathbf{x}$ satisfies filter conditions we then set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha \mathbf{d}_k$ and $\alpha_k = \alpha$.

On the other hand if the QP step is rejected by the filter and to overcome the difficulties associated with the Maratos effect, we can then construct a second order correction step. By adapting of an idea used by Herskovits [16] and also by Panier and Tits [20, 21] in the context of feasible SQP methods we first calculate a step $\widetilde{\mathbf{d}}_k$ (if it exists) from solving the following modified QP subproblem

$$\widetilde{QP}(\mathbf{x}_k) \begin{cases} \underset{\widetilde{\mathbf{d}}\in\mathbb{R}^n}{\text{minimize}} & \nabla f(\mathbf{x}_k)^T (\widetilde{\mathbf{d}} + \mathbf{d}_k) + \frac{1}{2}(\widetilde{\mathbf{d}} + \mathbf{d}_k)^T \mathbf{W}_k (\widetilde{\mathbf{d}} + \mathbf{d}_k) \\ \text{subject to} & \nabla c_i(\mathbf{x}_k)^T \widetilde{\mathbf{d}} + c_i(\mathbf{x}_k + \mathbf{d}_k) = -\|\mathbf{d}_k\|^\nu \ , \quad i \in \widetilde{\mathcal{A}}(\mathbf{x}_k) \\ & \nabla c_j(\mathbf{x}_k)^T \widetilde{\mathbf{d}} + c_j(\mathbf{x}_k + \mathbf{d}_k) \leq -\|\mathbf{d}_k\|^\nu \ , \quad j \in \mathcal{A}(\mathbf{x}_k) - \widetilde{\mathcal{A}}(\mathbf{x}_k) \end{cases}$$

where $\nu \in (2,3)$, $\widetilde{\mathcal{A}}(\mathbf{x}_k) = \{i \in \mathcal{A}(\mathbf{x}_k) \ : \ \lambda_k^{(i)} > 0\}$ and $\mathcal{A}(\mathbf{x}_k) = \{i \ : \ \nabla c_i(\mathbf{x}_k)^T \mathbf{d}_k + c_i(\mathbf{x}_k) = 0\}$. Using the inherited step size value $\alpha \in (0,1]$ from the previous trial iterate $\mathbf{x}_k + \alpha \mathbf{d}_k$, we then form a new trial iterate using an *SOC* step, $\mathbf{d}_k + \widetilde{\mathbf{d}}_k$ where

$$\mathbf{x} \ = \ \mathbf{x}_k + \alpha \mathbf{d}_k + \alpha^2 \widetilde{\mathbf{d}}_k.$$

and then subject $\mathbf{x}$ to the required filter test. If the new trial iterate $\mathbf{x}$ satisfies all the filter conditions we then set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha \mathbf{d}_k + \alpha^2 \widetilde{\mathbf{d}}_k$ and $\alpha_k = \alpha$. However, if the new trial iterate fails to be accepted by the filter we then reduce the step size α via a backtracking strategy and return to test the QP/SOC steps again until either $\mathbf{x}_k + \alpha \mathbf{d}_k$ or $\mathbf{x}_k + \alpha \mathbf{d}_k + \alpha^2 \widetilde{\mathbf{d}}_k$ satisfies the filter requirements or temporarily exits from the main filter algorithm to the feasibility restoration phase.

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

$$\Delta f \ = \ \begin{cases} f(\mathbf{x}_k) - f(\mathbf{x}_k + \alpha \mathbf{d}_k) & \text{if } QP \text{ step is used} \\ f(\mathbf{x}_k) - f(\mathbf{x}_k + \alpha \mathbf{d}_k + \alpha^2 \widetilde{\mathbf{d}}_k) & \text{if } SOC \text{ step is used.} \end{cases}$$

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 $\sigma \in (0, \frac{1}{2})$ 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 this paper we propose $\sigma \in (0, \frac{1}{2})$ instead of $\sigma \in (0,1)$ as in the global convergence theorem given in Chin [4] and the main reason for choosing the former is that this parameter range plays an important role in constructing the local convergence proof. Take note that even with this new reduced parameter range, the global convergence proof [4] still holds true.

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 [9] 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.

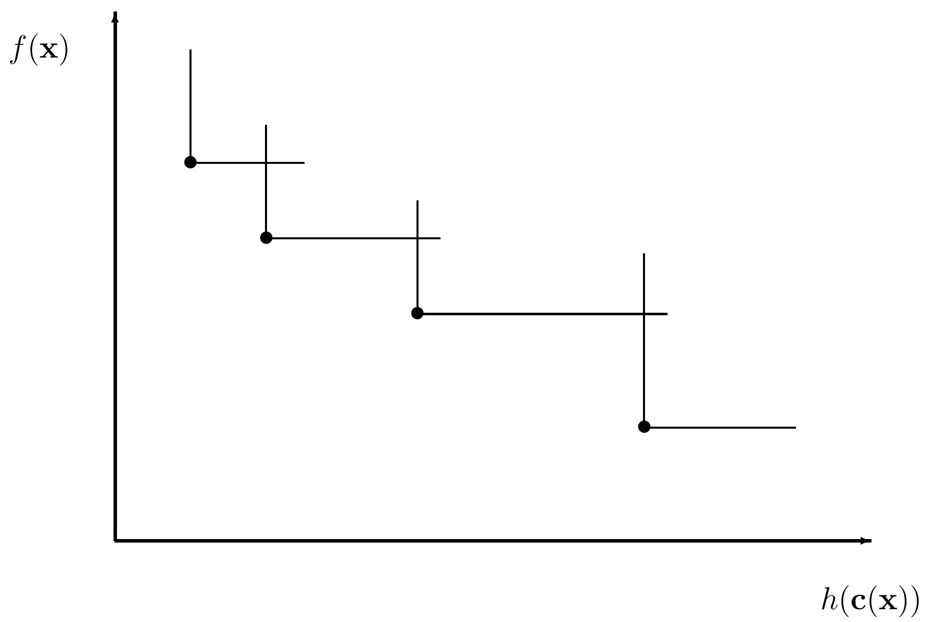

Figure 1: An NLP filter with four pairs of points

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 Karush-Kuhn-Tucker (KKT) points. This is readily corrected by defining a small envelope around the current filter entries. To enable global convergence to be proved (see Chin [4]), we use a slight modification of the "slanting" envelope described in Chin [3] and Chin and Fletcher [5]. 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.

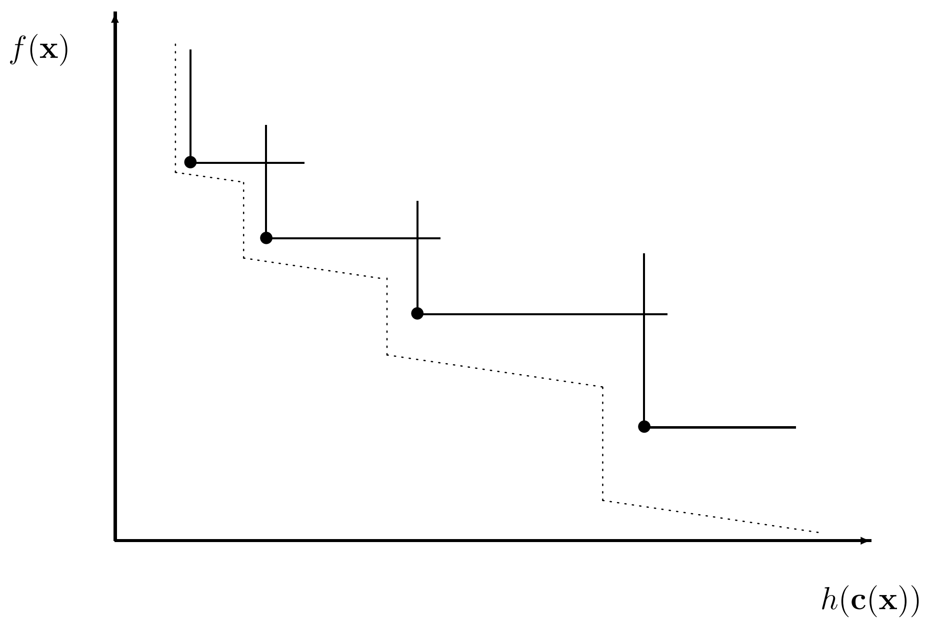

Figure 2: An NLP filter with "slanting" envelope strategy

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.

