On the Global Convergence of an SLP–Filter Algorithm that takes EQP steps

Choong Ming Chin and Roger Fletcher*

Department of Mathematics, University of Dundee, Dundee DD1 4HN, Scotland, UK.

Numerical Analysis Report NA/199, August 1999, revised April 2001

Abstract

A global convergence proof is presented for a class of trust region filter–type methods for nonlinear programming. 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 methods are based on successively solving linear programming subproblems for which effective software is readily available. The methods also permit the use of steps calculated on the basis of an equality constrained quadratic programming model, which enables rapid convergence to take place for problems in which second order information is important.

The proof technique is presented in a fairly general context, allowing a range of specific algorithm choices associated with choosing the quadratic model, updating the trust region radius and with feasibility restoration.

Keywords nonlinear programming, global convergence, filter, multiobjective optimization, SLP, EQP.

1 Introduction

Sequential quadratic programming (SQP) methods are well known to provide a robust and efficient approach to the solution of nonlinear programming (NLP) problems. Moreover, filter methods, first introduced by Fletcher and Leyffer [4], have proved to be an effective way of inducing global convergence. One major drawback, however, is the need to have a robust and efficient QP solver to underpin the method. For very large scale problems, QP solvers can be less efficient, certainly when compared against the performance of linear programming (LP) solvers on problems of similar size. There is also a potential difficulty of finding global solutions of non-convex QP subproblems. Thus it is attractive to consider using sequential linear programming (SLP) methods as a means of solving NLP problems.

Unfortunately SLP methods can prove very slow when the NLP solution is not at a vertex and second order information is required to formulate the optimality conditions. A way of resolving this difficulty in the context of a range of nonsmooth optimization problems (including l_1 penalty methods for NLP) is suggested by Fletcher and Sainz de la Maza [7], and a related approach is used here. Basically the idea is to solve a trust region LP subproblem to obtain an estimate of the active constraints of the NLP problem. Then an equality constrained QP (EQP) problem is solved to determine the step that is taken in the method. The EQP problem only requires the solution of a system of linear equations with special structure and efficient codes exist for this purpose. If the EQP step fails to make progress, then other alternatives such as a second order correction (SOC) step can be tried. If all else fails, then a step based on the SLP solution is tried. Convergence for trust region SLP methods has been established in the report of Fletcher, Leyffer and Toint [5] and their methods have been adapted in this paper to allow the possibility of taking EQP steps.

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. The motivation for this is that it provides certain conditions used in the convergence proof. To a large extent however, the approach allows conventional ideas to be used of halving or doubling (say) the previous trust region radius. The use of a trust region leads to a possible difficulty that the constraints in the trust region LP subproblem may become incompatible as the trust region radius is reduced. We treat this difficulty in the same way as [4] by introducing a feasibility restoration phase into the algorithm.

This paper only deals with the aspects of the method as they relate to global convergence. This allows flexibility in the design of an actual code. Questions such as how to provide a quadratic model, what to do if the reduced Hessian is indefinite, how to implement feasibility restoration, etc. are not considered here. Such questions are addressed in the PhD thesis of Chin [2] where a wide range of numerical evidence on CUTE test problems may be found. It is hoped to publish this work in the near future. These results show that the approach has considerable promise and is well suited to provide the basis of a commercial NLP code.

There exist some other papers, Fletcher, Gould, Leyffer and Toint [3] and Fletcher, Leyffer and Toint [6], that address the issue of global convergence for trust region SQP filter methods. Our paper here has some new features. One useful idea that first features in [2] is that of a *slanting filter*, which enables a stronger statement about convergence to a feasible point to be made. This idea is also taken up in [6]. Another way in which this paper differs from earlier work is in the details of which points are added to the filter. This is related to the unblocking issue described in [4]. These changes, and the need to allow EQP or SOC steps to be taken, cause non-trivial differences to the arguments in earlier proofs.

Section 2 describes the prototype algorithm that we are considering and Section 3 presents the global convergence proof. Section 4 discusses issues that arise about the limitations of the proof and the assumptions on which it is based.

2 An SLP-EQP Filter Algorithm

In this paper we consider an NLP problem 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 refer to a local solution by $\mathbf{x}^*$. We assume for the purposes of our convergence proof that all the points sampled by the algorithm lie in a non–empty bounded region X, which is readily ensured by including simple upper and lower bounds on $\mathbf{x}$ on all the variables. The LP 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

$$LP(\mathbf{x},\rho) \begin{cases} \underset{\mathbf{d}\in\mathbb{R}^n}{\text{minimize}} & \mathbf{g}^T\mathbf{d} \\ \text{subject to} & c_i + \mathbf{a}_i^T\mathbf{d} \leq 0 \quad i = 1, 2, \ldots, m \\ & \|\mathbf{d}\|_\infty \leq \rho, \end{cases}$$

where we denote $\mathbf{g} = \boldsymbol{\nabla} f(\mathbf{x})$, $c_i = c_i(\mathbf{x})$ and $\mathbf{a}_i = \boldsymbol{\nabla} c_i(\mathbf{x})$. The l_∞ norm is used to define the trust region because it is readily implemented by adding simple bounds to the LP subproblem. Let $\mathbf{d}_{LP}$ denote the solution (if it exists) of $LP(\mathbf{x},\rho)$. Then we denote

$$\Delta l = -\mathbf{g}^T\mathbf{d}_{LP} \tag{2.1}$$

as the *LP predicted reduction* in $f(\mathbf{x})$. Our algorithm uses the solution of $LP(\mathbf{x},\rho)$ to identify a set of active constraints (other than the trust region constraint) for problem P. These are used to determine a step $\mathbf{d}_{EQP}$ from an equality constrained QP calculation (the *EQP step*) based on any quadratic model

$$q(\mathbf{d}) = f(\mathbf{x}) + \mathbf{d}^T\mathbf{g}(\mathbf{x}) + \tfrac{1}{2}\mathbf{d}^T B\mathbf{d} \tag{2.2}$$

about the current point $\mathbf{x}$ that is available to us, in which B is some approximate Hessian matrix. We also allow the possibility of taking a *second order correction step* $\mathbf{d}_{SOC}$ (see for example [4]) to help improve the asymptotic properties of the method. For the purposes of this convergence proof, we do not need to be very specific about how these steps are calculated, other than to require that $\mathbf{x}^{(k)} + \mathbf{d} \in X$.

