On the global convergence of a filter-SQP algorithm

Fletcher, Roger; Leyffer, Sven; Toint, Philippe

Published in:
SIAM Journal on Optimization

DOI:
10.1137/S105262340038081X

Publication date:
2003

Document Version
Early version, also known as pre-print

On the Global Convergence of a Filter–SQP Algorithm

Roger Fletcher*, Sven Leyffer* and Philippe L. Toint†

*Department of Mathematics, University of Dundee, Dundee DD1 4HN, Scotland, UK.
†Department of Mathematics, University of Namur, 61 rue de Bruxelles, B-5000 Namur, Belgium.

Numerical Analysis Report NA/197, October 2000
Revised 23/07/01. Corollary to Lemma 1 corrected, 30/08/05

Abstract

A mechanism for proving global convergence in SQP–filter methods for nonlinear programming is described. Such methods are characterized by their use of the dominance concept of multiobjective optimization, instead of a penalty parameter whose adjustment can be problematic. The main point of interest is to demonstrate how convergence for NLP can be induced without forcing sufficient descent in a penalty-type merit function.

The proof relates to a prototypical algorithm, within which is allowed a range of specific algorithm choices associated with the Hessian matrix representation, updating the trust region radius, and feasibility restoration.

Keywords nonlinear programming, global convergence, filter, multiobjective optimization, SQP.
AMS(2000) subject classifications: 65K05, 49M37, 90C30, 90C26

1 Introduction

In Fletcher and Leyffer [5] a new technique for globalizing methods for nonlinear programming (NLP) is presented. The idea is referred to as an NLP filter and is motivated by the aim of avoiding the need to choose penalty parameters, such as would occur with the use of l_1 penalty functions or augmented Lagrangian functions. Numerical experience with the technique in a sequential quadratic programming (SQP) trust region algorithm is reported in [5] and is very promising. However, no global convergence proof is given in [5], although a number of heuristics are suggested to eliminate obvious situations in which the method might fail to converge.

This paper shows that the filter technique does provide a mechanism for forcing global convergence when used in an appropriate way. The proof relates to an NLP problem with both equations and inequality constraints, and shows that there exists an accumulation point that satisfies first order (Kuhn–Tucker) conditions. The result requires that a Mangasarian–Fromowitz constraint qualification holds at the accumulation point. Other non-trivial assumptions that are made are that the Hessian matrices of the quadratic programming (QP) subproblems are uniformly bounded and that a global solution of the subproblem is found by the QP solver. None of these qualifications to the result are readily circumvented.

The proposed algorithm contains an inner iteration for calculating a suitable trust region radius. In some ways this resembles the use of a backtracking line search along a piecewise linear trajectory. This approach enables us to guarantee that certain conditions used in the convergence proof are met. To a large extent however, the approach allows conventional ideas to be used of halving or doubling (say) the previous trust region radius.

An interesting feature of the proof is that various of the heuristics used in [5] are shown to be unnecessary. These include the NW corner rule, the need to unblock the filter in some cases, and the consequent decision to reduce the strict upper bound on constraint infeasibility. In this paper we also use a slightly different way of defining the sufficient reduction condition to that used in [5]. Another new feature of some interest is that some points may be accepted by the algorithm, without a new entry in the filter being made. This contributes to the non-monotonic properties of the algorithm. In common with [5], we do use a feasibility restoration technique, but are not prescriptive as to how this is done.

Subsequent to the work described in this paper, there have been a number of more recent developments in regard to global convergence of filter-related methods for nonlinear programming. The authors have contributed to other papers that prove global convergence for different algorithmic structures such as an SLP-EQP approach or an approach in which approximate solutions of the SQP step are used, based on a decomposition into normal and tangential steps. Recent work of other authors proves global convergence of filter-related methods in a variety of other contexts such as interior point and line search barrier methods. A brief discussion of these developments is given in Section 4.

2 A Filter–SQP Algorithm

In this paper we consider an NLP problem of the form

$$P \left\{ \begin{array}{lll} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} & f(\mathbf{x}) & \\ \text{subject to} & c_i(\mathbf{x}) = 0 & i \in \mathcal{E} \\ & c_i(\mathbf{x}) \leq 0 & i \in \mathcal{I}, \end{array} \right.$$

where the index sets $\mathcal{E}$ and $\mathcal{I}$ reference the equality and inequality constraints respectively. We denote the cardinality of $\mathcal{E} \cup \mathcal{I}$ by m. We assume for the purposes of our convergence proof that all points that are sampled by the algorithm lie in a non-empty closed and bounded set X. Because the points generated by our algorithm satisfy the linear constraints of the problem, it is readily possible to ensure that this condition holds by including suitable simple upper and lower bounds on $\mathbf{x}$ amongst the constraints of P. The QP subproblem in our algorithm depends upon the value of the current iterate $\mathbf{x}$ and trust region radius ρ, ($\rho > 0$), and is defined by

$$QP(\mathbf{x}, \rho) \left\{ \begin{array}{lll} \underset{\mathbf{d}\in\mathbb{R}^n}{\text{minimize}} & q(\mathbf{d}) := \mathbf{g}^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T B\mathbf{d} & \\ \text{subject to} & c_i + \mathbf{a}_i^T\mathbf{d} = 0 & i \in \mathcal{E} \\ & c_i + \mathbf{a}_i^T\mathbf{d} \leq 0 & i \in \mathcal{I} \\ & \quad \|\mathbf{d}\|_\infty \leq \rho. & \end{array} \right.$$

where we denote $\mathbf{g} = \nabla f(\mathbf{x})$, $c_i = c_i(\mathbf{x})$ and $\mathbf{a}_i = \nabla c_i(\mathbf{x})$. The l_∞ norm is used to define the trust region because it is readily implemented by adding simple bounds to the QP subproblem. The QP subproblem also requires the specification of a matrix B, although this plays a relatively minor part in the analysis of global convergence. For this reason we do not make the dependence on B explicit in the notation. We let $\mathbf{d}$ denote the global solution (if it exists) of $QP(\mathbf{x}, \rho)$. Then we denote

$$\Delta q = q(\mathbf{0}) - q(\mathbf{d}) = -\mathbf{g}^T\mathbf{d} - \tfrac{1}{2}\mathbf{d}^T B\mathbf{d} \tag{2.1}$$

as the *predicted reduction* in $f(\mathbf{x})$ and

$$\Delta f = f(\mathbf{x}) - f(\mathbf{x} + \mathbf{d}). \tag{2.2}$$

as the *actual reduction* in $f(\mathbf{x})$. The measure of constraint infeasibility that we use in this paper is

$$h(\mathbf{c}) = \|\mathbf{c}_\mathcal{I}^+\|_1 + \|\mathbf{c}_\mathcal{E}\|_1 \tag{2.3}$$

where $c_i^+ = \max(0, c_i)$, using the notation that $\mathbf{c}_\mathcal{E}$ and $\mathbf{c}_\mathcal{I}$ are partitions of $\mathbf{c}$ corresponding to equality and inequality constraints, respectively.

The algorithm that we propose is iterative, and the index k is used throughout to refer to the iteration number. The sequence of points accepted by the algorithm is referred to by $\{\mathbf{x}^{(k)}\}$, and quantities derived from $\mathbf{x}^{(k)}$ are superscripted in a similar manner, for example $h^{(k)}$ refers to $h(\mathbf{c}(\mathbf{x}^{(k)}))$ and $f^{(k)}$ to $f(\mathbf{x}^{(k)})$. The matrix B usually differs from iteration to iteration and is generally referred to as $B^{(k)}$. Within the inner loop of the iterative process, $B^{(k)}$ is a constant matrix.

