A GLOBALLY CONVERGENT FILTER METHOD FOR NONLINEAR PROGRAMMING*

CLÓVIS C. GONZAGA[†], ELIZABETH KARAS[‡], AND MÁRCIA VANTI[§¶]

Abstract. In this paper we present a filter algorithm for nonlinear programming and prove its global convergence to stationary points. Each iteration is composed of a feasibility phase, which reduces a measure of infeasibility, and an optimality phase, which reduces the objective function in a tangential approximation of the feasible set. These two phases are totally independent, and the only coupling between them is provided by the filter. The method is independent of the internal algorithms used in each iteration, as long as these algorithms satisfy reasonable assumptions on their efficiency. The main algorithm uses no trust regions, and no properties of the linear models. Under standard hypotheses, we show two results: for a filter with minimum size, the algorithm generates a stationary accumulation point; for a slightly larger filter, all accumulation points are stationary.

1. Introduction. We shall study the nonlinear programming problem

$$(P) \qquad \begin{array}{rl} \text{minimize} & f_0(x) \\ \text{subject to} & f_{\mathcal{E}}(x) = 0 \\ & f_{\mathcal{I}}(x) \le 0, \end{array}$$

where the index sets $\mathcal{E}$ and $\mathcal{I}$ refer to the equality and inequality constraints respectively. Let the cardinality of $\mathcal{E} \cup \mathcal{I}$ be m, and assume that the functions $f_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 0, 1, \dots, m$ are continuously differentiable. The Jacobian matrices of $f_{\mathcal{E}}$ and $f_{\mathcal{I}}$ are denoted respectively $A_{\mathcal{E}}(\cdot)$ and $A_{\mathcal{I}}(\cdot)$.

We define the function $f^+ : \mathbb{R}^n \to \mathbb{R}^m$ by

$$f_i^+(x) = \begin{cases} f_i(x) & \text{if } i \in \mathcal{E} \\ \max\{0, f_i(x)\} & \text{if } i \in \mathcal{I}. \end{cases} \tag{1.1}$$

The i^{th} constraint, $i = 1, \dots, m$, is satisfied at $x \in \mathbb{R}^n$ if $f_i^+(x) = 0$. We consider a measure of constraint infeasibility $x \in \mathbb{R}^n \mapsto h(x)$, which is an exact penalty applied to the constraints. Usually this measure is given by

$$h(x) = \|f^+(x)\|, \tag{1.2}$$

where $\|\cdot\|$ denotes an arbitrary norm.

A nonlinear programming algorithm must deal with two conflicting criteria, f_0 and h, which must be simultaneously minimized, with preference given to the infeasibility measure h, which must be driven to zero.

Optimality and feasibility can be combined using penalty functions or augmented Lagrangians, or can be treated more or less independently. The methods studied in this paper belong to the class in which f_0 and h are treated as two independent objectives. Each iteration of these methods is composed of two phases: a feasibility phase, which decreases h, followed by an optimization phase, which decreases f_0.

---

*REVISED IN AUGUST 2002

[†]Department of Mathematics, Federal University of Santa Catarina. Cx. Postal 5210, 88040-970 Florianópolis, SC, Brazil; e-mail: clovis@mtm.ufsc.br.

[‡]Department of Mathematics, Federal University of Paraná and Federal University of Santa Catarina. Cx. Postal 19081, 81531-990 Curitiba, PR, Brazil; e-mail: karas@mat.ufpr.br.

[§]Department of Electrical Engineering, Federal University of Santa Catarina. Cx. Postal 5210, 88040-970 Florianópolis, SC, Brazil; e-mail: marcia@labplan.ufsc.br.

[¶]The authors are supported by CNPq and CAPES - Brazil.