The main idea is to accept either one of these steps if it satisfies certain tests, one of which (if $\Delta l > 0$) is a sufficient reduction criterion in the objective function. If these tests fail, then we fall back on the use of the LP solution. However, as we have second order information available to us, our fall-back position is to use a so-called *Cauchy step* $\mathbf{d}_C = \alpha_C\mathbf{d}_{LP}$ $(0 < \alpha_C \leq 1)$, which is the step along the LP solution vector at which the quadratic model (2.2) is minimized. To be able to calculate α_C, we need the component

$$b = \mathbf{d}_{LP}^T B\mathbf{d}_{LP} \tag{2.3}$$

of the Hessian approximation B in the direction $\mathbf{d}_{LP}$ (see Lemma 3 in Section 3 below).

For any displacement $\mathbf{d}$ we denote

$$\Delta f = f(\mathbf{x}) - f(\mathbf{x} + \mathbf{d}) \tag{2.4}$$

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

$$\Delta q = -\mathbf{g}^T \mathbf{d} - \tfrac{1}{2}\mathbf{d}^T B \mathbf{d} \tag{2.5}$$

as the *quadratic predicted reduction* in $f(\mathbf{x})$. Either of the above may be subscripted with EQP, SOC or C to indicate the type of step being used. Particularly we shall need the value Δq_C calculated at the Cauchy step $\mathbf{d}_C$. Our sufficient reduction condition is that

$$\Delta f \geq \sigma \Delta q_C \tag{2.6}$$

where $\sigma \in (0, 1)$ is a preassigned constant. It is important not to compare Δf with Δl in this test. This is because Δl can be much larger than Δq_C in situations where second order information is significant, in which case it may become impossible for the tests involving $\mathbf{d}_{EQP}$ or $\mathbf{d}_{SOC}$ to succeed.

The measure of constraint infeasibility which we use in this paper is

$$h(\mathbf{c}) = \|\mathbf{c}^+\|_1 \tag{2.7}$$

where $c_i^+ = \max(0, c_i)$, using an obvious notation. 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 $B^{(k)}$.

We now turn to the definition of an NLP filter as introduced in [4]. 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. We use $\mathcal{F}^{(k)}$ to denotes the set of iteration indices j $(j \leq k)$ 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

$$\text{either} \qquad h < h^{(j)} \qquad \text{or} \qquad f < f^{(j)} \tag{2.8}$$

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 our 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 [4]. In our paper, a slightly different way of defining the envelope is used. In our test, the new pair (h, f) must satisfy

$$\text{either} \qquad h \leq \beta h^{(j)} \qquad \text{or} \qquad f + \gamma h \leq f^{(j)} \tag{2.9}$$

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. The outline of this envelope 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. This construction is referred to as the *slanting envelope* and is first used in the work described in Chin [?].

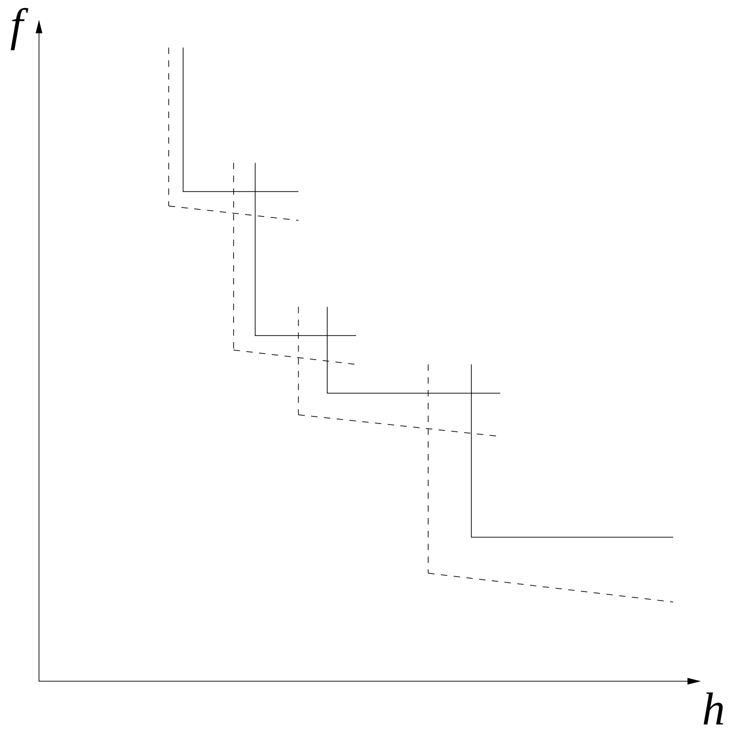

Figure 1: An NLP Filter with Slanting Envelope

The first of the inequalities in (2.9) is just sufficient reduction in h. The dependence on h in the second inequality is an artefact that enables convergence of $\{h^{(k)}\}$ to zero to be proved. Because γ is small the extra term has negligible practical impact. Compared with the test used in [5], the new test has the property that pairs with the same f value have the same envelope in the f direction. This provides an important *inclusion property* that if a new pair 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. The slanting envelope test allows a stronger statement about the convergence of $\{h^{(k)}\}$ to zero to be made, than the envelope test given in [5]. This is expressed in the following lemma.

Lemma 1 *Consider an infinite sequence of iterations on which* $(h^{(k)}, f^{(k)})$ *is entered into the filter, and* $\{f^{(k)}\}$ *is bounded below. It follows that* $h^{(k)} \to 0$.