We now turn to the definition of an NLP filter as introduced in [5]. The two aims in an NLP problem are to minimize $f(\mathbf{x})$, and to satisfy the constraints, that is to minimize $h(\mathbf{c}(\mathbf{x}))$. In a filter we consider pairs of values (h, f) obtained by evaluating $h(\mathbf{c}(\mathbf{x}))$ and $f(\mathbf{x})$ for various values of $\mathbf{x}$. A pair $(h^{(i)}, f^{(i)})$ obtained on iteration i is said to *dominate* another pair $(h^{(j)}, f^{(j)})$ if and only if both $h^{(i)} \leq h^{(j)}$ and $f^{(i)} \leq f^{(j)}$, indicating that the point $\mathbf{x}^{(i)}$ is at least as good as $\mathbf{x}^{(j)}$ in respect of both measures. The NLP filter is defined to be a list of pairs $(h^{(i)}, f^{(i)})$ such that no pair dominates any other. This is illustrated by the solid lines in Figure 1. We use $\mathcal{F}^{(k)}$ to denotes the set of iteration indices j ($j < k$) such that $(h^{(j)}, f^{(j)})$ is an entry in the current filter. (In practice we do not need to store the index set $\mathcal{F}^{(k)}$, the notation is just for theoretical convenience.) A point $\mathbf{x}$ is said to be "acceptable for inclusion in the filter" if its (h, f) pair is not dominated by any entry in the filter. This is the condition that

$$\text{either} \qquad h < h^{(j)} \qquad \text{or} \qquad f < f^{(j)} \tag{2.4}$$

for all $j \in \mathcal{F}^{(k)}$. We may also wish to "include a point $\mathbf{x}$ in the filter", by which we mean that its (h, f) pair is added to the list of pairs in the filter, and any pairs in the filter that are dominated by the new pair are removed. We use the filter as an alternative to a penalty function as a means of deciding whether or not to accept a new point in an NLP algorithm.

In fact this definition of a filter is not adequate for proving convergence as it allows points to accumulate in the neighbourhood of a filter entry that has $h^{(i)} > 0$. This is readily corrected by defining a small envelope around the current filter in which points are not accepted. This idea is suggested in the original paper of Fletcher and Leyffer [5]. A similar acceptability test is analysed by Fletcher, Leyffer and Toint [6] in proving global convergence of an SLP–filter algorithm. This is the condition that a point is acceptable to the filter if its (h, f) pair satisfies

$$\text{either} \qquad h \leq \beta h^{(j)} \qquad \text{or} \qquad f \leq f^{(j)} - \gamma h^{(j)} \tag{2.5}$$

for all $j \in \mathcal{F}^{(k)}$, where β and γ are preset parameters such that $1 > \beta > \gamma > 0$, with β close to 1 and γ close to zero. Because $1 - \beta$ and γ are very small, there is negligible difference in practice between (2.5) and (2.4).

In fact, it has more recently become apparent that a slightly different form of the acceptability test, due to Chin and Fletcher [2], allows stronger convergence results to be proved, and it is this that we analyse here. In this test a pair (h, f) is acceptable if

$$\text{either} \qquad h \leq \beta h^{(j)} \qquad \text{or} \qquad f + \gamma h \leq f^{(j)} \tag{2.6}$$

for all $j \in \mathcal{F}^{(k)}$. This *slanting envelope test* ensures that pairs with the same f value have the same envelope in the f direction. This is illustrated in Figure 1, using the values $\gamma = 0.1$ and $\beta = 1 - \gamma$, although in practice a value of γ much closer to zero would be used. (Typical values that we have used are $\gamma = 10^{-5}$ and $\beta = 1 - \gamma$.) The test provides an important *inclusion property* that if a pair (h, f) is added to the filter, then the set of unacceptable points for the new filter always includes the set of unacceptable points for the old filter. This is not always the case for (2.5).

The left hand inequality in (2.6) and also in (2.5) is an obvious way of defining a sufficient reduction in h. The right hand inequality in (2.6) asks for a sufficient reduction in f, defined in such a way that it provides a mechanism whereby iterates are forced towards feasibility. This is shown in the following lemma and its corollary.

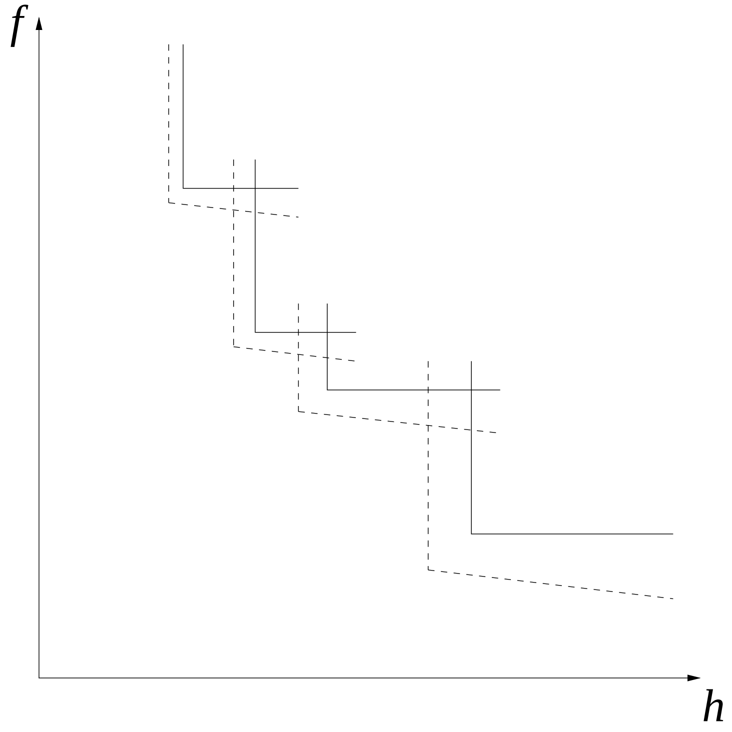

Figure 1: An NLP Filter with Slanting Envelope

Lemma 1 *Consider sequences* $\{h^{(k)}\}$ *and* $\{f^{(k)}\}$ *such that* $h^{(k)} \geq 0$ *and* $f^{(k)}$ *is monotonically decreasing and bounded below. Let constants* β *and* γ *satisfy* $0 < \gamma < \beta < 1$*. If, for all* k*,*

$$\text{either} \qquad h^{(k+1)} \leq \beta h^{(k)} \qquad \text{or} \qquad f^{(k)} - f^{(k+1)} \geq \gamma h^{(k+1)},$$

then $h^{(k)} \to 0$*.*

Proof If $h^{(k+1)} \leq \beta h^{(k)}$ for all k sufficiently large, then $h^{(k)} \to 0$. Otherwise there exists an infinite subsequence $\mathcal{S}$ on which $f^{(k)} - f^{(k+1)} \geq \gamma h^{(k+1)}$. Because $f^{(k)}$ is monotonically decreasing and bounded below, it follows that $\sum_{k \in \mathcal{S}} h^{(k+1)}$ is bounded, and hence $h^{(k+1)} \to 0$ for $k \in \mathcal{S}$. But $h^{(k+1)} \leq \beta h^{(k)}$ holds for iterations $k \notin \mathcal{S}$, so it follows that $h^{(k)} \to 0$ on the main sequence. *q.e.d.*

Corollary *Consider an infinite sequence of iterations on which* $(h^{(k)}, f^{(k)})$ *is entered into the filter, where* $h^{(k)} > 0$ *and* $\{f^{(k)}\}$ *is bounded below. It follows that* $h^{(k)} \to 0$*.*

Proof Consider the set $\mathcal{K} = \{k \mid f^{(j)} > f^{(k)} \ \forall \ j > k\}$. For any $k \in \mathcal{K}$, it follows by filter acceptability that $h^{(j)} \leq \beta h^{(k)}$ for all $j > k$. If $\mathcal{K}$ contains an infinite number of entries, then any $k \in \mathcal{K}$ has a successor $k^+ \in \mathcal{K}$ which is the least $j > k$ such that $j \in \mathcal{K}$. It follows by filter acceptability that $h^{(k^+)} \leq \beta h^{(k)}$. Hence $h^{(k)} \to 0$ for $k \in \mathcal{K}$. But any intermediate iterations j such that $k < j < k^+$ also have the property that $h^{(j)} \leq \beta h^{(k)}$ so it follows that $h^{(k)} \to 0$ on the main sequence.