With the inclusion of calculating second order correction steps, we are now in a position to state an extended version of 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 $\sigma \in (0, \frac{1}{2})$, $\gamma \in (0,1)$ and $k := 0$. 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**

- **if** QP step is used **then**

  - Goto Step 6.

  else if SOC step is used **then**

  - Goto Step 10.

- **endif**

Else

- Goto Step 11.

Endif

Step 6 **If** $\alpha = 1$ **Then**

- Goto Step 7.

Else If $\widetilde{\mathbf{d}}_k$ exists **Then**

- Goto Step 9.

Else

- Goto Step 10.

Endif

Step 7 Solve $\widetilde{QP}(\mathbf{x}_k)$ subproblem to obtain a correction step $\widetilde{\mathbf{d}}_k$

Step 8 **If** the $\widetilde{QP}(\mathbf{x}_k)$ subproblem is incompatible **Then**

- Goto Step 10.

Else

- Goto Step 9.

End If

Step 9 **If** $\mathbf{x}_k + \alpha\mathbf{d}_k + \alpha^2\widetilde{\mathbf{d}}_k$ satisfies the filter test (2.1) and upper bound criteria **Then**

- Goto Step 5.

Else

- Goto Step 10.

End If

Step 10 Set $\alpha := \alpha t$ and goto Step 3.

Step 11 Set $\alpha_k = \alpha$, $\Delta f_k = \Delta f$, $\Delta l_k = \Delta l$.

$$\text{Set } \mathbf{x}_{k+1} = \begin{cases} \mathbf{x}_k + \alpha_k\mathbf{d}_k & \text{if } QP \text{ step is used} \\ \mathbf{x}_k + \alpha_k\mathbf{d}_k + \alpha_k^2\widetilde{\mathbf{d}}_k & \text{if } SOC \text{ step is used.} \end{cases}$$

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.

The algorithm presented is fashioned to some extent on that given in Chin [4] and the most important additional feature is that it allows SOC steps to be tested if the QP steps are rejected by the filter for each choice of $\alpha \in (0, 1]$. Beginning with an initial guess $\mathbf{x}_0$ of the solution $\mathbf{x}^\infty$, we would include $(h(\mathbf{c}(\mathbf{x}_0)), f(\mathbf{x}_0))$ in the filter if $h(\mathbf{c}(\mathbf{x}_0)) > 0$ and at every iteration k there is an inner loop (Step 10) in which backtracking strategy is used. When the inner loop terminates the algorithm will satisfy either one of the following scenarios:

(a) a new iterate is found where the current values of α, Δf and Δl are denoted by α_k, Δf_k and Δl_k respectively;

(b) the step size $\alpha < \alpha_{\min}$ and the algorithm exits to the feasibility restoration phase.

As for proving global convergence, we use the terminology first introduced in Fletcher, Leyffer and Toint [11] and also in Chin and Fletcher [5]. First of all, a step that satisfies $\Delta l > 0$ is referred as an *f-type step* and if $\alpha \mathbf{d}_k$ or $\alpha \mathbf{d}_k + \alpha^2 \widetilde{\mathbf{d}}_k$ is accepted by the algorithm to become $\alpha_k \mathbf{d}_k$ or $\alpha_k \mathbf{d}_k + \alpha_k^2 \mathbf{d}_k$ respectively, 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. Otherwise we deem the step to be an *h-type step* and if the h-type step is accepted then the resulting iteration is known as an *h-type iteration* which also includes points generated from the restoration phase.

Our algorithm also follows the analysis discussed in Chin and Fletcher [5] where all acceptable points with $h(\mathbf{c}(\mathbf{x})) > 0$ are included into the filter inclusive of f-type and h-type iterations. 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. In addition the 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 [3], Fletcher and Leyffer [9, 10] and Madsen [18].

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 Local Convergence

In this section we present the local convergence proof of the filter line search SQP algorithm. In order to be consistent with the convergence to a KKT point under Strict Mangasarian - Fromowitz constraints qualification (SMFCQ) (see Kyparisis [17]) used in the global convergence proof [4], here we consider a local minimizer $\mathbf{x}^\infty$ that satisfies the second order sufficiency conditions without strict complementary slackness and we make the following assumptions.

Standard Assumptions

(**A1**) Let $\{\mathbf{x}_k\}$ be generated by the line search filter algorithm and suppose that $QP(\mathbf{x}_k)$ and $\widetilde{QP}(\mathbf{x}_k)$ are always feasible for all $\mathbf{x}_k$ in a neighbourhood $\mathcal{N}^\infty$ of $\mathbf{x}^\infty$. In addition, suppose that the feasibility restoration phase is not invoked for all $\mathbf{x}_k \in \mathcal{N}^\infty$. Assume further that $\{\mathbf{x}_k\}$, $\{\mathbf{x}_k + \alpha\mathbf{d}_k\}$ and $\{\mathbf{x}_k + \alpha\mathbf{d}_k + \alpha^2\widetilde{\mathbf{d}}_k\}$ 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 three times continuously differentiable in a neighbourhood of $\mathbf{x}^\infty$.

(**A3**) For all $\mathbf{x}$ in a neighbourhood of $\mathbf{x}^\infty$, the vectors $\{\nabla c_i(\mathbf{x}), i \in \mathcal{J}(\mathbf{x}^\infty)\}$ are linearly independent where $\mathcal{J}(\mathbf{x}^\infty) = \{i \in \mathcal{E}(\mathbf{x}^\infty) \ : \ \lambda_i^\infty > 0\}$ and $\mathcal{E}(\mathbf{x}^\infty) = \{i \ : \ c_i(\mathbf{x}^\infty) = 0\}$. This implies a weakened form of complementarity slackness, that is $\lambda_i^\infty c_i(\mathbf{x}^\infty) = 0$, $\lambda_i^\infty > 0$ for $i \in \mathcal{J}(\mathbf{x}^\infty)$ and $\lambda_j^\infty c_j(\mathbf{x}^\infty) = 0$, $\lambda_j^\infty = 0$ for $j \in \mathcal{J}^\perp(\mathbf{x}^\infty) = \mathcal{E}(\mathbf{x}^\infty) - \mathcal{J}(\mathbf{x}^\infty)$.

(**A4**) Assume $\mathbf{W}_k$ is bounded for all k.

(**A5**) The sequence $\{\mathbf{W}_k\}$ converges to a matrix $\mathbf{W}^\infty$ satisfying

$$\mathbf{P}^\infty \mathbf{W}^\infty \mathbf{P}^\infty = \mathbf{P}^\infty \nabla^2 \mathcal{L}(\mathbf{x}^\infty, \boldsymbol{\lambda}^\infty) \mathbf{P}^\infty$$

where $\mathbf{P}^\infty = \mathbf{I} - \mathbf{A}^\infty \left[(\mathbf{A}^\infty)^T \mathbf{A}^\infty\right]^{-1} (\mathbf{A}^\infty)^T$ and $\mathbf{A}^\infty = \{\nabla c_i(\mathbf{x}^\infty), i \in \mathcal{J}(\mathbf{x}^\infty)\}$.

(**A6**) For k large enough, $\mathbf{x}_k$ in the neighbourhood of $\mathbf{x}^\infty$, the matrices $\mathbf{W}_k$ are uniformly positive-definite on the null space of $\mathbf{A}(\mathbf{x}_k)^T = \left\{\nabla c_i(\mathbf{x}_k)^T, i \in \mathcal{J}(\mathbf{x}^\infty)\right\}$, that is there exists a $\rho > 0$ such that

$$\mathbf{d}^T \mathbf{P}_k \mathbf{W}_k \mathbf{P} \mathbf{d} \geq \rho \|\mathbf{P}_k \mathbf{d}\|^2$$

for all $\mathbf{d} \neq \mathbf{0}$ with $\nabla c_i(\mathbf{x}_k)^T \mathbf{P}_k \mathbf{d} = 0$, $i \in \mathcal{J}(\mathbf{x}^\infty)$ and $\nabla c_j(\mathbf{x}_k)^T \mathbf{P}_k \mathbf{d} \leq 0$, $j \in \mathcal{J}^\perp(\mathbf{x}^\infty)$.