Proof If $h^{(k+1)} \le \beta h^{(k)}$ for all k sufficiently large, then $h^{(k)} \to 0$. Otherwise we define a subsequence $\mathcal{S}$ as follows. The initial index in $\mathcal{S}$ is the first iteration on which $h^{(k+1)} > \beta h^{(k)}$. For any $k \in \mathcal{S}$, its successor $k^+ \in \mathcal{S}$ is the least $j > k$ such that $h^{(j)} > \beta h^{(k)}$. 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 $f^{(k)} - f^{(k^+)} \ge \gamma h^{(k^+)} > 0$. Thus $f^{(k)}$ is monotonically decreasing for $k \in \mathcal{S}$ and, because $f^{(k)}$ is bounded below, it follows that $\sum_{k \in \mathcal{S}} h^{(k)}$ is bounded above, and hence that $h^{(k)} \to 0$ for $k \in \mathcal{S}$. Moreover, intermediate iterations j such that $k < j < k^+$ have the property that $h^{(j)} \le \beta h^{(k)}$, so it follows that $h^{(k)} \to 0$ on the main sequence. *q.e.d.*

It is also convenient to allow an upper bound

$$h(\mathbf{c}(\mathbf{x})) \le u \tag{2.10}$$

($u > 0$) on constraint infeasibility, and this is readily implemented by initializing the filter with the entry $(u, -\infty)$.

We are now in a position to state our SLP-EQP filter algorithm, which we do by means of the flow diagram of Figure 2. On each major iteration (indexed by k) the trust region radius is initialized with any value of $\rho \ge \rho^\circ$, where $\rho^\circ > 0$ is a preset parameter. If $h^{(k)} > 0$, the current pair $(h^{(k)}, f^{(k)})$ is included in the filter. Within iteration k there is an inner loop in which the trust region radius ρ is successively reduced in geometric progression, and corresponding values of $\mathbf{d}$, Δl, Δq_C and Δf (unsuperscripted), are generated. The inner loop is terminated when either certain tests are satisfied, or the current LP subproblem becomes incompatible. In the latter case, a *feasibility restoration phase* is entered to find a new point $\mathbf{x}^{(k+1)}$. Otherwise, the current values of ρ, $\mathbf{d}$, Δl, Δq_C and Δf on termination are denoted respectively by $\rho^{(k)}$, $\mathbf{d}^{(k)}$, $\Delta l^{(k)}$, $\Delta q_C^{(k)}$ and $\Delta f^{(k)}$, and $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{d}^{(k)}$ is set. We observe that all points that are generated by the algorithm lie in the region X.

Within the inner loop is an innermost loop that allows a finite number of trial steps $\mathbf{d}$ to be evaluated. In our case these trial steps are (i) $\mathbf{d}_{EQP}$, (ii) $\mathbf{d}_{SOC}$, and (iii) $\mathbf{d}_C$. The aim of (i) and (ii) is to provide rapid progress to the solution, and it is envisaged that one of these steps will normally be accepted by the algorithm. If both these steps are unacceptable then the Cauchy step is tried. Only if the Cauchy step fails does the algorithm proceed to reduce ρ. Use of the Cauchy step in this way enables us to prove a convergence result by modifying arguments from an earlier SLP proof [5].

Trust region algorithms for NLP differ from those for unconstrained optimization in that it cannot necessarily be assumed that the predicted reduction in f is positive. We use the terminology introduced in [5] in which a step $\mathbf{d}$ that satisfies $\Delta l > 0$ is referred to as 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 (2.6) is satisfied. Thus a necessary condition for a step $\mathbf{d}$ to give rise to an f–type iteration is that both

$$\Delta l > 0 \qquad \text{and} \qquad \Delta f \geq \sigma \Delta q_C \tag{2.11}$$

are satisfied. If $\Delta l \leq 0$, or if the LP subproblem becomes incompatible, then the primary aim of the step 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 Δl is reduced (a consequence of having found the minimizer of $LP(\mathbf{x}^{(k)}, \rho)$). Thus the status of the test $\Delta l > 0$ 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.

Our algorithm differs in one important respect from that analysed in [6] In [6], 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. In our algorithm, all acceptable points with $h^{(k)} > 0$ are entered into the filter, both from f-type and h-type iterations, and this is done at the start of the iteration. This change of strategy affects the convergence proof quite significantly, although it is not clear which approach is best in practice (arguments in favour of either method can be advanced). We note that if $h^{(k)} = 0$, then the resulting iteration is always an f-type iteration. We also introduce the terminology of an *f-monotonic step* (or iteration) for a step $\mathbf{d}$ which satisfies $f(\mathbf{x}^{(k)} + \mathbf{d}) \leq f^{(k)}$.

In common with [3], [4], [5] and [6], our algorithm includes a feasibility restoration phase to handle the situation in which the LP subproblem becomes incompatible. The restoration phase aims to find a new point, acceptable to the filter, at which the LP subproblem is compatible. Because there are no entries in the filter with $h^{(j)} = 0$, it is always possible to obtain acceptability if a sufficiently small value of h is generated. Also, the LP subproblem is always compatible at a feasible point of P, and usually so if h is sufficiently small. Thus the restoration phase can be implemented by employing any method for solving a nonlinear system of equations and inequalities. It is of course possible that the value of f at the point generated by the restoration phase is significantly worse. The convergence proof of Section 3 provides an assurance that this cannot prevent convergence to a local solution. There is also the possibility that the restoration phase may converge to a point at which $h > 0$, without generating an acceptable point at which the LP subproblem is compatible. This we regard as an indication that problem P is incompatible (in a local sense). More discussion about the restoration phase is given in [5] and [6].

3 A Global Convergence Proof

In this section we present a proof of global convergence of the SLP-EQP filter algorithm of Figure 2 when applied to problem P.