Conversely, if $\mathcal{K}$ is not infinite, we let K be the largest index in $\mathcal{K}$ (or $K = 0$ if $\mathcal{K}$ is empty). Because $\mathcal{K}$ is finite, for any $k > K$ there exist indices $j > k$ for which $f^{(j)} \leq f^{(k)}$. Thus we may define an infinite subsequence $\mathcal{S}$ as follows. The first index in $\mathcal{S}$ is $k = K + 1$, and for any $k \in \mathcal{S}$ its successor $k^+ \in \mathcal{S}$ is the least value of $j > k$ such that $f^{(j)} \leq f^{(k)}$. Thus $\{f^{(k)}\}$ is monotonic decreasing for $k \in \mathcal{S}$. Also it is a consequence of the inclusion property that $(h^{(k^+)}, f^{(k^+)})$ is acceptable to $(h^{(k)}, f^{(k)})$, even if the latter pair has been deleted from the filter on an intermediate iteration. Hence the conditions of Lemma 1 are satisfied, so that $h^{(k)} \to 0$ for $k \in \mathcal{S}$. Moreover, it also follows by filter acceptability that $h^{(j)} \leq \beta h^{(k)}$ for all j such that $k < j < k^+$. Hence $h^{(k)} \to 0$ on the main sequence. *q.e.d.*

It is also convenient to allow an upper bound

$$h(\mathbf{c}(\mathbf{x})) \leq \beta u \tag{2.7}$$

($u > 0$) on constraint infeasibility, and this is readily implemented by initializing the filter with the entry $(u, -\infty)$. Existence of this upper bound is not necessary to the proof of convergence, but is a useful practical feature that can be used to prevent iterates from becoming too infeasible. In practice we have set a large default value of $u = 10^4$, which usually has negligible impact on performace, but there are a few problems for which a much smaller value is desirable, say $u = 1$.

A common feature in a trust region algorithm for unconstrained minimization is the use of a sufficient reduction criterion

$$\Delta f \geq \sigma \Delta q, \tag{2.8}$$

where Δq is positive, and $\sigma \in (0, 1)$ is a preset parameter. However, in an NLP algorithm, Δq may be negative or even zero, in which case this test is no longer appropriate. A feature of the algorithm in this paper is that it uses (2.8) only when Δq is positive. A typical value of σ that we have used is $\sigma = 0.1$.

We are now in a position to state our filter–SQP algorithm, which we do by means of the flow diagram of Figure 2. We observe that at the start of iteration k, the pair $(h^{(k)}, f^{(k)})$ is not in the current filter $\mathcal{F}^{(k)}$ but must be acceptable to it. It can be seen that there is an inner loop in which the trust region radius ρ is successively reduced until either certain tests are satisfied, or the current QP subproblem becomes incompatible (for clarity we avoid the use of the word 'infeasible' in this context). The inner loop is initialized with any value of $\rho \geq \rho^\circ$, where $\rho^\circ > 0$ is a preset parameter. The inner loop chooses a decreasing geometric sequence of values of ρ and generates corresponding values of $\mathbf{d}$, Δq and Δf (unsubscripted). The inner loop contains a test "is $\mathbf{x}^{(k)} + \mathbf{d}$ acceptable to the filter and $(h^{(k)}, f^{(k)})$". By this we mean that $\mathbf{x}^{(k)} + \mathbf{d}$ has to be acceptable to the filter formed of the current filter and $(h^{(k)}, f^{(k)})$, so that if $(h^{(k)}, f^{(k)})$ is subsequently entered into the filter, then $(h^{(k+1)}, f^{(k+1)})$ will still be acceptable to the new filter. When the inner iteration terminates, the current values of ρ, $\mathbf{d}$, Δq and Δf are denoted respectively by

$\rho^{(k)}$, $\mathbf{d}^{(k)}$, $\Delta q^{(k)}$ and $\Delta f^{(k)}$. We observe that all points that are generated by the algorithm lie in the region generated by the subset of linear constraints in the NLP problem.

Following our multiobjective thinking, we regard a step $\mathbf{d}$ that satisfies $\Delta q > 0$ as being an *f–type step* (having the primary aim of improving f, and possibly allowing an increase in h). If $\mathbf{d}$ is accepted and becomes $\mathbf{d}^{(k)}$, then an *f–type iteration* is said to have occurred. In this case we insist that the sufficient reduction condition (2.8) is satisfied. Thus a necessary condition for a step $\mathbf{d}$ to give rise to an f–type iteration is that both

$$\Delta f \geq \sigma \Delta q \qquad \text{and} \qquad \Delta q > 0 \tag{2.9}$$

are satisfied. If $\Delta q^{(k)} \leq 0$, or if the current QP subproblem is incompatible, then the primary aim of the iteration is to reduce h (possibly allowing an increase in f) and we refer to the resulting iteration as an *h–type iteration*. As ρ is reduced in the inner loop, the value of Δq is reduced (a consequence of having found a global minimizer of $QP(\mathbf{x}^{(k)}, \rho)$). Thus the status of the test $\Delta q > 0$ may go from true to false, but not vice-versa. Consequently, the inner loop always samples the possibility for an f–type iteration before that of an h–type iteration. This is a key argument in the convergence proof.

This algorithm differs in one important respect from that in [5] in that not all points $\mathbf{x}^{(k)}$ are included in the filter, even though they are acceptable to the filter. The point $\mathbf{x}^{(k)}$ is included in the filter at the end of the iteration if and only if that iteration is an h–type iteration. A consequence is that all the current filter entries have $h^{(j)} > 0$, $j \in \mathcal{F}^{(k)}$. This is because if $h^{(k)} = 0$ then $QP(\mathbf{x}^{(k)}, \rho)$ must be compatible and hence, if $\mathbf{x}^{(k)}$ is not a KT point, then $\Delta q > 0$ holds. Thus if $h^{(k)} = 0$, the resulting iteration is an f-type iteration and $\mathbf{x}^{(k)}$ is not entered into the filter. It is convenient to denote

$$\tau^{(k)} = \min_{j \in \mathcal{F}^{(k)}} h^{(j)} > 0. \tag{2.10}$$

It can be seen that our algorithm includes the provision for a feasibility restoration phase if the current QP subproblem becomes incompatible. Any method for solving a nonlinear algebraic system of inequalities can be used to implement this calculation, such as for example a Newton-like scheme for minimizing $h(\mathbf{c}(\mathbf{x}))$. The restoration phase terminates if it finds a point that is both acceptable to the filter, and for which $QP(\mathbf{x}, \rho)$ is compatible for some $\rho \geq \rho^{\circ}$. (Essentially the latter condition only requires that $QP(\mathbf{x}^{(k)}, \infty)$ is compatible, since we can always take $\rho = \infty$.) There are various existing algorithms that might be used to implement this calculation: that of Madsen [12] (with suitable changes to include inequality constraints) has a convergence proof and is close to the spirit of this paper. Alternatively we can make use of the ideas expressed in [5] which have performed well in practice. Note that the restoration phase makes no demands on the resulting value of $f(\mathbf{x})$, which could be significantly worse than that at the previous point. If the restoration phase does terminate, then the point of termination becomes $\mathbf{x}^{(k+1)}$ and the resulting step from $\mathbf{x}^{(k)}$ to $\mathbf{x}^{(k+1)}$ is deemed to be an h–type iteration.

Of course it may not always be possible to find a point which satisfies both the above conditions, and the restoration phase might converge to an infeasible point, for example if there exists a non–zero local minimum of $h(\mathbf{c}(\mathbf{x}))$. This is often an indication that the original problem P is incompatible. This is the situation typified by case ($\mathbb{A}$) of Theorem 1 that follows in the next section. If, on the other hand, the restoration phase is converging to a feasible point, then it is usually able to terminate. This is so because $QP(\mathbf{x}, \infty)$ is usually compatible if $\mathbf{x}$ is sufficiently close to the feasible region, and because $\tau^{(k)} > 0$ allows such a point to be acceptable to the filter. However this outcome is not guaranteed, as it is possible for $QP(\mathbf{x}, \infty)$ to be incompatible for any infeasible point $\mathbf{x}$. Such an example is the pathological problem $\min(x_2 - 1)^2$ subject to $x_1^2 = 0$ and $x_1^3 = 0$, starting from $\mathbf{x} = (1, 0)^T$. A Newton-like iteration for feasibility restoration is likely to converge to the feasible point $\mathbf{x} = \mathbf{0}$, which is not a solution of the NLP, without finding a point at which the QP subproblem is compatible. However such a pathological problem (P) has the property that there exists an arbitrary small perturbation to P for which P is incompatible. Thus in this paper we content ourselves with the possibility that the restoration phase may fail to terminate, and regard this as an indication that the constraints of P are incompatible (in a local sense) to within round-off error.

3 A Global Convergence Proof

In this section we present a proof of global convergence of the SQP–filter algorithm of Figure 2 when applied to problem P. We make the following assumptions.

Standard Assumptions

1. All points $\mathbf{x}$ that are sampled by the algorithm lie in a non-empty closed and bounded set X.

2. The problem functions $f(\mathbf{x})$ and $\mathbf{c}(\mathbf{x})$ are twice continuously differentiable on an open set containing X.