Remark A consequence of assumption (**A2**) and (**A4**) 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 $\|\mathbf{s}\| = 1$.

We first summarise some preliminary results. The following lemma shows an important property of the step $\widetilde{\mathbf{d}}_k$ that plays a crucial role in the local convergence analysis.

Lemma 1 *Let the standard assumptions hold and consider the step $\widetilde{\mathbf{d}}_k$ from the solution of the $\widetilde{QP}(\mathbf{x}_k)$ subproblem. For k sufficiently large, $\mathbf{x}_k \in \mathcal{N}^\infty$ and $\widetilde{\mathbf{d}}_k \in \mathbb{R}^n$ then*

$$\|\widetilde{\mathbf{d}}_k\| \quad = \quad O(\|\mathbf{d}_k\|^2).$$

Proof From the $\widetilde{QP}(\mathbf{x}_k)$ subproblem, the optimal step $\widetilde{\mathbf{d}}_k$ satisfies

$$\nabla \widehat{\mathbf{c}}(\mathbf{x}_k)^T \widetilde{\mathbf{d}}_k + \widehat{\mathbf{c}}(\mathbf{x}_k + \mathbf{d}_k) \quad = \quad -\|\mathbf{d}_k\|^\nu \mathbf{e}$$

where $\nabla \widehat{\mathbf{c}}(\mathbf{x}_k) = \{\nabla c_i(\mathbf{x}_k), i \in \widetilde{\mathcal{A}}(\mathbf{x}_k)\}$, $\widehat{\mathbf{c}}(\mathbf{x}_k + \mathbf{d}_k) = \{c_i(\mathbf{x}_k + \mathbf{d}_k), i \in \widetilde{\mathcal{A}}(\mathbf{x}_k)\}$ and $\mathbf{e} = (1, 1, \ldots, 1)^T$ a vector of $|\widetilde{\mathcal{A}}(\mathbf{x}_k)|$ - tuples.

Following assumption (**A3**) and by continuity,

$$\widetilde{\mathbf{d}}_k \quad = \quad -\left[\nabla \widehat{\mathbf{c}}(\mathbf{x}_k)^T\right]^+ \{\|\mathbf{d}_k\|^\nu \mathbf{e} + \widehat{\mathbf{c}}(\mathbf{x}_k + \mathbf{d}_k)\} \tag{3.1}$$

where $\left[\nabla \widehat{\mathbf{c}}(\mathbf{x}_k)^T\right]^+$ is the pseudo-inverse of $\nabla \widehat{\mathbf{c}}(\mathbf{x}_k)^T$. Expanding (3.1) using Taylor's theorem and taking note that $\nabla c_i(\mathbf{x}_k)^T \mathbf{d}_k + c_i(\mathbf{x}_k) = 0$, $i \in \mathcal{A}(\mathbf{x}_k)$ and since $\widetilde{\mathcal{A}}(\mathbf{x}_k) \subseteq \mathcal{A}(\mathbf{x}_k)$ we then have

$$\widetilde{\mathbf{d}}_k \quad = \quad -\left[\nabla \widehat{\mathbf{c}}(\mathbf{x}_k)^T\right]^+ \left(\|\mathbf{d}_k\|^\nu + O(\|\mathbf{d}_k\|^2\right) \mathbf{e}.$$

where $\nu \in (2, 3)$. Hence from the boundedness assumption $\|\widetilde{\mathbf{d}}_k\| = O(\|\mathbf{d}_k\|^2)$.

q.e.d

The next lemma concerns with the global convergence proof by showing that if there exists subsequences of $\{\mathbf{x}_k\}_{k\in\mathcal{K}}$ that converge to a Karush-Kuhn-Tucker (KKT) point then $\{\mathbf{d}_k\}_{k\in\mathcal{K}}$ tends to zero.

Lemma 2 *Let the standard assumptions hold and let $\{\mathbf{x}_k\}_{k\in\mathcal{K}}$ be any subsequences converging to an accumulation point $\mathbf{x}^\infty$. Hence $\mathbf{x}^\infty$ satisfies KKT conditions and $\{\mathbf{d}_k\}_{k\in\mathcal{K}} \to \mathbf{0}$.*

Proof The result of the first part of the lemma can be found in Theorem 2 of Chin [4]. To show that $\{\mathbf{d}_k\}_{k\in\mathcal{K}} \to \mathbf{0}$, we let $\mathbf{x}^\infty$ be the KKT point of Problem P. Then $\mathbf{d}^\infty = \mathbf{0}$ solves the problem

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

where conditions (**A3**) - (**A6**) are satisfied.

For $\mathbf{x}_k \in \mathcal{N}^\infty$, k sufficiently large, $k \in \mathcal{K}$, let $\mathbf{d}_k$ solves the problem

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

Hence from the first-order necessary conditions and by continuity, $\mathbf{d}^\infty$ and $\mathbf{d}_k$ satisfy

$$\mathbf{W}^\infty\mathbf{d}^\infty + \nabla f(\mathbf{x}^\infty) + \sum_{i\in\mathcal{J}(\mathbf{x}^\infty)} \lambda_i^\infty c_i(\mathbf{x}^\infty) = \mathbf{0} \tag{3.2}$$

$$\mathbf{W}_k\mathbf{d}_k + \nabla f(\mathbf{x}_k) + \sum_{i\in\mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} c_i(\mathbf{x}_k) = \mathbf{0} \tag{3.3}$$

respectively. Deducting (3.2) from (3.3), we then have

$$\mathbf{W}_k\mathbf{d}_k - \mathbf{W}^\infty\mathbf{d}^\infty + \nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}^\infty) + \sum_{i\in J(\mathbf{x}^\infty)} \left\{\lambda_k^{(i)}\nabla c_i(\mathbf{x}_k) - \lambda_i^\infty \nabla c_i(\mathbf{x}^\infty)\right\} = \mathbf{0}$$

or

$$\begin{array}{c} (\mathbf{W}_k - \mathbf{W}^\infty)\mathbf{d}_k + \mathbf{W}^\infty(\mathbf{d}_k - \mathbf{d}^\infty) + \nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}^\infty) + \\ \displaystyle\sum_{i\in\mathcal{J}(\mathbf{x}^\infty)} \left\{(\lambda_k^{(i)} - \lambda_i^\infty)\nabla c_i(\mathbf{x}_k) + \lambda_i^\infty(\nabla c_i(\mathbf{x}_k) - \nabla c_i(\mathbf{x}^\infty))\right\} = \mathbf{0}. \end{array} \tag{3.4}$$

Taking an inner product (3.4) with $\mathbf{d}_k$,

$$\begin{array}{c} \mathbf{d}_k^T(\mathbf{W}_k - \mathbf{W}^\infty)\mathbf{d}_k + \mathbf{d}_k^T\mathbf{W}^\infty(\mathbf{d}_k - \mathbf{d}^\infty) + (\nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}^\infty))^T\mathbf{d}_k + \\ \displaystyle\sum_{i\in\mathcal{J}(\mathbf{x}^\infty)} \left\{(\lambda_k^{(i)} - \lambda_i^\infty)\nabla c_i(\mathbf{x}_k)^T\mathbf{d}_k + \lambda_i^\infty(\nabla c_i(\mathbf{x}_k) - \nabla c_i(\mathbf{x}^\infty))^T\mathbf{d}_k\right\} = 0. \end{array} \tag{3.5}$$

For k sufficiently large, $k \in \mathcal{K}$, by continuity $\lambda_k^{(i)} > 0$ for $i \in \mathcal{J}(\mathbf{x}^\infty)$. We then have $\nabla c_i(\mathbf{x}_k)^T\mathbf{d}_k + c_i(\mathbf{x}_k) = 0$, $i \in \mathcal{J}(\mathbf{x}^\infty)$ and hence (3.5) can also be written as

$$\begin{array}{c} \mathbf{d}_k^T(\mathbf{W}_k - \mathbf{W}^\infty)\mathbf{d}_k + \mathbf{d}_k^T\mathbf{W}^\infty(\mathbf{d}_k - \mathbf{d}^\infty) + (\nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}^\infty))^T\mathbf{d}_k + \\ \displaystyle\sum_{i\in\mathcal{J}(\mathbf{x}^\infty)} \left\{(\lambda_i^\infty - \lambda_k^{(i)})c_i(\mathbf{x}_k) + \lambda_i^\infty(\nabla c_i(\mathbf{x}_k) - \nabla c_i(\mathbf{x}^\infty))^T\mathbf{d}_k\right\} = 0. \end{array}$$