Our global convergence theorem concerns Fritz–John necessary conditions (see for example Bazarra and Shetty [1]). Necessary conditions for $\mathbf{x}^*$ to solve P are that $\mathbf{x}^*$ is a feasible point, and that the set of directions

$$\{\mathbf{s} \mid \mathbf{s}^T\mathbf{g}^* < 0 \tag{3.1}$$

$$\mathbf{s}^T\mathbf{a}_i^* < 0 \quad i \in \mathcal{A}^*\} \tag{3.2}$$

is empty, where $\mathcal{A}^*$ denotes the set of active constraints at $\mathbf{x}^*$. Such a point is referred to as an *FJ point.* It follows that if $\mathbf{x}^\circ$ is a feasible point, but not an FJ point, then there exists a neighbourhood $\mathcal{N}^\circ$ of $\mathbf{x}^\circ$, a scalar $\varepsilon > 0$ and a vector $\mathbf{s}$ such that $\|\mathbf{s}\|_\infty = 1$ for which

$$\mathbf{s}^T\mathbf{g}(\mathbf{x}) \le -\varepsilon \tag{3.3}$$

$$\mathbf{s}^T\mathbf{a}_i(\mathbf{x}) \le -\varepsilon \quad i \in \mathcal{A}^\circ \tag{3.4}$$

for all $\mathbf{x} \in \mathcal{N}^\circ$. This result is a staightforward consequence of (3.1), (3.2) and continuity of the vectors $\mathbf{g}(\mathbf{x})$ and $\mathbf{a}_i(\mathbf{x})$.

The standard assumptions that are used in the proofs are summarized as follows.

Standard Assumptions *The set X is non-empty and bounded, the functions $f(\mathbf{x})$ and $\mathbf{c}(\mathbf{x})$ are twice continuously differentiable on an open set containing X, and the matrices $B^{(k)}$ are uniformly bounded.*

Remark A consequence of the boundedness assumptions is that there exists a constant $M > 0$, independent of $\mathbf{x}$ and k, such that for all $\mathbf{x} \in X$ and all k it follows that $\frac{1}{2}\mathbf{s}^T B^{(k)}\mathbf{s} \le M$, $\frac{1}{2}\mathbf{s}^T\nabla^2 c_i(\mathbf{x})\mathbf{s} \le M$ for $i = 1, 2, \ldots, m$, and $\frac{1}{2}\mathbf{s}^T(\nabla^2 f(\mathbf{x}) - B^{(k)})\mathbf{s} \le M$ hold for all vectors $\mathbf{s}$ such that $\|\mathbf{s}\|_\infty = 1$.

First we give a result that bounds the effects of nonlinearity in f and $\mathbf{c}$.

Lemma 2 *Let $\mathbf{d}_C$ be the Cauchy step arising from the solution of $LP(\mathbf{x}^{(k)}, \rho)$. If the standard assumptions hold then*

$$\Delta f_C \ge \Delta q_C - \alpha_C^2\rho^2 M \tag{3.5}$$

and

$$c_i(\mathbf{x}^{(k)} + \mathbf{d}_C) \le (1 - \alpha_C)h^{(k)} + \alpha_C^2\rho^2 M \qquad i = 1, 2, \ldots, m. \tag{3.6}$$

Proof By Taylor's theorem

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

where $\mathbf{y}$ denotes some point on the line segment from $\mathbf{x}^{(k)}$ to $\mathbf{x}^{(k)} + \mathbf{d}$. Hence

$$\Delta f_C = \Delta q_C - \tfrac{1}{2}\mathbf{d}_C^T(\nabla^2 f(\mathbf{y}) - B^{(k)})\mathbf{d}_C.$$

It follows by existence of the constant M and the trust region constraint on $\mathbf{d}_{LP}$ that (3.5) holds. Likewise by Taylor's theorem there exist points $\mathbf{y}_i$ such that

$$c_i(\mathbf{x}^{(k)} + \mathbf{d}) = c_i^{(k)} + \mathbf{a}_i^{(k)T}\mathbf{d} + \tfrac{1}{2}\mathbf{d}^T\nabla^2 c_i(\mathbf{y}_i)\mathbf{d} \qquad i = 1, 2, \dots, m. \tag{3.7}$$

Now

$$c_i^{(k)} + \mathbf{a}_i^{(k)T}\mathbf{d}_C = (1-\alpha_C)c_i^{(k)} + \alpha_C(c_i^{(k)} + \mathbf{a}_i^{(k)T}\mathbf{d}_{LP}) \le (1-\alpha_C)h^{(k)}$$

by definition of α_C and feasibility of $\mathbf{d}_{LP}$ in $LP(\mathbf{x}^{(k)}, \rho)$. Thus (3.6) follows from (3.7) using the trust region constraint and existence of M. *q.e.d.*

We now give a result relating to Δq_C.

Lemma 3 *Let standard assumptions hold and let $\mathbf{d}_{LP}$ solve $LP(\mathbf{x}^{(k)}, \rho)$. If $\Delta l \ge \eta\rho > 0$, where $\eta > 0$ is constant, then the resulting value of Δq_C satisfies the condition*

$$\Delta q_C \ge \tfrac{1}{2}\alpha_C \Delta l \ge \tfrac{1}{2}\min(\eta\rho,\ \tfrac{1}{2}\eta^2/M). \tag{3.8}$$

where M is defined as in the remark before Lemma 2. Moreover, if

$$\rho \le \tfrac{1}{2}(1-\sigma)\eta/M, \tag{3.9}$$

then $\alpha_C = 1$ and

$$\Delta f_C \ge \sigma\Delta q_C \ge \tfrac{1}{2}\sigma\eta\rho. \tag{3.10}$$

Proof We need to find the value of α in $[0,1]$ which minimizes the quadratic $-\alpha\Delta l + \frac{1}{2}\alpha^2 b$ where $b = \mathbf{d}_{LP}^T B^{(k)}\mathbf{d}_{LP} \le 2\rho^2 M$. Because $\Delta l > 0$, we only need consider $\alpha > 0$. The unconstrained minimizer, if it exists, is at $\alpha = \Delta l/b$. Thus if $\Delta l < b$, the minimizer is $\alpha_C = \Delta l/b < 1$ and $\Delta q_C = \frac{1}{2}\Delta l^2/b$. Otherwise if $\Delta l \ge b$, the minimizer is $\alpha_C = 1$ and