3. There exists an $M > 0$ such that the Hessian matrices $B^{(k)}$ satisfy $\|B^{(k)}\|_2 \leq M$ for all k.

It is a consequence of the standard assumptions that the Hessian matrices of f and the c_i are bounded on X and without loss of generality we may assume that they also satisfy bounds $\|\nabla^2 f(\mathbf{x})\|_2 \leq M$, $\|\nabla^2 c_i(\mathbf{x})\|_2 \leq M$, $i \in \mathcal{E} \cup \mathcal{I}$, for all $\mathbf{x} \in X$.

Our global convergence theorem concerns Kuhn–Tucker (KT) necessary conditions under a Mangasarian–Fromowitz constraint qualification (MFCQ), (see for example, Mangasarian [9]). This is essentially an extended form of the Fritz-John conditions for a problem that includes equality constraints. A feasible point $\mathbf{x}^\circ$ of problem P satisfies MFCQ if and only if both (i) the vectors $\mathbf{a}_i^\circ$, $i \in \mathcal{E}$ are linearly independent, and (ii) there exists a vector $\mathbf{s}$ that satisfies $\mathbf{s}^T\mathbf{a}_i^\circ = 0$, $i \in \mathcal{E}$ and $\mathbf{s}^T\mathbf{a}_i^\circ < 0$, $i \in \mathcal{A}^\circ$, where $\mathcal{A}^\circ \subset \mathcal{I}$ denotes the set of active inequality constraints at $\mathbf{x}^\circ$. Necessary conditions for $\mathbf{x}^\circ$ to solve

P are that $\mathbf{x}^\circ$ is a feasible point, and, if MFCQ holds, then the set of directions

$$\{\mathbf{s} \ | \ \mathbf{s}^T\mathbf{g}^\circ < 0 \tag{3.1}$$

$$\mathbf{s}^T\mathbf{a}_i^\circ = 0 \quad i \in \mathcal{E} \tag{3.2}$$

$$\mathbf{s}^T\mathbf{a}_i^\circ < 0 \quad i \in \mathcal{A}^\circ \ \} \tag{3.3}$$

is empty. If $\mathbf{x}^\circ$ solves P, and MFCQ holds, then these conditions are equivalent to the existence of KT multipliers (although we do not use that result in this paper), and it has been shown (Gauvin [8]) that the multiplier set is bounded.

Before proving our main theorem we need some results that describe the behaviour of QP subproblems in the neighbourhood of a feasible point $\mathbf{x}^\circ$ at which the vectors $\mathbf{a}_i^\circ$, $i \in \mathcal{E}$ are linearly independent. First however we prove two simple lemmas that enable us to handle the second order terms in the analysis.

Lemma 2 *Consider minimizing a quadratic function $\phi(\alpha)$ ($\mathbb{R} \to \mathbb{R}$) on the interval $\alpha \in [0,1]$, when $\phi'(0) < 0$. A necessary and sufficient condition for the minimizer to be at $\alpha = 1$ is $\phi'' + \phi'(0) \leq 0$. In this case it follows that $\phi(0) - \phi(1) \geq -\frac{1}{2}\phi'(0)$.*

Proof Using $\phi(\alpha) = \phi(0) + \alpha\phi'(0) + \frac{1}{2}\alpha^2\phi''$, the minimizer is at $\alpha = 1$ either if $\phi'' \leq 0$ or if $\phi'' > 0$ and $-\phi'(0)/\phi'' \geq 1$, from which the result follows. *q.e.d.*

Lemma 3 *Let the standard assumptions hold, and let $\mathbf{d}$ be a feasible point of $QP(\mathbf{x}^{(k)}, \rho)$. It then follows that*

$$\Delta f \geq \Delta q - n\rho^2 M, \tag{3.4}$$

$$|c_i(\mathbf{x}^{(k)} + \mathbf{d})| \leq \tfrac{1}{2}n\rho^2 M \qquad i \in \mathcal{E}, \tag{3.5}$$

and

$$c_i(\mathbf{x}^{(k)} + \mathbf{d}) \leq \tfrac{1}{2}n\rho^2 M \qquad i \in \mathcal{I} \tag{3.6}$$

Proof These results follow from the intermediate value form of Taylor's theorem, for example

$$f(\mathbf{x}^{(k)} + \mathbf{d}) = f^{(k)} + \mathbf{g}^{(k)T}\mathbf{d} + \tfrac{1}{2}\mathbf{d}^T\nabla^2 f(\mathbf{y})\mathbf{d}$$

where $\mathbf{y}$ denotes some point on the line segment from $\mathbf{x}^{(k)}$ to $\mathbf{x}^{(k)} + \mathbf{d}$. It follows from (2.2) and (2.1) that

$$\Delta f = \Delta q + \tfrac{1}{2}\mathbf{d}^T(B^{(k)} - \nabla^2 f(\mathbf{y}))\mathbf{d},$$

and (3.4) follows from the Hessian bounds and the inequality $\|\mathbf{d}\|_2^2 \leq n\|\mathbf{d}\|_\infty^2 \leq n\rho^2$. Also, for $i \in \mathcal{I}$, it follows that

$$c_i(\mathbf{x}^{(k)} + \mathbf{d}) = c_i^{(k)} + \mathbf{a}_i^{(k)T}\mathbf{d} + \tfrac{1}{2}\mathbf{d}^T\nabla^2 c_i(\mathbf{y}_i)\mathbf{d} \leq \tfrac{1}{2}\mathbf{d}^T\nabla^2 c_i(\mathbf{y}_i)\mathbf{d}$$

by feasibility of $\mathbf{d}$, and (3.6) then follows in a similar way. The result (3.5) follows for $i \in \mathcal{E}$ by regarding an equation as two opposed inequality constraints. *q.e.d.*

Lemma 4 *Let standard assumptions hold. If* $\mathbf{d}$ *solves* $QP(\mathbf{x}^{(k)}, \rho)$, *then* $\mathbf{x}^{(k)} + \mathbf{d}$ *is acceptable to the filter if* $\rho^2 \leq 2\beta\tau^{(k)}/(mnM)$.

Proof It follows from (2.3), (3.5) and (3.6) that $h(\mathbf{c}(\mathbf{x}^{(k)} + \mathbf{d})) \leq \frac{1}{2} mn\rho^2 M$. If $\rho^2 \leq 2\beta\tau^{(k)}/(mnM)$, it then follows that $h(\mathbf{c}(\mathbf{x}^{(k)} + \mathbf{d})) \leq \beta\tau^{(k)}$. Hence, by definition of $\tau^{(k)}$, the filter acceptance test (2.6) is satisfied. *q.e.d.*

Lemma 5 *Let standard assumptions hold and let* $\mathbf{x}^\circ \in X$ *be a feasible point of problem* P *at which MFCQ holds, but which is not a KT point. Then there exists a neighbourhood* $\mathcal{N}^\circ$ *of* $\mathbf{x}^\circ$ *and positive constants* ε, μ *and* κ *such that for all* $\mathbf{x} \in \mathcal{N}^\circ \cap X$ *and all* ρ *for which*

$$\mu h(\mathbf{c}(\mathbf{x}))) \leq \rho \leq \kappa, \tag{3.7}$$

it follows that $QP(\mathbf{x}, \rho)$ *has a feasible solution* $\mathbf{d}$ *at which the predicted reduction (2.1) satisfies*

$$\Delta q \geq \tfrac{1}{3}\rho\varepsilon, \tag{3.8}$$

the sufficient reduction condition (2.8) holds, and the actual reduction (2.2) satisfies

$$\Delta f \geq \gamma h(\mathbf{c}(\mathbf{x} + \mathbf{d})). \tag{3.9}$$

Proof Since $\mathbf{x}^\circ$ is a feasible point at which MFCQ holds, but not a KT point, it follows that the vectors $\mathbf{a}_i^\circ$, $i \in \mathcal{E}$ are linearly independent, and there exists a vector $\mathbf{s}^\circ$ for which $\|\mathbf{s}^\circ\|_2 = 1$ that satisfies (3.1), (3.2) and (3.3). We note that these conditions imply that the cardinality $|\mathcal{E}| < n$. We use the notation $A^+ = (A^T A)^{-1} A^T$ and let $A_\mathcal{E}$ denote the matrix with columns $\mathbf{a}_i$, $i \in \mathcal{E}$, evaluated at some point $\mathbf{x}$. By linear independence and continuity there exists a neighbourhood of $\mathbf{x}^\circ$ in which $A_\mathcal{E}^+$ is bounded. If $\mathcal{E}$ is not empty, we denote $\mathbf{p} = -A_\mathcal{E}^{+T}\mathbf{c}_\mathcal{E}$, which is the closest point in the linearized equality constraint manifold to $\mathbf{d} = \mathbf{0}$, and let $p = \|\mathbf{p}\|_2$. Also we denote $\mathbf{s} = (I - A_\mathcal{E} A_\mathcal{E}^+)\mathbf{s}^\circ / \|(I - A_\mathcal{E} A_\mathcal{E}^+)\mathbf{s}^\circ\|_2$, which is the closest unit vector to $\mathbf{s}^\circ$ in the null space of $A_\mathcal{E}^T$. If $\mathcal{E}$ is empty, we let $\mathbf{p} = \mathbf{0}$, $p = 0$ and $\mathbf{s} = \mathbf{s}^\circ$. It follows from (3.1) and (3.3) by continuity that there exists a (smaller) neighbourhood $\mathcal{N}^\circ$ and a constant $\varepsilon > 0$ such that