For k sufficiently large, $k \in \mathcal{K}$, $\mathbf{x}_k \to \mathbf{x}^\infty$ and by continuity $\mathbf{W}_k \to \mathbf{W}^\infty$, $\nabla f(\mathbf{x}_k) \to \nabla f(\mathbf{x}^\infty)$, $\nabla c_i(\mathbf{x}_k) \to \nabla c_i(\mathbf{x}^\infty)$, $c_i(\mathbf{x}_k) \to c_i(\mathbf{x}^\infty) = 0$, $i \in \mathcal{J}(\mathbf{x}^\infty)$, thus $\{\mathbf{d}_k\}_{k\in\mathcal{K}} \to \mathbf{0}$.

q.e.d

The following lemma presents a stronger result such that if the accumulation point $\mathbf{x}^\infty$ of any subsequences satisfies Karush-Kuhn-Tucker (KKT) conditions then the entire sequence $\{\mathbf{x}_k\} \to \mathbf{x}^\infty$.

Lemma 3 *Let the standard assumptions hold and if the accumulation point $\mathbf{x}^\infty$ of any subsequences satisfies KKT conditions then for $\mathbf{x}_k \in \mathcal{N}^\infty$, k sufficiently large the entire sequence $\{\mathbf{x}_k\}$ converges to $\mathbf{x}^\infty$.*

Proof To prove this lemma we need to show the accumulation point $\mathbf{x}^\infty$ of any subsequences that satisfies KKT conditions is isolated under the standard assumptions. Let $\{\mathbf{x}_k\}_{k\in\mathcal{K}}$ be any subsequence converging to $\mathbf{x}^\infty$ and if $\mathbf{x}_k \in \mathcal{N}^\infty$ then there exists $\varepsilon > 0$ such that

$$\|\mathbf{x}_k - \mathbf{x}^\infty\| \;\; \leq \;\; \varepsilon$$

for all $k \in \mathcal{K}$. We then have

$$\|\mathbf{x}_{k+1} - \mathbf{x}^\infty\| \;\; \leq \;\; \|\mathbf{x}_{k+1} - \mathbf{x}_k\| + \|\mathbf{x}_k - \mathbf{x}^\infty\|.$$

Thus for k large enough, $k \in \mathcal{K}$, we have

$$\begin{aligned} \|\mathbf{x}_{k+1} - \mathbf{x}_k\| \;\; &\leq \;\; \begin{cases} \alpha_k\|\mathbf{d}_k\| & \text{if } QP \text{ step is used} \\ \alpha_k\|\mathbf{d}_k\| + \alpha_k^2\|\widetilde{\mathbf{d}}_k\| & \text{if } SOC \text{ step is used.} \end{cases} \\ &\leq \;\; 2\alpha_k\|\mathbf{d}_k\| \end{aligned}$$

due to Lemma 1 and Lemma 2.

From the definition of $\mathbf{d}_k$ in Lemma 2, we then have $\alpha_k\mathbf{d}_k \to \mathbf{0}$ for $k \to \infty$, $k \in \mathcal{K}$. Therefore for k large enough, $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| \leq \varepsilon/2$ and there also exists a subsequence $\{\mathbf{x}_k\}_{k\in\mathcal{K}}$ such that $\|\mathbf{x}_k - \mathbf{x}^\infty\| \leq \varepsilon/2$. Hence

$$\|\mathbf{x}_{k+1} - \mathbf{x}^\infty\| \;\; \leq \;\; \varepsilon$$

and it is then impossible for the iterates $\{\mathbf{x}_k\}$ to leave the neighbourhood $\mathcal{N}^\infty$ without generating another accumulation point, and hence a KKT point in that neighbourhood. Thus for $\mathbf{x}_k \in \mathcal{N}^\infty$, k large enough $\{\mathbf{x}_k\} \to \mathbf{x}^\infty$.

q.e.d

The next lemma is a consequence of Lemma 3 where as $k \to \infty$, $\{\mathbf{x}_k\} \to \mathbf{x}^\infty$ then the QP step converges to zero, the multipliers converge to $\boldsymbol{\lambda}^\infty$ and the identification of the correct active set can be determined.

Lemma 4 *Let the standard assumptions hold and for k sufficiently large let $\{\mathbf{x}_k\} \to \mathbf{x}^\infty$ where $\mathbf{x}^\infty$ satisfies KKT conditions. Hence for k large enough* *(i)* $\{\mathbf{d}_k\} \rightarrow \mathbf{0}$;

(ii) $\{\boldsymbol{\lambda}_k\} \rightarrow \boldsymbol{\lambda}^\infty$;

(iii) $\mathcal{A}(\mathbf{x}_k) = \mathcal{E}(\mathbf{x}^\infty)$;

(iv) $\widetilde{\mathcal{A}}(\mathbf{x}_k) = \mathcal{J}(\mathbf{x}^\infty)$ *where*

$\mathcal{A}(\mathbf{x}_k) = \{i \ : \ c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T \mathbf{d}_k = 0\}$, $\mathcal{E}(\mathbf{x}^\infty) = \{i \ : \ c_i(\mathbf{x}^\infty) = 0\}$, $\widetilde{\mathcal{A}}(\mathbf{x}_k) = \{i \in \mathcal{A}(\mathbf{x}_k) \ : \ \lambda_k^{(i)} > 0\}$, $\mathcal{J}(\mathbf{x}^\infty) = \{i \in \mathcal{E}(\mathbf{x}^\infty) \ : \ \lambda_i^\infty > 0\}$.

Proof The result of part (i) follows from Lemma 2 and Lemma 3. As for part (ii), from (3.4) of Lemma 2, we have

$$\begin{aligned}(\mathbf{W}_k - \mathbf{W}^\infty)\mathbf{d}_k + \mathbf{W}^\infty(\mathbf{d}_k - \mathbf{d}^\infty) + \nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}^\infty) + \\ \nabla \mathbf{c}(\mathbf{x}_k)(\boldsymbol{\lambda}_k - \boldsymbol{\lambda}^\infty) + (\nabla \mathbf{c}(\mathbf{x}_k) - \nabla \mathbf{c}(\mathbf{x}^\infty))\boldsymbol{\lambda}^\infty &= \mathbf{0}.\end{aligned}$$

Since $\mathbf{x}_k \rightarrow \mathbf{x}^\infty$, $\mathbf{d}_k \rightarrow \mathbf{0}$ and from continuity we then have $\boldsymbol{\lambda}_k \rightarrow \boldsymbol{\lambda}^\infty$.

To prove part (iii) and (iv), we note that as $k \rightarrow \infty$, $\mathbf{x}_k \rightarrow \mathbf{x}^\infty$, $\mathbf{d}_k \rightarrow \mathbf{0}$, $\boldsymbol{\lambda}_k \rightarrow \boldsymbol{\lambda}^\infty$, we therefore have $\mathcal{A}(\mathbf{x}_k) = \mathcal{E}(\mathbf{x}^\infty)$ and also $\widetilde{\mathcal{A}}(\mathbf{x}_k) = \mathcal{J}(\mathbf{x}^\infty)$. Thus the result is proved.

q.e.d

Using the previous results we will show that in the neighbourhood of $\mathbf{x}^\infty$, asymptotically the nonlinear constraints are feasible when *SOC* steps are used with step size $\alpha = 1$.

Lemma 5 *Let the standard assumptions hold and let* $\mathbf{x}_k$ *be in the neighbourhood* $\mathcal{N}^\infty$ *of* $\mathbf{x}^\infty$. *For* k *sufficiently large*

$$h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)) = 0.$$

Proof The proof of this result consists of two parts.