$$\Delta q_C = \Delta l - \tfrac{1}{2}b \ge \tfrac{1}{2}\Delta l.$$

Putting these together gives (3.8). For the second part, if $\alpha_C < 1$ then

$$\alpha_C = \Delta l/b \ge \eta\rho/(2\rho^2 M) > 1$$

if (3.9) holds, which is a contradiction. Thus $\alpha_C = 1$. Hence from (3.5), (3.8) and $\Delta l \ge \eta\rho$ it follows that

$$\Delta f_C \ge \sigma\Delta q_C + (1-\sigma)\Delta q_C - \rho^2 M \ge \sigma\Delta q_C + \tfrac{1}{2}(1-\sigma)\eta\rho - \rho^2 M.$$

Then (3.10) follows from (3.9). *q.e.d.*

Now we prove an important lemma which shows that an f-type step, acceptable to $(h^{(k)}, f^{(k)})$, is always generated in the neighbourhood of a feasible point that is not an FJ point.

Lemma 4 *Let standard assumptions hold and let $\mathbf{x}^\circ \in X$ be a feasible point of problem P but not an FJ point. Then there exist positive constants ε and κ and a neighbourhood $\mathcal{N}^\circ$ of $\mathbf{x}^\circ$, such that for all $\mathbf{x} \in \mathcal{N}^\circ \cap X$ and all ρ for which*

$$h(\mathbf{c}(\mathbf{x}))/\varepsilon \leq \rho \leq \kappa, \tag{3.11}$$

it follows that $LP(\mathbf{x}, \rho)$ has a feasible solution $\mathbf{d}_{LP}$ at which certain conditions hold. These are that the LP-predicted reduction satisfies

$$\Delta l \geq \rho\varepsilon, \tag{3.12}$$

the Cauchy step is given by $\mathbf{d}_C = \mathbf{d}_{LP}$ (that is $\alpha_C = 1$), the sufficient reduction condition (2.6) holds with

$$\Delta f_C \geq \sigma \Delta q_C \geq \tfrac{1}{2}\sigma\rho\varepsilon, \tag{3.13}$$

and finally

$$\Delta f_C \geq \gamma h(\mathbf{c}(\mathbf{x} + \mathbf{d}_C)), \tag{3.14}$$

which is the second condition in the envelope test (2.9) for $j = k$.

Proof We consider the solution of $LP(\mathbf{x}, \rho)$, and let $\mathbf{s}$ be a vector defined by (3.3) and (3.4) that exists for all $\mathbf{x} \in \mathcal{N}^\circ$. For active constraints at $\mathbf{x}^\circ$ we have from (3.4) that

$$c_i + \rho \mathbf{a}_i^T \mathbf{s} \leq h - \rho\varepsilon \qquad i \in \mathcal{A}^\circ, \tag{3.15}$$

where $c_i = c_i(\mathbf{x})$, etc. For inactive constraints $i \notin \mathcal{A}^\circ$, there exist positive constants $\bar{c}$ and $\bar{a}$, independent of $\mathbf{x}$, such that

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

by continuity of c_i and boundedness of $\mathbf{a}_i$ on X. It follows that

$$c_i + \rho \mathbf{a}_i^T \mathbf{s} \leq -\bar{c} + \rho\bar{a} \qquad i \notin \mathcal{A}^\circ. \tag{3.16}$$

Thus, if

$$h/\varepsilon \leq \rho \leq \bar{c}/\bar{a}, \tag{3.17}$$

it follows from (3.15) and (3.16) that

$$c_i + \rho \mathbf{a}_i^T \mathbf{s} \leq 0 \qquad i = 1, 2, \ldots, m,$$

and we are assured that $\rho\mathbf{s}$ is a feasible step, and hence that $LP(\mathbf{x},\rho)$ is a compatible subproblem.

It therefore follows by optimality of $\mathbf{d}_{LP}$ that

$$\Delta l \geq -\rho\mathbf{g}^T\mathbf{s} \geq \rho\varepsilon$$

from (3.3), which gives (3.12). Hence we deduce from Lemma 3 that if $\rho \leq \frac{1}{2}(1-\sigma)\varepsilon/M$ then $\alpha_C = 1$ and $\Delta f_C \geq \sigma\Delta q_C \geq \frac{1}{2}\sigma\rho\varepsilon$, which is (3.13). Consequently, using (3.6),

$$\Delta f_C - \gamma h(\mathbf{c}(\mathbf{x}+\mathbf{d}_C)) \geq \tfrac{1}{2}\sigma\rho\varepsilon - \gamma m\rho^2 M \geq 0$$

if $\rho \leq \sigma\varepsilon/(2\gamma mM)$, giving (3.14). Thus we may define the constant κ in (3.11) to be the least of $\sigma\varepsilon/(2\gamma mM)$ and the values $\bar{c}/\bar{a}$ and $\frac{1}{2}(1-\sigma)\varepsilon/M$ required earlier in the proof. *q.e.d.*

We also need a lemma giving a condition under which a sufficient reduction in h is obtained.