$$\mathbf{s}^T\mathbf{g} \leq -\varepsilon \qquad \text{and} \qquad \mathbf{s}^T\mathbf{a}_i \leq -\varepsilon, \quad i \in \mathcal{A}^\circ \tag{3.10}$$

when $\mathbf{g}$, $\mathbf{a}_i$ and $\mathbf{s}$ are evaluated for any $\mathbf{x} \in \mathcal{N}^\circ$. By definition of $\mathbf{p}$, it follows that $p = O(h(\mathbf{c}))$ so we can choose the constant μ in (3.7) sufficiently large so that $\rho > p$ for all $\mathbf{x} \in \mathcal{N}^\circ$.

We now consider the solution of $QP(\mathbf{x},\rho)$, and in particular the line segment defined by

$$\mathbf{d}_\alpha = \mathbf{p} + \alpha(\rho - p)\mathbf{s}, \quad \alpha \in [0,1], \tag{3.11}$$

for a fixed value of $\rho > p$. We note that $\mathbf{d}_\alpha$ satisfies the equality constraints $\mathbf{c}_{\mathcal{E}} + A_{\mathcal{E}}^T\mathbf{d} = \mathbf{0}$ of $QP(\mathbf{x},\rho)$ for any value of α. Because the vectors $\mathbf{p}$ and $\mathbf{s}$ are orthogonal, it follows that

$$\|\mathbf{d}_1\|_2 = \sqrt{p^2 + (\rho - p)^2} = \sqrt{\rho^2 - 2\rho p + 2p^2} \le \rho$$

since $\rho > p$. Consequently $\|\mathbf{d}_1\|_\infty \le \rho$, and hence $\mathbf{d}_1$ satisfies the trust region constraint of $QP(\mathbf{x},\rho)$.

Next we look at inactive constraints $i \in \mathcal{I}/\mathcal{A}^\circ$. If $\mathbf{x} \in \mathcal{N}^\circ \cap X$ then there exist positive constants $\bar{c}$ and $\bar{a}$, independent of ρ, such that

$$c_i \le -\bar{c} \qquad \text{and} \qquad \mathbf{a}_i^T\mathbf{s} \le \bar{a}$$

for all vectors $\mathbf{s}$ such that $\|\mathbf{s}\|_\infty \le 1$, by continuity of c_i and boundedness of $\mathbf{a}_i$ on X. It follows that

$$c_i + \mathbf{a}_i^T\mathbf{d} \le -\bar{c} + \rho\bar{a} \qquad i \in \mathcal{I}/\mathcal{A}^\circ$$

for all vectors $\mathbf{d}$ such that $\|\mathbf{d}\|_\infty \le \rho$. Thus inactive constraints do not affect the solution to $QP(\mathbf{x},\rho)$ if ρ satisfies $\rho \le \bar{c}/\bar{a}$.

For active inequality constraints $i \in \mathcal{A}^\circ$, we have from (3.10) and (3.11) that

$$c_i + \mathbf{a}_i^T\mathbf{d}_1 = c_i + \mathbf{a}_i^T\mathbf{p} + (\rho - p)\mathbf{a}_i^T\mathbf{s} \le c_i + \mathbf{a}_i^T\mathbf{p} - (\rho - p)\varepsilon \le 0$$

if

$$\rho \ge p + (c_i + \mathbf{a}_i^T\mathbf{p})/\varepsilon.$$

By definition of $\mathbf{p}$, the right hand side of this inequality is $O(h(\mathbf{c}))$ so we can choose the constant μ in (3.7) sufficiently large so that $c_i + \mathbf{a}_i^T\mathbf{d}_1 \le 0$, $i \in \mathcal{A}^\circ$. Thus $\mathbf{d}_1$ is feasible in $QP(\mathbf{x},\rho)$ with respect to the active inequality constraints, and hence to all the constraints, using results from above. Hence we have shown that $QP(\mathbf{x},\rho)$ is compatible for all $\mathbf{x} \in \mathcal{N}^\circ$ and all ρ satisfying (3.7) for any value of $\kappa \le \bar{c}/\bar{a}$.

Next we aim to obtain a bound on the predicted reduction Δq and hence show that (3.8), (2.8) and (3.9) hold. First we consider the line segment (3.11) and define $\phi(\alpha) = q(\mathbf{p} + \alpha(\rho - p)\mathbf{s})$. It follows that

$$\phi'(\alpha) = (\rho - p)\mathbf{s}^T\nabla q(\mathbf{p} + \alpha(\rho - p)\mathbf{s}) = (\rho - p)\mathbf{s}^T(\mathbf{g} + B(\mathbf{p} + \alpha(\rho - p)\mathbf{s})).$$

Hence, using (3.10), bounds on B and $\mathbf{p}$, and $\rho > p$

$$\phi'(0) = (\rho - p)\mathbf{s}^T(\mathbf{g} + B\mathbf{p}) \le (\rho - p)(\mathbf{s}^T B\mathbf{p} - \varepsilon) \le (\rho - p)(Mp - \varepsilon) < (\rho - p)(M\rho - \varepsilon) \le 0$$

if $\rho \le \varepsilon/M$. Now $\phi'' = (\rho - p)^2\mathbf{s}^T B\mathbf{s} \le (\rho - p)^2 M$ so

$$\phi'' + \phi'(0) \le (\rho - p)^2 M + (\rho - p)(Mp - \varepsilon) = (\rho - p)((\rho - p)M + Mp - \varepsilon) \le 0$$

if $\rho \leq \varepsilon/M$. In this case, applying Lemma 2, the minimum value of $\phi(\alpha)$ occurs at $\alpha = 1$ and the reduction in q satisfies $\phi(0) - \phi(1) \geq -\frac{1}{2}\phi'(0)$. After adding in a contribution for the change in q along $\mathbf{p}$, we may express

$$q(\mathbf{0}) - q(\mathbf{d}_1) \geq \tfrac{1}{2}(\rho - p)(\varepsilon - \mathbf{s}^T B\mathbf{p}) + O(p) \geq \tfrac{1}{2}\rho\varepsilon + O(p).$$

Since $\mathbf{d}_1$ is feasible and $p = O(h(\mathbf{c}))$, it follows that the predicted reduction (2.1) satisfies

$$\Delta q \geq \tfrac{1}{2}\rho\varepsilon + O(h(\mathbf{c})) \geq \tfrac{1}{2}\rho\varepsilon - \xi h(\mathbf{c})$$

for some ξ sufficiently large and independent of ρ. Thus (3.8) is satisfied if $\rho \geq 6\xi h(\mathbf{c})/\varepsilon$. This condition can be achieved by making the constant μ in (3.7) sufficiently large. It follows from (3.4) and (3.8) that

$$\frac{\Delta f}{\Delta q} \geq 1 - \frac{n\rho^2 M}{\Delta q} \geq 1 - \frac{3n\rho^2 M}{\rho\varepsilon} = 1 - \frac{3n\rho M}{\varepsilon}.$$

Then, if $\rho \leq (1-\sigma)\varepsilon/(3nM)$ it follows that (2.8) holds.