For $i \notin \mathcal{E}(\mathbf{x}^\infty)$ then for k sufficiently large, $\mathbf{x}_k \in \mathcal{N}^\infty$, there exists a constant $c > 0$ such that

$$c_i(\mathbf{x}_k) \leq -c.$$

Hence using Taylor's theorem

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

where $\mathbf{y}_i$ is between $\mathbf{x}_k + \mathbf{d}_k$ and $\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k$. Hence

$$\begin{aligned}c_i(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k) &= c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T \mathbf{d}_k + \nabla c_i(\mathbf{x}_k)^T \widetilde{\mathbf{d}}_k + \\ &\quad O(\|\mathbf{d}_k\|^2) + O(\|\mathbf{d}_k\|\|\widetilde{\mathbf{d}}_k\|) + O(\|\widetilde{\mathbf{d}}_k\|^2) \\ &\leq -c + O(\|\mathbf{d}_k\|).\end{aligned}$$

Take note that the right-hand-side term is negative since $\|\mathbf{d}_k\| \to 0$ as $k \to \infty$. Hence for $i \notin \mathcal{E}(\mathbf{x}^\infty)$, $c_i(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k) \leq 0$.

As for $i \in \mathcal{E}(\mathbf{x}^\infty)$, using Taylor's theorem

$$\begin{aligned} c_i(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k) &= c_i(\mathbf{x}_k + \mathbf{d}_k) + \nabla c_i(\mathbf{x}_k + \mathbf{d}_k)^T \widetilde{\mathbf{d}}_k + \frac{1}{2}\widetilde{\mathbf{d}}_k^T \nabla^2 c_i(\mathbf{z}_i)\widetilde{\mathbf{d}}_k \\ &= c_i(\mathbf{x}_k + \mathbf{d}_k) + \nabla c_i(\mathbf{x}_k)^T \widetilde{\mathbf{d}}_k + \frac{1}{2}\mathbf{d}_k^T \nabla^2 c_i(\mathbf{w}_i)\widetilde{\mathbf{d}}_k + \frac{1}{2}\widetilde{\mathbf{d}}_k^T \nabla^2 c_i(\mathbf{z}_i)\widetilde{\mathbf{d}}_k \end{aligned}$$

where $\mathbf{w}_i$ is between $\mathbf{x}_k$ and $\mathbf{x}_k + \mathbf{d}_k$ and $\mathbf{z}_i$ is between $\mathbf{x}_k + \mathbf{d}_k$ and $\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k$. From the boundedness assumption we have

$$\begin{aligned} c_i(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k) &\leq -\|\mathbf{d}_k\|^\nu + O(\|\mathbf{d}_k\|\|\widetilde{\mathbf{d}}_k\|) + O(\|\widetilde{\mathbf{d}}_k\|^2) \\ &= -\|\mathbf{d}_k\|^\nu + O(\|\mathbf{d}_k\|^3) \end{aligned}$$

where $\nu \in (2, 3)$ and $\|\widetilde{\mathbf{d}}_k\| = O(\|\mathbf{d}_k\|^2)$. The last term is also negative since $\mathbf{d}_k \to \mathbf{0}$, $k \to \infty$.

Thus from the two cases, we can deduce that $h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)) = 0$ for $k \to \infty$, $\mathbf{x}_k \in \mathcal{N}^\infty$.

q.e.d

The next two lemmas provide some intermediate results that will be used in the local convergence proof.

Lemma 6 *Let the QP step* $\mathbf{d}_k$ *be decomposed into* $\mathbf{d}_k = \mathbf{P}_k\mathbf{d}_k + \mathbf{s}_k$ *where* $\nabla c_i(\mathbf{x}_k)^T\mathbf{P}_k\mathbf{d}_k = 0$ *and* $\nabla c_i(\mathbf{x}_k)^T\mathbf{s}_k + c_i(\mathbf{x}_k) = 0$ *for* $i \in \mathcal{J}(\mathbf{x}^\infty)$. *Then for* k *sufficiently large*

$$\begin{aligned} \|\mathbf{s}_k\| &= O(\|\mathbf{c}_k\|) \\ \|\mathbf{d}_k\| &= O(\|\mathbf{c}_k\|) \end{aligned}$$

where $\|\mathbf{c}_k\| = \sqrt{\sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} c_i(\mathbf{x}_k)^2}$.

Proof Let $\mathbf{A}_k$ consists of column vectors $\nabla c_i(\mathbf{x}_k)$, $i \in \mathcal{J}(\mathbf{x}^\infty)$, then by linear independence and continuity $\mathbf{A}_k$ has full column rank. Hence for the linear system

$$\mathbf{A}_k^T\mathbf{s}_k + \mathbf{c}_k = \mathbf{0}$$

where $\mathbf{c}_k$ is the corresponding column vector, we can write $\mathbf{s}_k$ as

$$\mathbf{s}_k = -(\mathbf{A}_k^T)^+\mathbf{c}_k$$

such that $(\mathbf{A}_k^T)^+$ is the pseudo-inverse of $\mathbf{A}_k^T$. By the boundedness assumption we then have

$$\|\mathbf{s}_k\| = O(\|\mathbf{c}_k\|).$$

Similarly because $\mathbf{A}_k^T \mathbf{d}_k + \mathbf{c}_k = \mathbf{0}$, the result

$$\|\mathbf{d}_k\| = O(\|\mathbf{c}_k\|)$$

also holds.

q.e.d

Lemma 7 *Let the standard assumptions hold and if* $\mathbf{x}_k$ *is a feasible point then for* k *sufficiently large,*

$$\sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} c_i(\mathbf{x}_k) \leq -\bar{\lambda}\|\mathbf{c}_k\|$$

where $\|\mathbf{c}_k\| = \sqrt{\sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} c_i(\mathbf{x}_k)^2}$ *and* $\bar{\lambda} > 0$.

Proof Since $\mathbf{x}_k$ is a feasible point therefore $c_i(\mathbf{x}_k) \leq 0$, $i = 1, 2, \ldots, m$. By continuity, $\lambda_k^{(i)} > 0$, $i \in \mathcal{J}(\mathbf{x}^\infty)$, we then have

$$\sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} c_i(\mathbf{x}_k) \leq \min\left\{\lambda_k^{(i)}, i \in \mathcal{J}(\mathbf{x}^\infty)\right\} \sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} c_i(\mathbf{x}_k).$$

From the convergence of the multipliers (Lemma 4) and from complementary slackness,

$$\lambda_k^{(i)} \to \lambda_{(i)}^\infty > 0$$

for $i \in \mathcal{J}(\mathbf{x}^\infty)$ as $k \to \infty$. We can then write

$$\begin{aligned} \min\left\{\lambda_k^{(i)}, i \in \mathcal{J}(\mathbf{x}^\infty)\right\} &\geq \min\left\{\frac{\lambda_{(i)}^\infty}{2}, i \in \mathcal{J}(\mathbf{x}^\infty)\right\} \\ &= \bar{\lambda} \end{aligned}$$

where $\bar{\lambda} > 0$.

Hence using the identity $\|\mathbf{z}\| \leq \|\mathbf{z}\|_1$ for $\mathbf{z} \in \mathbb{R}^n$, we then have

$$\begin{aligned} \sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} c_i(\mathbf{x}_k) &\leq \bar{\lambda} \sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} c_i(\mathbf{x}_k) \\ &\leq -\bar{\lambda}\|\mathbf{c}_k\| \end{aligned}$$

where $\|\mathbf{c}_k\| = \sqrt{\sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} c_i(\mathbf{x}_k)^2}$.

q.e.d

The following lemma is a special case in which for a non-KKT point $\mathbf{x}_k \in \mathcal{N}^\infty$ such that $h(\mathbf{c}(\mathbf{x}_k)) = 0$, then for k large enough the predicted reduction is positive.

Lemma 8 *Let the standard assumptions hold and if $\mathbf{x}_k$ is a feasible point then for k sufficiently large, $\mathbf{x}_k$ in the neighbourhood $\mathcal{N}^\infty$ of $\mathbf{x}^\infty$,*

$$\nabla f(\mathbf{x}_k)^T \mathbf{d}_k \quad \leq \quad -\rho \|\mathbf{d}_k\|^2$$

where $\rho > 0$.