Lemma 5 *Let standard assumptions hold, let $\tau > 0$, and let $\mathbf{d}_C$ be the Cauchy step arising from the solution of $LP(\mathbf{x}^{(k)},\rho)$. Then $h(\mathbf{c}(\mathbf{x}^{(k)}+\mathbf{d}_C))$ satisfies the test $h(\mathbf{c}(\mathbf{x}^{(k)}+\mathbf{d}_C)) \leq \beta\tau$ if $\rho^2 \leq \beta\tau/(mM)$.*
Proof It follows from (2.7) and (3.6) that $h(\mathbf{c}(\mathbf{x}^{(k)}+\mathbf{d}_C)) \leq m\rho^2 M$, and the result follows if $\rho^2 \leq \beta\tau/(mM)$. *q.e.d.*

Now we proceed to analyse the algorithm of Figure 2. First we need to show that the inner iteration is a finite calculation. Here $\mathbf{x}^{(k)}$ is fixed and we consider what happens to the solution of $LP(\mathbf{x}^{(k)},\rho)$ 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 $\mathbf{d}_{LP} = \mathbf{0}$ solves $LP(\mathbf{x}^{(k)},\rho)$ 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 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 $LP(\mathbf{x}^{(k)},\rho)$ is incompatible. 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 $\|\mathbf{d}\|_\infty \leq \rho$, for sufficiently small ρ. Thus we need only consider constraints $i \in \mathcal{A}^{(k)}$. Also there is ultimately no difficulty in obtaining acceptability to the filter because $h(\mathbf{c}(\mathbf{x}^{(k)}+\mathbf{d})) \leq \beta\mathbf{h}^{(j)}$ for all $j \in \mathcal{F}^{(k)}$ will be satisfied for sufficiently small ρ. Because $\mathbf{x}^{(k)}$ is not a KT point, there exists an LP-feasible descent direction $\mathbf{s}$, $\|\mathbf{s}\|_\infty = 1$, and an $\eta > 0$ such that $\mathbf{s}^T\mathbf{g}^{(k)} = -\eta$ and $\mathbf{s}^T\mathbf{a}_i^{(k)} \leq 0$, $i \in \mathcal{A}^{(k)}$. Thus the solution vector $\mathbf{d}_{LP}$ is such that $\Delta l \geq \rho\eta$, and the results of Lemma 3 can be invoked. It then follows as in the proof of Lemma 4 that $\mathbf{d}_C$ gives rise to an acceptable f-type step for sufficiently small ρ, and hence the 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 SLP-EQP filter 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* $LP(\mathbf{x}, \rho)$ *is compatible for some* $\rho \geq \rho^\circ$.

($\mathbb{B}$) *A KT point of problem P is found* ($\mathbf{d}_{LP} = \mathbf{0}$ *solves* $LP(\mathbf{x}^{(k)}, \rho)$ *for some* k).

($\mathbb{C}$) *All accumulation points are feasible and there exists an accumulation point that is an FJ point.*

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, or any infinite subsequence, has one or more accumulation points.

All iterates in the outer iteration either have $h^{(k)} = 0$ or else $(h^{(k)}, f^{(k)})$ is entered into the filter. It follows from Lemma 1 that $h^{(k)} \to 0$ and hence any accumulation point is a feasible point.

We first consider the case that there exists an iteration K such that all iterations are f-monotonic for all $k \geq K$, and denote

$$\tau^{(K)} = \min_{\substack{j \in \mathcal{F}^{(K)} \\ h^{(j)} > h^{(K)}}} h^{(j)}. \tag{3.18}$$

We let $\mathbf{x}^\infty$ be any accumulation point of the main sequence and examine the proposition (to be contradicted) that $\mathbf{x}^\infty$ that is not an FJ point. It follows that there exists a neighbourhood $\mathcal{N}^\infty$ in which the set defined by (3.1) and (3.2) is not empty, and Lemma 4 applies. We consider any subsequence $k \in \mathcal{S}$ for which $\mathbf{x}^{(k)} \to \mathbf{x}^\infty$, and assume without loss of generality that $\mathbf{x}^{(k)} \in \mathcal{N}^\infty$ and $k \geq K$ for all $k \in \mathcal{S}$.

Now let $k \in \mathcal{S}$ and consider the interval

$$\frac{h^{(k)}}{\varepsilon} \leq \rho \leq \min\left\{\sqrt{\frac{\beta\tau^{(K)}}{mM}},\ \kappa,\ \rho^\circ\right\}. \tag{3.19}$$

It follows from Lemma 4 that any value of ρ in the interval would give rise to a step $\mathbf{d}_C$ that satisfies (3.12), (3.13) and (3.14). Let (h, f) denote the entry arising from the step $\mathbf{d}_C$. It follows from (3.14) that (h, f) would be acceptable to $(h^{(k)}, f^{(k)})$. It therefore follows by the f-monotonicity of iterations $K, K+1, \ldots, k-1$ that (h, f) would be acceptable to all filter entries $j \in \mathcal{F}^{(k)}$ for which $j \geq K$, and also to all entries with $j < K$ for which $f^{(j)} > f^{(K)}$. Any other entries in $\mathcal{F}^{(k)}$ must have $h^{(j)} > h^{(K)}$, and because (3.19) implies $\rho^2 \leq \beta\tau^{(K)}/(mM)$, it follows from Lemma 5 and (3.18) that (h, f) would be acceptable with respect to these entries also. Thus the Cauchy step would give rise to an f-type step that satisfies the sufficient reduction condition (3.13) and be acceptable to all entries in the filter.

The right hand side of (3.19) is a number, $\bar{\rho}$ say ($\bar{\rho} > 0$), independent of k, whilst the left hand side converges to zero. Thus there exists some iteration L such that $h^{(k)}/\varepsilon \leq \frac{1}{2}\bar{\rho}$ for all $k \geq L \geq K$. Thus, for any $k \geq L$, as ρ is reduced in the inner loop, either it must eventually fall within the interval (3.19) or a value to the right of the interval is accepted. If ρ falls within the interval, then the Cauchy step will be accepted if the EQP or SOC steps fail. Hence we can guarantee that a value $\rho^{(k)} > \frac{1}{2}\bar{\rho}$ will be chosen. If $\rho^{(k)} \leq \bar{\rho}$ then (3.13) holds. If a value to the right of the interval is accepted, then (2.6) must hold, and $\Delta l^{(k)} \geq \bar{\rho}\varepsilon$ holds because Δl is a monotonic increasing function of ρ. We therefore deduce from (3.8) that $\Delta f^{(k)} \geq \frac{1}{2}\sigma \min(\varepsilon\bar{\rho},\ \frac{1}{2}\varepsilon^2/M)$. Thus, in either case, $\Delta f^{(k)}$ is uniformly bounded away from zero, which, together with the f-monotonicity of the main sequence for $k \geq K$, contradicts the fact that the sequence $\{f^{(k)}\}$ is bounded below. Thus, in this case, any accumulation point is an FJ point.