Such methods can be traced back to Rosen's gradient projection method [16] and Abadie and Carpentier's GRG [1]. They are surveyed in Martínez and Pilotta [12]. Combining the ideas of sequential quadratic programming and trust region algorithms for problems with equality constraints only, Celis, Dennis and Tapia [6] started a line of research which led to the method of Byrd and Omojokum [3, 15]: each iteration of this method works in a trust region centered at the current iterate x^k, and is composed of a *normal* step (feasibility step) followed by a *tangential* step (optimality step). The tangential step must follow a direction in the null space of the constraint Jacobian at x^k.

The feasibility and optimality phases become more independent in the *inexact restoration* algorithms described by Martínez [10] and by Martínez and Pilotta [11, 12], who place the trust region used in each iteration around the point obtained *after* the feasibility phase. Any method for reducing h can be used in the feasibility phase: they describe an algorithm for problems with nonlinear equality constraints and box inequality constraints. Methods for the feasibility phase can also use ideas from Byrd, Gilbert and Nocedal [4] and from Byrd, Hribar and Nocedal [5], who rewrite the problem using equality constraints and nonnegative slack variables.

In these algorithms, the progress is usually measured by a *merit function* $\psi = f_0 + \nu h$, where ν is a positive weight. At iteration k, the points in $\{x \in \mathbb{R}^n \mid \psi(x) \geq \psi(x^k)\}$ are *forbidden*, and the step tries to decrease the value of ψ. The choice of ν may be tricky: small values of ν may forbid the optimal solutions; large values of ν may slow down the algorithm.

As a rule, algorithms must include some procedure to increase ν when needed, increasing the importance of h in ψ. This choice of ν usually depends on both the feasibility and optimality steps, reducing their independence.

Filter algorithms. Filter algorithms define a *forbidden region* in a clever way, by memorizing the pairs $(f_0(x^k), h(x^k))$ from well chosen former iterations, and then avoiding points dominated by these by the usual Pareto domination rule:

$$\text{" } x \text{ dominates } y \quad \text{if and only if} \quad f_0(y) \geq f_0(x) \quad \text{and} \quad h(y) \geq h(x)\text{"}.$$

We cannot construct the set of forbidden points, but it very easy to check whether a point belongs to it by performing a small number of comparisons in $\mathbb{R}^2$.

These methods were introduced by Fletcher and Leyffer in their important paper [8], and a global convergence proof was obtained by Fletcher, Gould, Leyffer, Toint and Wächter [7]. The approach was also applied to interior point algorithms by Ulbrich, Ulbrich and Vicente [19]. In these papers, each feasibility phase must reduce h until a property called *compatibility* is verified, which depends on a trust region radius and on the linear model of the constraints.

Our method is an inexact restoration algorithm in the sense of Martínez and Pilotta [11], which uses a filter. The method has the following characteristics:

- Each iteration starts with a filter and its associated forbidden region.
- The feasibility and optimality phases are totally independent, and may be based on any algorithms satisfying some reasonable hypotheses. The only connection between both phases is that they are not allowed to generate forbidden points.
- Differently from the filter algorithms cited above, no compatibility is required after a feasibility step: the only requirement is that h decreases by at least a fixed ratio.

- In our first algorithm, the number of pairs $(h(x^k), f_0(x^k))$ introduced in the filter is perhaps the minimum possible to guarantee the existence of a stationary accumulation point.

Local convergence. In this paper we deal with the global convergence of filter algorithms, without discussing details of the internal algorithms. Fletcher and Leyffer [8] comment that filter algorithms may suffer from the Maratos effect, and propose a second order correction to remedy this shortcoming. Wächter and Biegler in their recent work [18] propose a filter method using line searches, and also discuss the usage of a second order correction. In our general approach, it is easy to show that Maratos effect will be present when the method is applied to Powell's example [17]. Although we believe that second order correction schemes can be devised for this general setting, this will not be discussed in this paper.