Proof From the first-order necessary conditions of the $QP(\mathbf{x}_k)$ subproblem, the step $\mathbf{d}_k$ satisfies

$$\nabla f(\mathbf{x}_k) + \mathbf{W}_k \mathbf{d}_k + \sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} \nabla c_i(\mathbf{x}_k) \quad = \quad \mathbf{0}.$$

Taking an inner product with $\mathbf{d}_k$ we then have

$$\begin{aligned} \nabla f(\mathbf{x}_k)^T \mathbf{d}_k &= -\sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} \nabla c_i(\mathbf{x}_k)^T \mathbf{d}_k - \mathbf{d}_k^T \mathbf{W}_k \mathbf{d}_k \\ &= \sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} c_i(\mathbf{x}_k) - \mathbf{d}_k^T \mathbf{W}_k \mathbf{d}_k. \end{aligned}$$

By replacing $\mathbf{d}_k$ by it's composition $\mathbf{d}_k = \mathbf{P}_k \mathbf{d}_k + \mathbf{s}_k$ we have

$$\begin{aligned} \nabla f(\mathbf{x}_k)^T \mathbf{d}_k &= \sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} c_i(\mathbf{x}_k) - (\mathbf{P}_k \mathbf{d}_k + \mathbf{s}_k)^T \mathbf{W}_k (\mathbf{P}_k \mathbf{d}_k + \mathbf{s}_k) \\ &= \sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} c_i(\mathbf{x}_k) - \mathbf{d}_k^T \mathbf{P}_k^T \mathbf{W}_k \mathbf{P}_k \mathbf{d}_k - 2\mathbf{d}_k^T \mathbf{P}_k^T \mathbf{W}_k \mathbf{s}_k - \mathbf{s}_k^T \mathbf{W}_k \mathbf{s}_k. \end{aligned}$$

From assumption (**A5**), Lemma 6 and 7,

$$\begin{aligned} \nabla f(\mathbf{x}_k)^T \mathbf{d}_k &\leq -\bar{\lambda}\|\mathbf{c}_k\| - \rho \|\mathbf{P}_k \mathbf{d}_k\|^2 + O(\|\mathbf{P}_k \mathbf{d}_k\| \|\mathbf{s}_k\|) + O(\|\mathbf{s}_k\|^2) \\ &= -\bar{\lambda}\|\mathbf{c}_k\| - \rho \|\mathbf{P}_k \mathbf{d}_k\|^2 + O(\|\mathbf{d}_k\| \|\mathbf{s}_k\|) + O(\|\mathbf{s}_k\|^2) \\ &= -\bar{\lambda}\|\mathbf{c}_k\| - \rho \|\mathbf{P}_k \mathbf{d}_k\|^2 + O(\|\mathbf{c}_k\|)^2). \end{aligned}$$

From the decomposition $\mathbf{P}_k \mathbf{d}_k = \mathbf{d}_k - \mathbf{s}_k$, we then have

$$\begin{aligned} \|\mathbf{P}_k \mathbf{d}_k\| &= \|\mathbf{d}_k - \mathbf{s}_k\| \\ &\geq |\ \|\mathbf{d}_k\| - \|\mathbf{s}_k\|\ |. \end{aligned}$$

Hence

$$\begin{aligned} \|\mathbf{P}_k \mathbf{d}_k\|^2 &\geq \|\mathbf{d}_k\|^2 - 2\|\mathbf{d}_k\| \|\mathbf{s}_k\| + \|\mathbf{s}_k\|^2 \\ &= \|\mathbf{d}_k\|^2 + O(\|\mathbf{c}_k\|^2). \end{aligned}$$

Therefore

$$\nabla f(\mathbf{x}_k)^T \mathbf{d}_k \quad \leq \quad -\bar{\lambda}\|\mathbf{c}_k\| - \rho \|\mathbf{d}_k\|^2 + O(\|\mathbf{c}_k\|^2)$$

and the term $-\bar{\lambda}\|\mathbf{c}_k\|+O(\|\mathbf{c}_k\|^2)$ is negative since $\{\mathbf{c}_k\} \to \mathbf{0}$ as $k \to \infty$. Thus $\nabla f(\mathbf{x}_k)^T \mathbf{d}_k \leq -\rho\|\mathbf{d}_k\|^2$.

q.e.d

We now come to the stage of proving the local convergence theorem. First we need to show that asymptotically as $\mathbf{x}_k \in \mathcal{N}^\infty$, k sufficiently large, the filter will not reject the SOC steps at step size $\alpha = 1$. Finally two-step superlinear convergence of the iterates is given in Theorem 2.

Theorem 1 *Let the standard assumptions hold, and for k large enough $(h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)), f(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k))$ is acceptable to the filter.*

Proof To prove this result we need to consider two cases depending on whether $h(\mathbf{c}(\mathbf{x}_k)) = 0$ or not.

case 1 $h(\mathbf{c}(\mathbf{x}_k)) > 0$
To show that $(h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)), f(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k))$ is acceptable to the filter for the case of $h(\mathbf{c}(\mathbf{x}_k)) > 0$, we focus on

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

the minimum of all $h(\mathbf{c}(\mathbf{x}))$ values in the current filter set $\mathcal{F}^{(k)}$ such that

$$\tau_k \;\leq\; h(\mathbf{c}(\mathbf{x}_k)).$$

From Lemma 5, for k sufficiently large we have $h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)) = 0$, therefore

$$h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)) \;\leq\; (1 - \eta_1)\tau_k$$

and hence

$$h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)) \;\leq\; (1 - \eta_1)h(\mathbf{c}(\mathbf{x}_k)).$$

Thus the trial point $\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k$ satisfies the filter test for all filter entries in $\mathcal{F}^{(k)}$ and also to $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$.

If the predicted reduction $\Delta l = -\nabla f(\mathbf{x}_k)^T \mathbf{d}_k \leq 0$ then according to Step 5 of the algorithm, the filter test is sufficient for $(h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)), f(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k))$ to be acceptable by the filter. Hence in this special case, the trial pair $(h(\mathbf{c}(\mathbf{x}_k+\mathbf{d}_k+\widetilde{\mathbf{d}}_k)), f(\mathbf{x}_k+\mathbf{d}_k + \widetilde{\mathbf{d}}_k))$ is acceptable to the filter algorithm.

On the other hand if $\Delta l = -\nabla f(\mathbf{x}_k)^T \mathbf{d}_k > 0$, then we need to show also that $\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k$ must satisfy the sufficient reduction test in order for $(h(\mathbf{c}(\mathbf{x}_k+\mathbf{d}_k+\widetilde{\mathbf{d}}_k)), f(\mathbf{x}_k+\mathbf{d}_k+\widetilde{\mathbf{d}}_k))$ to be acceptable by the filter. Because $\mathbf{x}_k$ is not a KKT point we can then denote the optimal QP solution $\mathbf{d}_k = \|\mathbf{d}_k\|\mathbf{s}$ where $\|\mathbf{s}\| = 1$ such that

$$\nabla f(\mathbf{x}_k)^T \mathbf{s} \;\leq\; -\varepsilon$$

where $\varepsilon > 0$. Hence

$$\nabla f(\mathbf{x}_k)^T \mathbf{d}_k \;\; \leq \;\; -\varepsilon \|\mathbf{d}_k\|.$$

Using the sufficient reduction condition we have

$$\begin{aligned} f(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k) - f(\mathbf{x}_k) - \sigma \nabla f(\mathbf{x}_k)^T \mathbf{d}_k &= f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T (\mathbf{d}_k + \widetilde{\mathbf{d}}_k) + \\ & \quad O(\|\mathbf{d}_k + \widetilde{\mathbf{d}}_k\|^2) - f(\mathbf{x}_k) - \sigma \nabla f(\mathbf{x}_k)^T \mathbf{d}_k \\ &= (1-\sigma) \nabla f(\mathbf{x}_k)^T \mathbf{d}_k + O(\|\widetilde{\mathbf{d}}_k\|) + O(\|\mathbf{d}_k\|^2) \\ &\leq -(1-\sigma)\varepsilon \|\mathbf{d}_k\| + O(\|\mathbf{d}_k\|^2). \end{aligned}$$