Now we now consider the alternative case that there exists an infinite subsequence of iterations that are not f-monotonic. We let $\mathbf{x}^\infty$ be any accumulation point of this subsequence and examine the proposition (to be contradicted) that $\mathbf{x}^\infty$ that is not an FJ point. We consider any infinite subsequence $\mathcal{S}$ of iterations that are not f-monotonic, and for which $\mathbf{x}^{(k)} \to \mathbf{x}^\infty$ and $\mathbf{x}^{(k)} \in \mathcal{N}^\infty$.

For any k, the entries in the current filter must satisfy

$$\text{either} \qquad h^{(j)} \geq h^{(k)}/\beta \qquad \text{or} \qquad f^{(j)} \geq f^{(k)} + \gamma h^{(k)}$$

for all $j \in \mathcal{F}^{(k)}$, $j \neq k$, for otherwise $(h^{(k)},\ f^{(k)})$ would not be acceptable to $\mathcal{F}^{(k-1)}$. Because $\mathbf{x}^{(k)} \in \mathcal{N}^\infty$, it follows from Lemma 4 for any $k \in \mathcal{S}$ that there exists an interval (3.11) such that any ρ from the interval would give rise to a Cauchy step $\mathbf{d}_C$ that is f-type and satisfies (3.14). The entry (h, f) arising from this step would be acceptable to $(h^{(k)},\ f^{(k)})$, and to any entries $j \in \mathcal{F}^{(k)}$ for which $f^{(j)} > f^{(k)} + \gamma h^{(k)}$. All other entries in $j \in \mathcal{F}^{(k)}$ have $h^{(j)} > h^{(k)}/\beta$. It follows from Lemma 5 that the step $\mathbf{d}_C$ would be acceptable to any such entries if $\rho^2 \leq \beta h^{(k)}/(mM)$.

Thus if ρ satisfies

$$\frac{h^{(k)}}{\varepsilon} \leq \rho \leq \min\left\{\sqrt{\frac{\beta h^{(k)}}{mM}},\ \kappa,\ \rho^\circ\right\} \tag{3.20}$$

we see that the entry (h, f) is acceptable to all entries in $\mathcal{F}^{(k)}$. Because $h^{(k)} \to 0$ it follows for sufficiently large k that the range (3.20) becomes

$$\frac{h^{(k)}}{\varepsilon} \leq \rho \leq \sqrt{\frac{\beta h^{(k)}}{mM}}. \tag{3.21}$$

In the limit, because of the square root, the upper bound in (3.21) is more than twice the lower bound. Hence, as in an earlier paragraph, reducing ρ in the inner loop will eventually locate a value in the interval (3.21), or to the right of this interval, that gives rise to an acceptable Cauchy step for which $\Delta l > 0$ and $\Delta f_C > 0$. Either the Cauchy step will be taken or an EQP or SOC step, any one of which satisfies $\Delta f > 0$. Thus if $k \in \mathcal{S}$ is sufficiently large, an f–monotonic step is taken. This contradicts the fact that the case being considered is formed by a subsequence of steps that are not f-monotonic. *q.e.d.*

4 Discussion

Of course the algorithm of Figure 2 is only a framework for what might be implemented in practice. It is necessary to make decisions about how the quadratic model is defined, particularly in regard to the Hessian $B^{(k)}$. Also it may not always be appropriate to use an SOC step on some occasions, or even the EQP step. It is also possible that it might be advantageous to repeat the SOC step a finite number of times, or even to try some other type of step, before falling back on the Cauchy step. Clearly the theory permits such changes. There are also decisions to be made about how to implement the feasibility restoration phase.

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, as required for the last part of Theorem 1.

Another issue is that of whether to put all pairs $(h^{(k)}, f^{(k)})$ with $h^{(k)} > 0$ into the filter (as in this paper) or just those arising from h-type steps, as in [6]. The latter strategy is in a sense more non-monotonic, in that it allows points to be accepted that are dominated by previous points from f-type steps, so is perhaps more risky. It also allows a bigger 'gap' for accepting points generated in the restoration phase. The strategy used in this paper is more in keeping with the idea of not accepting points that are dominated by previous points. In practice there seems little to choose between the two strategies. It would be easy to implement and to prove convergence for an SLP-EQP algorithm based on just including h-type steps in the filter.

The convergence proof requires very few a-priori assumptions, for example no qualifications about linear independence and non-degeneracy are required, and there is no difficulty in finding global solutions of the LP subproblem (such as might exist for non-convex QP). The assumption that the iterates are bounded is fairly common and is a convenient way of ensuring that accumulation points exist and avoids the need to discuss unbounded NLP problems. We do not see this as a very restrictive assumption. The other assumption that is made is that the matrices $B^{(k)}$ are uniformly bounded. Usually $B^{(k)}$ will be some approximation to the Hessian of a Lagrangian function, so there is an implicit assumption here that multiplier estimates are bounded. A simple way to handle this is to require in the proof that the value of α_C used in calculating the Cauchy step satisfies a bound $\alpha_C \geq \delta > 0$, where δ is a fixed small constant. That is, if $\Delta l/b < \delta$ in Lemma 3, then $\alpha_C = \delta$ would be used in the calculation of Δq_C. The only danger here is that too large a value of δ might inhibit superlinear convergence of the algorithm. However, if unbounded multipliers arise then superlinear convergence is not always to be expected, so in practice this modification is unlikely to be disadvantageous.