Structure of the paper. In this section we present some general definitions and hypotheses. Sec. 2 describes the main algorithm and proves that under a very general hypothesis on the behavior of a complete step of the algorithm, any sequence generated by it has a stationary accumulation point. This section also discusses how to break this general hypothesis into reasonable independent assumptions for the feasibility and optimality phases. Sec. 3 describes the internal algorithms, and shows how to satisfy the hypotheses used in Sec. 2. Sec. 4 deepens the convergence analysis, showing that the objective values always converge under the hypotheses in Sec. 2, and presents two improvements on the algorithms: first, using a slightly larger filter, we prove that all accumulation points are stationary; second, we discuss a simplified optimality step, using the Jacobian matrices already calculated in the feasibility phase. Sec. 5 shows a graphical example, and an Appendix proves some continuity properties.

Hypotheses. We shall develop algorithms which generate sequences (x^k) and (z^k) in $\mathbb{R}^n$. Here are the general hypotheses used in this paper.

(H1) The iterates (x^k) and (z^k) remain in a convex compact domain $X \subset \mathbb{R}^n$.

(H2) All the functions $f_i(\cdot)$ for $i = 0, 1, \dots, m$ are Lipschitz continuously differentiable in an open set containing X.

(H3) All feasible accumulation points $\bar{x} \in X$ of (x^k) satisfy the Mangasarian-Fromovitz (M-F) qualification condition, namely, the gradients $\nabla f_i(\bar{x})$ for $i \in \mathcal{E}$ are linearly independent, and there exists a direction $d \in \mathbb{R}^n$ such that $A_{\mathcal{E}}(\bar{x})d = 0$ and $A_{\bar{\mathcal{I}}}(\bar{x})d < 0$, where $\bar{\mathcal{I}} = \{i \in \mathcal{I} \mid f_i(\bar{x}) = 0\}$.

The first hypothesis is quite usual. It can be enforced by adding a large box constraint to the problem. If the set $\{x \in \mathbb{R}^n \mid h(x) \leq \bar{H}\}$ is bounded for some $\bar{H} \geq h(x^0)$, then the filter may start with a pair $(-\infty, \bar{H})$ (see below for the filter structure), thus forbidding forever points x with $h(x) \geq \bar{H}$. Similarly, if $\{x \in \mathbb{R}^n \mid f_0(x) \leq \bar{F}\}$ for some upper bound $\bar{F}$ for the value of an optimal solution, then the pair $(\bar{F}, -\infty)$ in the filter ensures (H1). Of course, both entries can be used if $\{x \in \mathbb{R}^n \mid h(x) \leq \bar{H}, f_0(x) \leq \bar{F}\}$ is bounded.

From (H2) we conclude that for $x, y \in X$ and $i = 0, 1, \dots, m$,

$$f_i(y) = f_i(x) + \nabla f_i(x)^T (y - x) + o(x, y), \tag{1.3}$$

where $|o(x, y)| \leq M \|x - y\|^2$ and $M > 0$ is a Lipschitz constant.

The linearized sets. We shall associate with each $z \in \mathbb{R}^n$ a linearization of the set $\{x \in \mathbb{R}^n \mid f_{\mathcal{E}}(x) = f_{\mathcal{E}}(z),\ f_{\mathcal{I}}(x) \leq f_{\mathcal{I}}^+(z)\}$:

$$L(z) = \{x \in \mathbb{R}^n \mid A_{\mathcal{E}}(z)\ (x - z) = 0,\ f_{\mathcal{I}}(z) + A_{\mathcal{I}}(z)\ (x - z) \leq f_{\mathcal{I}}^+(z)\}. \tag{1.4}$$

At a feasible point z, $L(z)$ is a linearization of the feasible set. The following facts are easily seen:

- The Mangasarian-Fromowitz condition at a feasible point z is equivalent to the following: $A_{\mathcal{E}}(z)$ has linearly independent rows and the set $L(z)$ satisfies a Slater condition, i.e., $L(z)$ has an interior point, a point $y \in L(z)$ such that $f_{\mathcal{I}}(z) + A_{\mathcal{I}}(z)(y-z) < f_{\mathcal{I}}^+(z)$.
- The Karush-Kuhn-Tucker (KKT) conditions for (P) at z coincide with the KKT conditions at z for the problem of minimizing $f_0(\cdot)$ in $L(z)$. These conditions are also equivalent to the inexistence of a feasible descent direction from z into $L(z)$.