Since $\mathbf{d}_k \to \mathbf{0}$ as $k \to \infty$, the last term is negative and hence $\Delta f \geq \sigma \Delta l$. Therefore for the case of $\Delta l > 0$, the trial pair $(h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)), f(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k))$ satisfies the filter test for all filter entries in $\mathcal{F}^{(k)}$ and also to the current filter pair $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$. Furthermore the trial pair also satisfies the sufficient reduction test and thus an f-type step is generated. Hence $(h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)), f(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k))$ is also acceptable to the filter.

case 2 $h(\mathbf{c}(\mathbf{x}_k)) = 0$
If $h(\mathbf{c}(\mathbf{x}_k)) = 0$ and since $h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)) = 0$ for k large enough then the trial point $\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k$ satisfies the filter test for $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$. By denoting

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

as the minimum of all $h(\mathbf{c}(\mathbf{x}))$ values in the current filter set $\mathcal{F}^{(k)}$ and since $h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)) = 0$, we then have

$$h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)) \;\; \leq \;\; (1-\eta_1)\tau_k.$$

Therefore the trial pair $(h(\mathbf{c}(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k)), f(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k))$ is also acceptable to all filter entries in $\mathcal{F}^{(k)}$. Hence the trial iterate $\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k$ satisfies the filter test for $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$ and also for all filter entries in $\mathcal{F}^{(k)}$.

Because $\mathbf{x}_k \in \mathcal{N}^\infty$ is a feasible point then from Lemma 8,

$$\nabla f(\mathbf{x}_k)^T \mathbf{d}_k \;\; \leq \;\; -\rho \|\mathbf{d}_k\|^2$$

where $\rho > 0$. We now need to show that $\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k$ must also satisfy the sufficient reduction test in order for it to acceptable by the filter. By definition, the actual and predicted reduction in f for $\alpha = 1$ are

$$\begin{aligned} \Delta f &= f(\mathbf{x}_k) - f(\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k) \\ \Delta l &= -\nabla f(\mathbf{x}_k)^T \mathbf{d}_k \end{aligned}$$

respectively. From Taylor's expansion

$$\begin{aligned}\Delta f &= f(\mathbf{x}_k) - f(\mathbf{x}_k) - \nabla f(\mathbf{x}_k)^T(\mathbf{d}_k + \widetilde{\mathbf{d}}_k) - \\ &\quad \frac{1}{2}(\mathbf{d}_k + \widetilde{\mathbf{d}}_k)^T \nabla^2 f(\mathbf{x}_k)(\mathbf{d}_k + \widetilde{\mathbf{d}}_k) + O(\|\mathbf{d}_k + \widetilde{\mathbf{d}}_k\|^3) \\ &= -\nabla f(\mathbf{x}_k)^T \mathbf{d}_k - \nabla f(\mathbf{x}_k)^T \widetilde{\mathbf{d}}_k + \frac{1}{2}\mathbf{d}_k^T \nabla^2 f(\mathbf{x}_k)\mathbf{d}_k + O(\|\mathbf{d}_k\|^3) \\ &= \sigma\Delta l + (1-\sigma)\Delta l - \nabla f(\mathbf{x}_k)^T \widetilde{\mathbf{d}}_k - \frac{1}{2}\mathbf{d}_k^T \nabla^2 f(\mathbf{x}_k)\mathbf{d}_k + O(\|\mathbf{d}_k\|^3) \\ &= \sigma\Delta l + \theta_k \end{aligned} \tag{3.6}$$

where

$$\begin{aligned}\theta_k &= \left(\frac{1}{2} - \sigma\right)\Delta l - \frac{1}{2}\nabla f(\mathbf{x}_k)^T \mathbf{d}_k - \nabla f(\mathbf{x}_k)^T \widetilde{\mathbf{d}}_k - \\ &\quad \frac{1}{2}\mathbf{d}_k^T \nabla^2 f(\mathbf{x}_k)\mathbf{d}_k + O(\|\mathbf{d}_k\|^3). \end{aligned} \tag{3.7}$$

From the first-order necessary conditions of $QP(\mathbf{x}_k)$ subproblem, for k sufficiently large

$$\nabla f(\mathbf{x}_k) + \mathbf{W}_k \mathbf{d}_k + \sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} \nabla c_i(\mathbf{x}_k) = \mathbf{0}. \tag{3.8}$$

Taking an inner product of (3.8) with the optimal QP point $\mathbf{d}_k$, and multiplying with one-half we have

$$\frac{1}{2}\nabla f(\mathbf{x}_k)^T \mathbf{d}_k + \frac{1}{2}\mathbf{d}_k^T \mathbf{W}_k \mathbf{d}_k + \sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \frac{1}{2}\lambda_k^{(i)} \nabla c_i(\mathbf{x}_k)^T \mathbf{d}_k = 0$$

or

$$\frac{1}{2}\nabla f(\mathbf{x}_k)^T \mathbf{d}_k = -\frac{1}{2}\mathbf{d}_k^T \mathbf{W}_k \mathbf{d}_k + \frac{1}{2}\sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} c_i(\mathbf{x}_k) \tag{3.9}$$

where $\nabla c_i(\mathbf{x}_k)^T \mathbf{d}_k + c_i(\mathbf{x}_k) = 0$, $i \in \mathcal{J}(\mathbf{x}^\infty)$.

In addition, by taking an inner product of (3.8) with the $\widetilde{QP}$ step $\widetilde{\mathbf{d}}_k$, we then have

$$\begin{aligned}\nabla f(\mathbf{x}_k)^T \widetilde{\mathbf{d}}_k &= -\widetilde{\mathbf{d}}_k^T \mathbf{W}_k \mathbf{d}_k - \sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} \nabla c_i(\mathbf{x}_k)^T \widetilde{\mathbf{d}}_k \\ &= -\sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} \nabla c_i(\mathbf{x}_k)^T \widetilde{\mathbf{d}}_k + O(\|\mathbf{d}_k\|^3) \\ &= \sum_{i \in \mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} c_i(\mathbf{x}_k + \mathbf{d}_k) + O(\|\mathbf{d}_k\|^\nu) + O(\|\mathbf{d}_k\|^3) \end{aligned} \tag{3.10}$$

where from Lemma 4, $\nabla c_i(\mathbf{x}_k)^T\widetilde{\mathbf{d}}_k + c_i(\mathbf{x}_k+\mathbf{d}_k) = -\|\mathbf{d}_k\|^\nu$, $i \in \mathcal{J}(\mathbf{x}^\infty)$. Using Taylor's expansion, we can write (3.10) as

$$\begin{aligned}\nabla f(\mathbf{x}_k)^T\widetilde{\mathbf{d}}_k &= \sum_{i\in\mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)}\left(c_i(\mathbf{x}_k) + \nabla c_i(\mathbf{x}_k)^T\mathbf{d}_k + \frac{1}{2}\mathbf{d}_k^T\nabla^2 c_i(\mathbf{x}_k)\mathbf{d}_k\right) + \\ &\quad O(\|\mathbf{d}_k\|^3) + O(\|\mathbf{d}_k\|^\nu) \\ &= \sum_{i\in\mathcal{J}(\mathbf{x}^\infty)} \frac{1}{2}\lambda_k^{(i)}\mathbf{d}_k^T\nabla^2 c_i(\mathbf{x}_k)\mathbf{d}_k + O(\|\mathbf{d}_k\|^\nu) \qquad (3.11)\end{aligned}$$