Finally, we deduce from (2.3), (3.5), (3.6), (2.8) and (3.8) that

$$f^{(k)} - f - \gamma h(\mathbf{c}(\mathbf{x}+\mathbf{d})) = \Delta f - \gamma h(\mathbf{c}(\mathbf{x}+\mathbf{d})) \geq \tfrac{1}{3}\sigma\rho\varepsilon - \tfrac{1}{2}\gamma mn\rho^2 M \geq 0$$

if $\rho \leq \frac{2}{3}\sigma\varepsilon/(\gamma mnM)$. Thus we may define the constant κ in (3.7) to be the least of $\frac{2}{3}\sigma\varepsilon/(\gamma mnM)$ and the values $(1-\sigma)\varepsilon/(3nM)$, ε/M and $\bar{c}/\bar{a}$, as required earlier in the proof. *q.e.d.*

Now we proceed to analyse the algorithm of Figure 2. First we need a result that is similar to Lemma 2 of [6]. Here $\mathbf{x}^{(k)}$ and $B^{(k)}$ are fixed and we consider what happens to the solution of $QP(\mathbf{x}^{(k)}, \rho)$ as ρ is reduced.

Lemma 6 *Let the standard assumptions hold, then the inner iteration terminates finitely.*

Proof If $\mathbf{x}^{(k)}$ is a KT point of problem P then $\mathbf{d} = \mathbf{0}$ solves $QP(\mathbf{x}^{(k)}, \rho)$ and the algorithm terminates. Otherwise, if the inner iteration does not terminate finitely then the rule for decreasing ρ ensures that $\rho \to 0$. Two cases need to be considered, depending on whether $h^{(k)} > 0$ or $h^{(k)} = 0$.

If $h^{(k)} > 0$ and $i \in \mathcal{E} \cup \mathcal{I}$ is an index for which $c_i^{(k)} > 0$ then for all $\mathbf{d}$ such that $\|\mathbf{d}\|_\infty \leq \rho$ it follows that

$$c_i^{(k)} + \mathbf{a}_i^{(k)T}\mathbf{d} \geq c_i^{(k)} - \rho\|\mathbf{a}_i^{(k)}\|_1 > 0$$

if either $\|\mathbf{a}_i^{(k)}\|_1 = 0$ or $\rho < c_i^{(k)}/\|\mathbf{a}_i^{(k)}\|_1$. Thus for sufficiently small ρ, constraint i cannot be satisfied and $QP(\mathbf{x}^{(k)}, \rho)$ is incompatible. A similar conclusion obtains for $i \in \mathcal{E}$ if $c_i^{(k)} < 0$. Thus the inner iteration terminates finitely if $h^{(k)} > 0$.

If $h^{(k)} = 0$, then by a similar argument, inactive constraints at $\mathbf{x}^{(k)}$ are inactive at any point for which $\|\mathbf{d}\|_\infty \le \rho$, for sufficiently small ρ. Thus we need only consider constraints $i \in \mathcal{E} \cup \mathcal{A}^{(k)}$. The rest of the proof is now similar to that of Lemma 5 in the case $p = 0$. Because $\mathbf{x}^{(k)}$ is not a KT point, there exists a vector $\mathbf{s}$, $\|\mathbf{s}\|_2 = 1$, and an $\eta > 0$ such that $\mathbf{s}^T\mathbf{g}^{(k)} = -\eta$, $\mathbf{s}^T\mathbf{a}_i^{(k)} = 0$, $i \in \mathcal{E}$, and $\mathbf{s}^T\mathbf{a}_i^{(k)} \le 0$, $i \in \mathcal{A}^{(k)}$. We consider the QP-feasible line segment $\mathbf{d}_\alpha = \alpha\rho\mathbf{s}$ for $\alpha \in [0,1]$, and construct the function $\phi(\alpha) = q(\mathbf{d}_\alpha)$. It follows that $\phi'(0) = -\rho\eta$ and $\phi'' = \rho^2\mathbf{s}^T B^{(k)}\mathbf{s} \le \rho^2 M$. Hence if $\rho \le \eta/M$, it follows that $\phi'' + \phi'(0) \le 0$. It then follows from Lemma 2 that $\phi(0) - \phi(1) \ge \frac{1}{2}\rho\eta$. Therefore, by global optimality of the solution $\mathbf{d}$ to $QP(\mathbf{x}^{(k)}, \rho)$, the actual reduction Δq also satisfies $\Delta q \ge \frac{1}{2}\rho\eta$, and if $\rho \le (1-\sigma)\eta/(2nM)$, it follows from (3.4) that $\Delta f \ge \sigma\Delta q > 0$ and the necessary condition (2.9) for an f–type iteration is satisfied. Also, from (3.4) (3.5) and (3.6),

$$f^{(k)} - f(\mathbf{x}^{(k)} + \mathbf{d}) - \gamma h(\mathbf{c}(\mathbf{x}^{(k)} + \mathbf{d})) = \Delta f - \gamma h(\mathbf{c}(\mathbf{x}^{(k)} + \mathbf{d})) \ge \tfrac{1}{2}\sigma\rho\eta - \tfrac{1}{2}\gamma m n \rho^2 M \ge 0$$

if $\rho \le \sigma\eta/(\gamma m n M)$. In this case it follows that $\mathbf{x}^{(k)} + \mathbf{d}$ is acceptable relative to $(h^{(k)}, f^{(k)})$. Finally, from Lemma 4, $\mathbf{x}^{(k)} + \mathbf{d}$ is acceptable to the filter if $\rho^2 \le 2\beta\tau^{(k)}/(mnM)$. Thus, if ρ is sufficiently small, all the conditions for an f–type step are satisfied and the inner iteration terminates finitely. *q.e.d.*

We are now in a position to state our main theorem.

Theorem 1 *If standard assumptions hold, the outcome of applying the filter–SQP algorithm of Figure 2 is one of the following.*

- ($\mathbb{A}$) *The restoration phase fails to find a point* $\mathbf{x}$ *which is both acceptable to the filter and for which* $QP(\mathbf{x}, \rho)$ *is compatible for some* $\rho \ge \rho^\circ$.
- ($\mathbb{B}$) *A KT point of problem P is found* ($\mathbf{d} = \mathbf{0}$ *solves* $QP(\mathbf{x}^{(k)}, \rho)$ *for some* k).
- ($\mathbb{C}$) *There exists an accumulation point that is feasible, and is either a KT point or fails to satisfy MFCQ.*

Proof We need only consider the case in which neither ($\mathbb{A}$) nor ($\mathbb{B}$) occurs. Because the inner loop of each iteration is finite (Lemma 6), the outer iteration sequence indexed by k is infinite. All iterates $\mathbf{x}^{(k)}$ lie in X, which is bounded, so it follows that the sequence has one or more accumulation points.

First, we consider the case that the main sequence contains an infinite number of h–type iterations, and we consider this subsequence. For an h–type iteration, $(h^{(k)}, f^{(k)})$ is always entered into the filter at the completion of the iteration, so it follows from the Corollary to Lemma 1 that $h^{(k)} \to 0$ on this subsequence. It must also follow that $\tau^{(k)} \to 0$. Moreover, only h–type iterations can reset $\tau^{(k)}$, so there exists a thinner infinite subsequence on which $\tau^{(k+1)} = h^{(k)} < \tau^{(k)}$ is set. Because X is bounded, there exists an accumulation point $\mathbf{x}^\infty$ and a subsequence indexed by $k \in \mathcal{S}$ of h–type iterations for which $\mathbf{x}^{(k)} \to \mathbf{x}^{\infty}$, $h^{(k)} \to 0$ and $\tau^{(k+1)} = h^{(k)} < \tau^{(k)}$. One consequence is that $\mathbf{x}^{\infty}$ is a feasible point. If MFCQ is not satisfied at $\mathbf{x}^{\infty}$, then $\mathbb{C}$ is established in this case. We therefore assume that MFCQ is satisfied and consider the proposition (to be contradicted) that $\mathbf{x}^{\infty}$ is not a KT point. In this case, the vectors $\mathbf{a}_i^{\infty}$, $i \in \mathcal{E}$ are linearly independent, and the set defined by (3.1), (3.2) and (3.3) is not empty. For sufficiently large $k \in \mathcal{S}$ it follows that $\mathbf{x}^{(k)}$ is in the neighbourhood $\mathcal{N}^{\infty}$, as defined in Lemma 5. We show that this leads to a contradiction.