Optimality conditions. Here we make some comments on optimality conditions and on our usage of the expression *stationary point.*

Let us define the *projected Cauchy direction* or *projected gradient direction* associated with each $z \in \mathbb{R}^n$

$$d_c(z) = P_{L(z)}(z - \nabla f_0(z)) - z, \tag{1.5}$$

where $P_\Gamma(w)$ denotes the orthogonal projection of $w \in \mathbb{R}^n$ onto the closed set $\Gamma \subset \mathbb{R}^n$.

The projected gradient direction is well known. See for instance Bertsekas [2]. It satisfies $d_c(z) = 0$ if and only if there exists no feasible descent direction from z into $L(z)$. We conclude from the facts above that at a feasible z, the KKT conditions are equivalent to $d_c(z) = 0$. If $d_c(z) \neq 0$, then $\nabla f_0(z)^T d_c(z) < 0$.

Actually, this direction is the main construct used by Martínez and Svaiter [13] to define an optimality condition which lies between KKT and Fritz-John in generality: a feasible point $\bar{x}$ satisfies a Martínez-Svaiter optimality condition if and only if

$$\liminf_{x \to \bar{x}} \|d_c(x)\| = 0. \tag{1.6}$$

This optimality condition is actually quite constructive: what we shall prove in this paper is that our algorithms produce feasible points satisfying (1.6). These points will be called *stationary.*

Here we have two possible courses of action: either we rely on their paper and do not use the Mangasarian-Fromowitz condition, or use it and the fact that in this case KKT and Martínez-Svaiter are equivalent conditions. We choose the second option.

For completeness, we now prove this equivalence, using continuity properties of the point to set map $L(\cdot)$ which are shown in Appendix A. We keep this treatment in the paper because we believe that it may have some interest in itself.

LEMMA 1.1. *Let $\bar{x}$ be a feasible point satisfying a M-F condition. Then*
(i) The map (1.5) *is continuous at $\bar{x}$.*
(ii) $\bar{x}$ satisfies the Karush-Kuhn-Tucker conditions if and only if it satisfies the Martínez-Svaiter conditions.

Proof. (i) follows directly from Lemmas A.1 and A.2: under a M-F condition, $z \mapsto L(z)$ is a continuous map at $\bar{x}$ by Lemma A.1, and Lemma A.2 ensures that $z \mapsto P_{L(z)}(z - \nabla f_0(z))$ is continuous because $\nabla f_0(\cdot)$ is continuous.
To prove (ii), note that for a continuous map $d_c(\cdot)$, (1.6) is equivalent to $d_c(\bar{x}) = 0$, which as we saw above is equivalent to the KKT conditions, completing the proof. ☐

Notation. Given two non-negative functions $g_1, g_2 : X \subseteq \mathbb{R}^n \to \mathbb{R}$ we say that:

- $g_1(x) = O(g_2(x))$ (or equivalently $g_2(x) = \Omega(g_1(x))$) in $\Gamma \subseteq X$ if there exists $M > 0$ such that for all $x \in \Gamma$, $g_1(x) \leq M g_2(x)$.
- $g_1(x) = o(g_2(x))$ in $\Gamma \subseteq X$ if $\displaystyle\lim_{g_2(x) \to 0^+} \frac{g_1(x)}{g_2(x)} = 0$.

2. The algorithm. In this section we present the method, with no specification of the internal algorithms used in the feasibility and optimality steps. Afterwards we state assumptions on the performance of these steps, and prove that any sequence generated by the algorithm has a stationary accumulation point. The next section will show that quite usual methods for the internal steps fulfill these assumptions.