By substituting (3.9) and (3.11) into (3.7) we have

$$\begin{aligned}\theta_k &= \left(\frac{1}{2}-\sigma\right)\Delta l + \frac{1}{2}\mathbf{d}_k^T\left(\mathbf{W}_k - \nabla^2\mathcal{L}(\mathbf{x}_k,\boldsymbol{\lambda}_k)\right)\mathbf{d}_k - \\ &\quad \frac{1}{2}\sum_{i\in\mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)} c_i(\mathbf{x}_k) + O(\|\mathbf{d}_k\|^\nu)\end{aligned}$$

where $\nabla^2\mathcal{L}(\mathbf{x}_k,\boldsymbol{\lambda}_k) = \nabla^2 f(\mathbf{x}_k) + \sum_{i\in\mathcal{J}(\mathbf{x}^\infty)} \lambda_k^{(i)}\nabla^2 c_i(\mathbf{x}_k)$. From Lemmas 7, 8 and from the convergence of Hessian matrix

$$\begin{aligned}\theta_k &\geq \left(\frac{1}{2}-\sigma\right)\rho\|\mathbf{d}_k\|^2 + \frac{\bar{\lambda}}{2}\|\mathbf{c}_k\| + o(\|\mathbf{d}_k\|^2) \\ &> 0\end{aligned}$$

since $(\frac{1}{2}-\sigma)\rho\|\mathbf{d}_k\|^2 + o(\|\mathbf{d}_k\|^2) > 0$ as $\mathbf{d}_k \to \mathbf{0}$, $k\to\infty$. Therefore from (3.6), $\Delta f \geq \sigma\Delta l$ such that $\Delta l \geq \rho\|\mathbf{d}_k\|^2$. Thus for k sufficiently large, $\mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k$ satisfies the sufficient reduction test, and also $(h(\mathbf{c}(\mathbf{x}_k+\mathbf{d}_k+\widetilde{\mathbf{d}}_k), f(\mathbf{x}_k+\mathbf{d}_k+\widetilde{\mathbf{d}}_k))$ satisfies the filter test for all filter entries in $\mathcal{F}^{(k)}$ and also to $(h(\mathbf{c}(\mathbf{x}_k)), f(\mathbf{x}_k))$. Hence an f-type step is generated and the trial pair $(h(\mathbf{c}(\mathbf{x}_k+\mathbf{d}_k+\widetilde{\mathbf{d}}_k)), f(\mathbf{x}_k+\mathbf{d}_k+\widetilde{\mathbf{d}}_k))$ is acceptable to the filter for this case also.

q.e.d

Theorem 2 *Let the standard assumptions hold and assume that* $\mathbf{x}_k \to \mathbf{x}^\infty$, $k\to\infty$ *where* $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k$. *If the condition*

$$\lim_{k\to\infty} \frac{\|\mathbf{P}_k(\mathbf{W}_k - \mathbf{W}^\infty)\mathbf{P}_k(\mathbf{x}_{k+1}-\mathbf{x}_k)\|}{\|\mathbf{x}_{k+1}-\mathbf{x}_k\|} = 0$$

holds then

$$\frac{\|\mathbf{x}_{k+1}-\mathbf{x}^\infty\|}{\|\mathbf{x}_{k-1}-\mathbf{x}^\infty\|} \to 0.$$

Proof The proof of this theorem is based on Powell [23] with some modifications where it can be shown that if the QP step is accepted at every iteration such that $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$ for all k and if

$$\|\mathbf{P}_k(\mathbf{W}_k - \mathbf{W}^\infty)\mathbf{P}_k\mathbf{d}_k\| \;=\; o(\|\mathbf{d}_k\|) \tag{3.12}$$

then the sequence $\{\mathbf{x}_k\}$ converges two-step superlinearly to $\mathbf{x}^\infty$.

In the context of our algorithm, by setting $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k + \widetilde{\mathbf{d}}_k$ and because $\|\widetilde{\mathbf{d}}_k\| = o(\|\mathbf{d}_k\|)$, we can write

$$\begin{aligned}\|\mathbf{P}_k(\mathbf{W}_k - \mathbf{W}^\infty)\mathbf{P}_k(\mathbf{x}_{k+1} - \mathbf{x}_k)\| &= \|\mathbf{P}_k(\mathbf{W}_k - \mathbf{W}^\infty)\mathbf{P}_k(\mathbf{d}_k + \widetilde{\mathbf{d}}_k)\| \\ &\leq \|\mathbf{P}_k(\mathbf{W}_k - \mathbf{W}^\infty)\mathbf{P}_k\mathbf{d}_k\| + \|\mathbf{P}_k(\mathbf{W}_k - \mathbf{W}^\infty)\mathbf{P}_k\widetilde{\mathbf{d}}_k\| \\ &= \|\mathbf{P}_k(\mathbf{W}_k - \mathbf{W}^\infty)\mathbf{P}_k\mathbf{d}_k\| + o(\|\mathbf{d}_k\|).\end{aligned}$$

Hence for k sufficiently large provided (3.12) holds the sequence $\{\mathbf{x}_k\}$ which uses *SOC* steps will not inhibit superlinear convergence. Therefore Theorem 2 is also true under the assumption that $\mathbf{d}_k + \widetilde{\mathbf{d}}_k$ is accepted for all k sufficiently large.

q.e.d

4 Conclusion

A prototypical algorithm of applying filter strategy in line search SQP methods has been described and global [4] as well as local convergence have been shown, demonstrating the fact that convergence for NLP can be achieved without the need to maintain sufficient descent in a traditional penalty type merit function approach. In addition the convergence proof also shows that two-step superlinear convergence of the iterates can be attained without the assumption of strict complementary slackness. 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. 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.

Another important topic that we expect to consider is to provide convergence proof when incorporating equality constraints into the formulation of our algorithm. The outcome is likely that the convergence to a KKT point is proved under the assumption that a Mangasarian-Fromowitz constraint qualification (MFCQ) (see Gauvin [13], Kyparisis

[17]) is satisfied at the accumulation point. However, convergence to a feasible non-KKT points that do not satisfy MFCQ cannot be ruled out. Finally, there is also 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. It is hoped that a simple and efficient restoration phase algorithm is proposed so that global and local convergence can be proved, and all these questions are still the subject of ongoing research.

References

[1] 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).

[2] R. Chamberlain, C. Lemarechal, H.C. Pedersen and M.J.D. Powell, *The watchdog technique for forcing convergence in algorithms for constrained optimization*, Mathematical Progamming Study **16**, (1982), pp. 1 – 17.

[3] 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).

[4] C.M. Chin, *A global convergence theory of a filter line search method for nonlinear programming*, Numerical Optimization Report, Department of Statistics, University of Oxford, England, (2002).

[5] 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, to appear in Mathematical Programming, (2001).

[6] 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).

[7] T.F. Coleman and A.R. Conn, *Nonlinear prrogramming via an exact penalty function: Asymptotic analysis*, Mathematical Programming, **24**, (1982), pp. 123 – 136.

[8] R. Fletcher, *Second-order corrections for non-differentiable optimization*, In: Numerical Analysis Proceedings Dundee 1981, edited by G.A. Watson, Lecture Notes in Mathematics 912, Springer-Verlag, Berlin and New York, (1982), pp. 85 – 114.

[9] 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).

[10] 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).

[11] 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).

[12] 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).

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

[14] D. Gabay, *Reduced quasi-Newton methods with feasibility improvement for nonlinearly constrained optimization*, Mathematical Programming Study, **16**, (1982), pp. 18 – 44.

[15] L. Grippo, F. Lampariello and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM Journal on Numerical Analysis, **23**(4), (1986), pp. 707 – 716.

[16] J. Herskovits, *A two-stage feasible direction algorithm for nonlinear constrained optimization*, Mathematical Programming, **36**(1), (1987), pp. 19 – 38.

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

[18] 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.

[19] D.Q. Mayne and E. Polak, *A superlinearly convergent algorithm for constrained optimization problems*, Mathematical Programming Study, **16**, (1982), pp. 45 – 61.

[20] E.R. Panier, A.L. Tits, *A superlinear convergent feasible method for the solution of inequality constrained optimization problems*, SIAM Journal on Control and Optimization, **25**(4), (1987), pp. 934 – 950.

[21] E.R. Panier, A.L. Tits, *On combining feasibility, descent and superlinear convergence in inequality constrained optimization*, Mathematical Programming, **59**, (1993), pp. 261 – 276.

[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.J.D. Powell, *The convergence of variable metric methods for nonlinear constrained optimization calculations*, In O.L. Mangasarian, R.R. Meyer and S.M. Robinson (editors), Nonlinear Programming 3, Symposium on Nonlinear Programming, Academic Press, (1978).

[24] S. Ulbrich, *On the superlinear local convergence of a filter-SQP method*, Technical Report, Technische Universität München, Germany, (2002).