Lemma 5 provides conditions on ρ which ensure that $QP(\mathbf{x}^{(k)}, \rho)$ is compatible, and the resulting step $\mathbf{d}$ satisfies $\Delta f \geq \sigma \Delta q > 0$ and $f^{(k)} \geq f + \gamma h$, where f and h denote $f = f(\mathbf{x}^{(k)} + \mathbf{d})$ and $h = h(\mathbf{c}(\mathbf{x}^{(k)} + \mathbf{d}))$, respectively. This shows that the necessary condition (2.9) for an f–type step is satisfied, and the entry (h, f) is acceptable to (not dominated by) $(h^{(k)}, f^{(k)})$. Moreover, it follows from Lemma 4 that $\mathbf{x}^{(k)} + \mathbf{d}$ is acceptable to the filter if $\rho^2 \leq 2\beta\tau^{(k)}/(mnM)$. Thus we deduce that if ρ satisfies

$$\mu h^{(k)} \leq \rho \leq \min\left\{ \sqrt{\frac{2\beta\tau^{(k)}}{mnM}}, \kappa \right\}, \tag{3.12}$$

then (h, f) satisfies all the conditions for an f–type iteration.

Now we need to show that a value of ρ in this range will be located by the inner iteration. It follows for $k \in \mathcal{S}$ sufficiently large that $\tau^{(k)} \to 0$ and the range (3.12) becomes

$$\mu h^{(k)} \leq \rho \leq \sqrt{\frac{2\beta\tau^{(k)}}{mnM}}. \tag{3.13}$$

In the limit, because $h^{(k)} < \tau^{(k)}$, and because of the square root, the upper bound in (3.13) is more than twice the lower bound. Now consider how the inner loop of the algorithm works. Initially a value $\rho \geq \rho^{\circ}$ is chosen, which in the limit will be greater than the upper bound in (3.13). Then, successively halving ρ in the inner loop will eventually locate a value in the interval (3.13), or to the right of this interval, which provides the conditions for an f–type step to occur. It is not possible for any value of $\rho \geq \mu h^{(k)}$ to produce an h–type step since Δq decreases monotonically as ρ decreases (this is a consequence of the global optimality of $\mathbf{d}$). Thus if $k \in \mathcal{S}$ is sufficiently large, an f–type iteration will result. This contradicts the fact that the subsequence is composed of h–type iterations. Thus $\mathbf{x}^{\infty}$ is a KT point and $\mathbb{C}$ is established in this case.

Next we consider the alternative case that the main sequence contains only a finite number of h–type iterations. Hence there exists an index K such that all iterations are f–type iterations for all $k \geq K$. It follows that $(h^{(k+1)}, f^{(k+1)})$ is always acceptable to $(h^{(k)}, f^{(k)})$, and also that $\Delta f^{(k)} \geq \sigma \Delta q^{(k)} > 0$, so that the sequence of function values $\{f^{(k)}\}$ is strictly monotonically decreasing for $k \geq K$. It therefore follows from Lemma 1 that $h^{(k)} \to 0$, and hence that any accumulation point $\mathbf{x}^{\infty}$ of the main sequence is a feasible point. Because $f(\mathbf{x})$ is bounded on X it also follows that $\sum_{k \geq K} \Delta f^{(k)}$ is convergent. As above, we now aim to contradict the proposition that there exists an accumulation point at which MFCQ holds that is not a KT point.

Because all iterations $k \geq K$ are f–type, no filter entries are made and so $\tau^{(k)} = \tau^{(K)}$ is constant. For sufficiently large $k \geq K$ it follows that $\mathbf{x}^{(k)}$ is in the neighbourhood $\mathcal{N}^{\infty}$ defined in Lemma 5. It follows as above that sufficient conditions for accepting an f-type step are that ρ satisfies

$$\mu h^{(k)} \leq \rho \leq \min\left\{ \sqrt{\frac{2\beta\tau^{(K)}}{mnM}},\ \kappa \right\}. \tag{3.14}$$

This time the right hand side of (3.14) is a constant, $\bar{\rho}$ say ($\bar{\rho} > 0$) independent of k, whilst the left hand side converges to zero. Thus, for sufficiently large k, the upper bound must be greater than twice the lower bound. In this case, as ρ is reduced in the inner loop, either it must eventually fall within this interval or a value to the right of the interval is accepted. Hence we can guarantee that a value $\rho^{(k)} \geq \min(\frac{1}{2}\bar{\rho},\ \rho^{\circ})$ will be chosen. We then deduce from (2.8) and (3.8) that $\Delta f^{(k)} \geq \frac{1}{3}\sigma\varepsilon \min(\frac{1}{2}\bar{\rho},\ \rho^{\circ})$ which contradicts the fact that $\sum_{k \geq K} \Delta f^{(k)}$ is convergent. Thus $\mathbf{x}^{\infty}$ is a KT point and $\mathbb{C}$ is established in this case also. *q.e.d.*

4 Discussion

Of course, the algorithm of Figure 2 is only a guide to what might be successfully implemented in practice, and is incomplete in various ways. For example, it is necessary to make a specific choice of algorithm to implement the restoration phase. Also the rule for adjusting ρ in the inner iteration could be more intricate, based partly on interpolation. Another possibility is to allow the pair $(h^{(k)}, f^{(k)})$ to be entered into the filter on an f–type step if $h^{(k)} \geq \tau^{(k)}$, as this does not affect the convergence proof. An overall strategic decision is that of how to specify the matrix $B^{(k)}$. One possibility is to use a Lagrangian Hessian based on exact second derivatives and estimates of Lagrange multipliers. A disadvantage of this is that the matrices $B^{(k)}$ may be indefinite, in which case finding the global minimizer of the QP subproblem is problematic. An alternative possibility is to use some quasi-Newton formula to update $B^{(k)}$, in which case it might be possible to ensure that $B^{(k)}$ is positive semi-definite, and hence any KT point of the QP subproblem is a global solution. It is also not easy to prove that $B^{(k)}$ is bounded. However, when MFCQ holds, it can be expected that Lagrange multiplier estimates are bounded, and hence that $B^{(k)}$ is bounded. In practice, the algorithm has been implemented with an exact Hessian with very satisfactory performance, akin to that reported in [5]. Preliminary practical experience with a quasi-Newton form of the algorithm is also promising. There are other ways in which the potential difficulty of finding the global minimizer of the QP subproblem might be avoided, whilst retaining the rapid convergence normally associated with an SQP algorithm, and some of these are described later in the section.

The choice of an initial value of ρ for the inner iteration requires that the condition $\rho \geq \rho^{\circ}$ is satisfied, but is otherwise unspecific. We envisage that in practice ρ° is close to zero (say 10^{-4}) so that the effect of this restriction is small. Thus to a large extent the algorithm of Figure 2 allows the more usual trust region procedure in which one may double or halve (say) the value of ρ from the previous iteration, only setting $\rho = \rho^\circ$ if it would otherwise be less than ρ°. The potential danger of just taking ρ from the previous iteration is that the existence of a successful f-type step may not be recognised. By starting with $\rho \geq \rho^\circ$, we ensure that ρ is greater than twice the lower bound $\mu h^{(k)}$ in the limit, and hence that an f-type step will be taken if the range allows. Adjusting the trust region in this sort of way has featured in other recent work, see for example [10], [11] and references contained therein.

Another important aspect that we have not addressed in this paper is to consider the asymptotic behaviour of the algorithm to ensure that the second order convergence property of the SQP iteration is not compromised. We have already given some thought to this, but is not yet clear how to make progress. The algorithm in [5] allows the use of a second order correction step, although it is not clear in practice whether this is necessary or even beneficial. We shall continue to study such issues in our future work.

The referees for the paper both make the point that the link between f and $\mathbf{c}$ that is implicit in the second inequality of (2.6) is undesirable. Aesthetically we agree that it would be preferable not to have this link, although we submit that its effect is minimal. We stress that the parameter γ is intended to be close to zero (typically 10^{-5}) so that this inequality is little different in practice to that in (2.4), in which case there is no linkage. We have successfully implemented this type of algorithm in practice, with results of similar quality to those in [5], and changing to $f < f^{(j)}$ causes negligible difference to the outcome. It may well be desirable to take the relative scaling of f and h into account, but this is readily done.

In any event it is by no means clear how to avoid the linkage between f and $\mathbf{c}$. The step $\mathbf{d}$ that solves the QP subproblem is not a descent direction for f when $\Delta q \leq 0$ so we cannot use any analogue of the Goldstein or Wolfe–Powell tests from unconstrained optimization. We feel that our proposals are noteworthy in that they enable a convergence proof to be made in such a way that the linkages between f and h are small and the impact on practical performance is negligible.