ALGORITHM 2.1. *Filter algorithm.*
Data: $x^0 \in \mathbb{R}^n$, $F_0 = \emptyset$, $\mathcal{F}_0 = \emptyset$, $\alpha \in (0,1)$.
$k = 0$
REPEAT
    $(\tilde{f}_0, \tilde{h}) = (f_0(x^k) - \alpha h(x^k),\ (1-\alpha)h(x^k))$.
    Construct the set $\bar{F}_k = F_k \bigcup \{(\tilde{f}_0, \tilde{h})\}$.
    Define the set $\bar{\mathcal{F}}_k = \mathcal{F}_k \bigcup \{x \in \mathbb{R}^n \mid f_0(x) \geq \tilde{f}_0, h(x) \geq \tilde{h}\}$.
    Feasibility phase:
        if $h(x^k) = 0$ then set $z^k = x^k$
        else compute $z^k \notin \bar{\mathcal{F}}_k$ such that $h(z^k) < (1-\alpha)\ h(x^k)$.
            if impossible then stop with insuccess.
    Optimality phase:
        if z^k is stationary then stop with success
        else compute $x^{k+1} \notin \bar{\mathcal{F}}_k$ such that $x^{k+1} \in L(z^k)$ and $f_0(x^{k+1}) \leq f_0(z^k)$.
    Filter update:
        if $f_0(x^{k+1}) < f_0(x^k)$ then
            $F_{k+1} = F_k, \quad \mathcal{F}_{k+1} = \mathcal{F}_k$ (f_0-iteration)
        else
            $F_{k+1} = \bar{F}_k, \quad \mathcal{F}_{k+1} = \bar{\mathcal{F}}_k$ (h-iteration)
    $k = k + 1$.

Sec. 5 shows a graphical example, where each step of the algorithm is depicted. The main feature of the algorithm is the construction of the filter: at the beginning of each iteration, the pair $(f_0(x^k) - \delta,\ h(x^k) - \delta)$, with $\delta = \alpha h(x^k)$, $\alpha \in (0,1)$, is *temporarily* introduced in the filter. After the complete iteration, this entry will become permanent in the filter only if the iteration *does not* produce a decrease in f_0.

The algorithm deals with the filter and with the forbidden set associated with it. One must keep in mind that the forbidden set is never constructed, but helps the understanding of the process.

Stopping rules. The algorithm can stop in two situations:
(i) A stationary point is obtained. In this case there is nothing to prove.
(ii) The feasibility algorithm fails. This may well happen, depending on the method used. A common condition that may cause the failure is the existence of a stationary point $\bar{x}$ for $h(\cdot)$, with $h(\bar{x}) \neq 0$.

Elimination of filter entries. Whenever a new entry (f_0^j, h^j) is introduced in the filter, one can eliminate from it all entries dominated by the incoming one. This saves comparisons when checking whether a point is forbidden. See the example in Sec. 5.

From now on we shall assume that the algorithm generates infinite sequences (x^k) and (z^k). We also assume that the Hypotheses (H1-H3) are satisfied, and now we state the main assumption on the performance of the algorithm at each iteration. We shall postpone the discussion of this assumption for the end of this section, where it will be thoroughly analyzed and replaced by simpler ones. In the next section we shall state methods which satisfy this assumption.

The main hypothesis.
Given an iterate x^k, we start by defining the *filter slack* at x^k:

$$H_k = \min \left\{ 1, \min\{h^j \mid (f_0^j, h^j) \in F_k, f_0^j \le f_0(x^k)\} \right\}, \tag{2.1}$$

illustrated in Figure 2.1. Our main hypothesis is:
(H4) Given a feasible non-stationary point $\bar{x} \in X$, there exists a neighborhood V of $\bar{x}$ such that for any iterate $x^k \in V$,

$$f_0(x^k) - f_0(x^{k+1}) = \Omega(\sqrt{H_k}). \tag{2.2}$$

Note that (H4) is a local condition. The relation (2.2) means that there exists $M > 0$ dependent on $\bar{x}$ such that whenever x^k is near $\bar{x}$, $f_0(x^k) - f_0(x^{k+1}) \ge M\sqrt{H_k}$.