The main limitation of the proof here is that it does not say anything about problems with equality constraints, whereas such problems form a significant proportion of practical problems. There are unlikely to be serious problems in extending the proof, along the lines of [6] in which the SLP proof of [5] is extended to include a treatment of NLP problems that include equality constraints. The outcome is likely to be the same here, namely that convergence to a Kuhn-Tucker point is proved, under the assumption that a Mangasarian-Fromowitz constraint qualification (MFCQ) is satisfied at the accumulation point. Convergence to feasible non-KT points that do not satisfy MFCQ cannot be ruled out.

Nothing is said in this paper about the rate of convergence of the algorithm, as this depends on the way that the Hessian approximation $B^{(k)}$ is calculated. Given sensible decisions here we would usually expect superlinear convergence to occur, and second order convergence if exact second derivatives were available. However, it is by no means certain that this is the case, as it is an open question for SQP filter algorithms as to whether acceptability to the filter is compatible with accepting the SQP step in the limit. In practice this does not seem to be a problem and second order convergence is observed whenever it would be expected.

A significant advantage of SLP-EQP algorithms is that they are readily able to accept approximate solutions of the EQP subproblems if they satisfy (2.11). Thus iterative algorithms such as (preconditioned) conjugate gradients or truncated Newton can readily be subsumed into the convergence framework. Together with the effectiveness of software for large scale LP, this permits very large NLP problems to be tackled. Also there is the possibility of using approximate solutions of the LP subproblem if the LP method is able to provide bounds on the nearness to the exact solution of the subproblem, such as with feasible primal-dual interior point methods. Our theory is readily modified to allow such approximate solutions if $\Delta l > 0$ and the approximate predicted reduction is within some fixed fraction of Δl.

5 References

[1] Bazaraa M.S. and Shetty C.M. (1979), *NONLINEAR PROGRAMMING Theory and Algorithms*, John Wiley and Sons, New York.

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

[3] 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).

[4] 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.

[5] Fletcher R., Leyffer S. and Toint Ph.L (1998), On the Global Convergence of an SLP–Filter Algorithm, Dundee University, Dept. of Mathematics, Report NA/183.

[6] Fletcher R., Leyffer S. and Toint Ph.L (2000), On the Global Convergence of a Filter-SQP Algorithm, Dundee University, Dept. of Mathematics, Report NA/197.

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

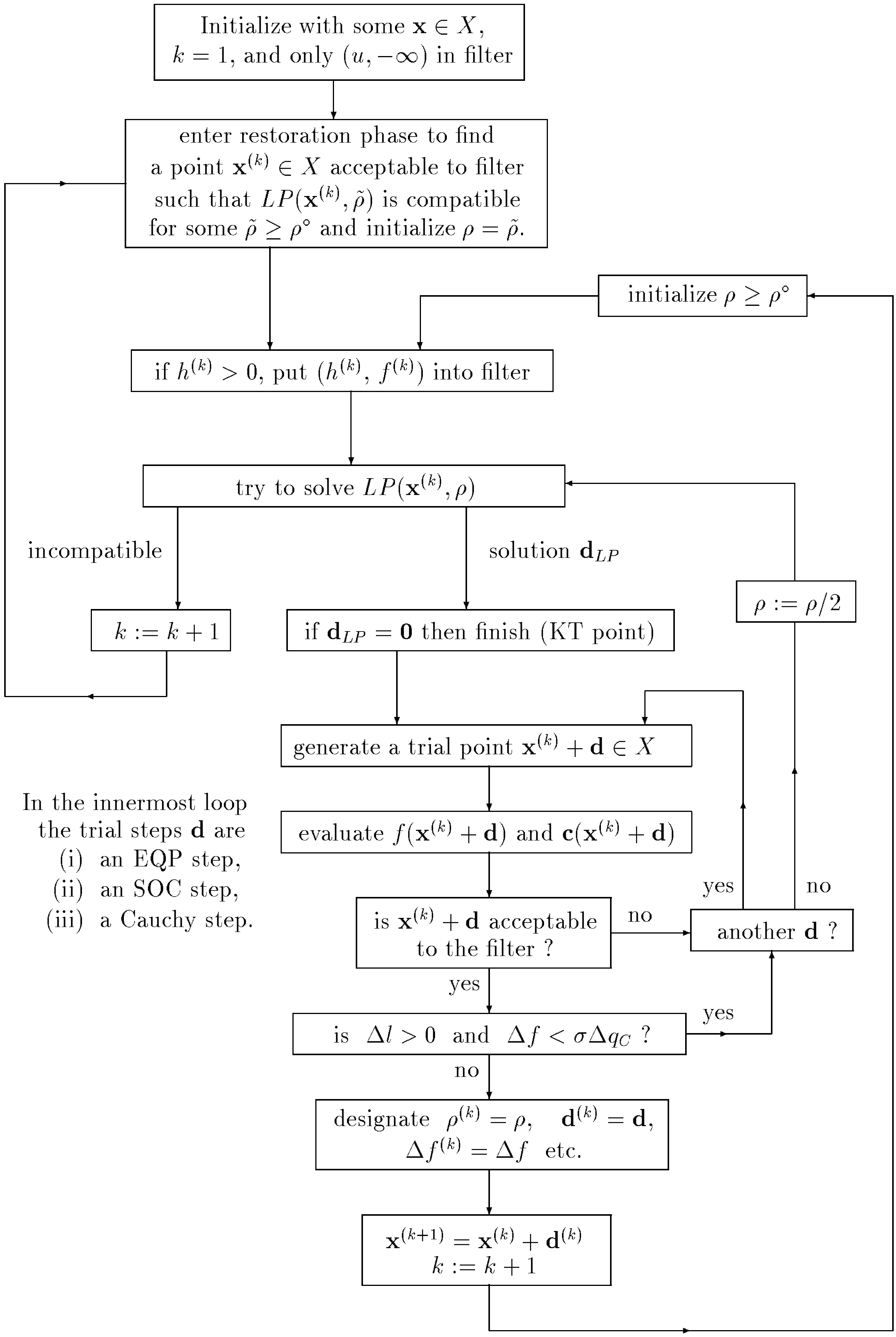

Figure 2: An SLP-EQP Filter Algorithm