The authors of this paper have also contributed to other papers that suggest filter-type algorithms for which global convergence can be proved. One paper uses ideas akin to those suggested by Fletcher and Sainz de la Maza [7], in which an LP trust region subproblem is solved in order to obtain an estimate of the active set, which can then be used in an equality QP calculation to determine a trial step. The theoretical and practical properties of this approach have been investigated by a student C.M. Chin and are reported in Chin and Fletcher [2], [3]. Another approach, suggested by Fletcher, Gould, Leyffer and Toint [4], is a trust region SQP algorithm using a filter, but which allows the use of an approximate solution $\mathbf{d}$ to the QP subproblem. The algorithm is based on a decomposition of the step $\mathbf{d}$ into its normal and tangential components. A proof of global convergence to a first order critical point is given. The proof is significantly different from that in this paper, and provides a different outlook on the problem, more related to the familiar Cauchy point decrease condition that appears elsewhere in the trust region literature (see, for example Conn, Gould and Toint [1]). It is an advantage that the proof allows an approximate solution to the QP subproblem, but there is also a disadvantage that it relies on certain conditions that may require an expensive projection calculation to verify. Also the filter envelope (2.5) is used, rather than the slanting filter envelope (2.6) used in this paper. No practical experience with the Cauchy-type of algorithm is as yet available.

Global convergence proofs for other filter-related algorithms that do not use merit functions have also been set out in recent papers. Ulbrich, Ulbrich and Vicente [14] use a decomposition into normal and tangential components of a primal-dual interior point step, and use a filter to decide on acceptability. The paper of Ulbrich and Ulbrich [13] uses non-monotonic improvement conditions on both the normal and tangential steps and obtains global convergence using an acceptance test based on comparing the normal and tangential predicted reductions with a suitably chosen weighting parameter. Encouraging numerical results of a preliminary MATLAB program on a range of CUTE test problems are presented. Wächter and Biegler [15] described a line search method in which the NLP problem is converted into equations and simple bounds, and a filter is used to balance the contributions of a barrier function for the simple bounds and a constraint violation function for the equations. Both [14] and [15] present additional results relating to second order convergence.

5 Acknowledgement

We gratefully acknowledge the contribution of Ademir Alves Ribeiro of the Federal University of Parana, Brazil, who pointed out an error in the previous proof of the corollary to Lemma 1. This error is corrected in the current paper.

6 References

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

[2] Chin C.M. and Fletcher R. (2001), On the Global Convergence of an SLP–Filter Algorithm that takes EQP Steps, Dundee University, Dept. of Mathematics, Report NA/199.

[3] Chin C.M. and Fletcher R. (2001), Numerical results of SLPSQP, filterSQP and LANCELOT on selected CUTE test problems, Dundee University, Dept. of Mathematics, Report NA/203.

[4] Fletcher R., Gould N.I.M, Leyffer S. and Toint Ph-.L. (1999), Global convergence of trust-region SQP-filter algorithms for nonlinear programming, Tech. Report 99/03, Dept. of Mathematics, University of Namur. (Revised 12th February, 2001).

[5] Fletcher R. and Leyffer S. (1997), Nonlinear Programming Without a Penalty Function, Dundee University, Dept. of Mathematics, Report NA/171, to appear in Mathematical Programming.

[6] Fletcher R., Leyffer S. and Toint Ph-.L. (1999), On the Global Convergence of an SLP–Filter Algorithm, Dundee University, Dept. of Mathematics, Report NA/183.

[7] Fletcher R. and Sainz de la Maza E. (1989), Nonlinear programming and nonsmooth optimization by successive linear programming, *Mathematical Programming*, **43**, pp. 235–256.

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

[9] Mangasarian O.L. (1969), *Nonlinear Programming*, McGraw-Hill, New York.

[10] Jiang H., Fukushima M., Qi L. and Sun D. (1998), A Trust Region Method for Solving Generalized Complementarity Problems, *SIAM J. Optim.*, **8**, pp. 140–158.

[11] Kanzow Ch. and Zupke M. (1997), Inexact Trust-Region Methods for Nonlinear Complementarity Problems, University of Hamburg, Institute of Applied Mathematics, Report Number 127.

[12] Madsen K., An Algorithm for Minimax Solution of Overdetermined Systems of Nonlinear Equations, *J. Inst. Maths Applns.*, **16**, pp. 321–328.

[13] Ulbrich M. and Ulbrich S. (2000), Non-monotone Trust Region Methods for Nonlinear Equality Constrained Optimization without a Penalty Function, Technical Report, Zentrum Mathematik, Technische Universität München.

[14] Ulbrich M., Ulbrich S. and Vicente L.N. (2000), A Globally Convergent Primal-dual Interior-point Filter Method for Nonconvex Nonlinear Programming, Technical Report TR00-12, Department of Computational and Applied Mathematics, Rice University.

[15] Wächter A. and Biegler L.T. (2001), 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.

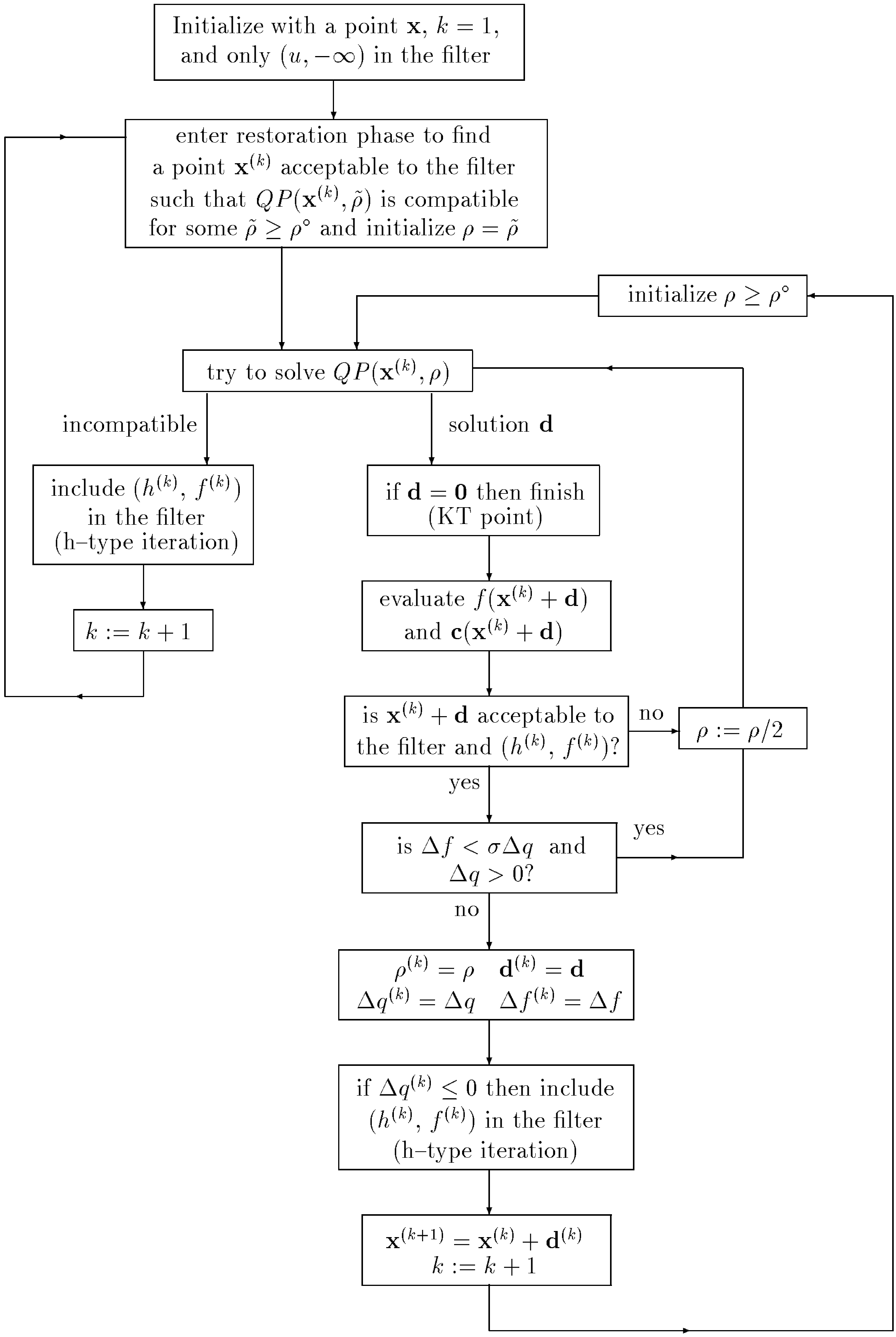

Figure 2: A Filter–SQP Algorithm