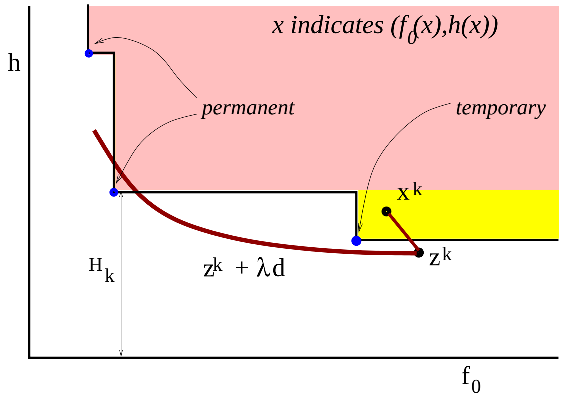

FIG. 2.1. *Example of the set $\bar{F}_k$ and of the quantity H_k.*

The following facts follow directly from the hypotheses and the construction made by the algorithm.

FACT 2.2. *Given $k \in \mathbb{N}$, $x^{k+p} \notin \mathcal{F}_{k+1}$, for all $p \ge 1$.*

FACT 2.3. *Given $k \in \mathbb{N}$, at least one of the following two situations occurs.*
(i) $h(x^{k+1}) \le (1-\alpha)\, h(x^k)$.
(ii) $f_0(x^{k+1}) \le f_0(x^k) - \alpha\, h(x^k)$.

FACT 2.4. *Given $k \in \mathbb{N}$, $h^j > 0$ for all $j \in \mathbb{N}$ such that $(f_0^j, h^j) \in F_k$. Consequently, $H_k > 0$ for all $k \in \mathbb{N}$.*

By the Algorithm 2.1, the pair $(\tilde{f}_0, \tilde{h})$ is included in the filter at the end of the iteration if and only if that iteration is an h-iteration. If $\tilde{h} = h(x^k) = 0$, then $z^k = x^k$ and $f_0(x^{k+1}) < f_0(z^k)$, so the iteration k is an f_0-iteration, and both statements in the Fact follow.

LEMMA 2.5. *Let $\bar{x} \in X$ be a non-stationary point. Then there exist $\bar{k} \in \mathbb{N}$ and a neighborhood V of $\bar{x}$ such that whenever $k > \bar{k}$ and $x^k \in V$, the iteration k is an f_0-iteration.*

Proof. If $\bar{x} \in X$ is a feasible point, then by (H4) and Fact 2.4 there exists a neighborhood V of $\bar{x}$ such that for all $x^k \in V$,

$$f_0(x^k) - f_0(x^{k+1}) = \Omega(\sqrt{H_k}) > 0,$$

and k is an f_0-iteration.

Assume that $\bar{x}$ is infeasible, i.e. $h(\bar{x}) > 0$. Assume by contradiction that there exists an infinite set $\mathcal{K} \subset \mathbb{N}$ such that $x^k \xrightarrow{\mathcal{K}} \bar{x}$ and all iterations in $\mathcal{K}$ are h-iterations. Since h and f_0 are continuous functions, we have

$$h(x^k) \xrightarrow{\mathcal{K}} h(\bar{x}) \qquad \text{and} \qquad f_0(x^k) \xrightarrow{\mathcal{K}} f_0(\bar{x}).$$

Then there must exist $k_1 \in \mathcal{K}$ such that for all $k \in \mathcal{K}$, $k \geq k_1$

$$|h(x^k) - h(\bar{x})| \;<\; \frac{\alpha}{2}\, h(x^{k_1}) \qquad \text{and} \qquad |f_0(x^k) - f_0(\bar{x})| \;<\; \frac{\alpha}{2}\, h(x^{k_1}). \tag{2.3}$$

For any given $k_2 \in \mathcal{K}$ such that $k_2 > k_1$,

$$|h(x^{k_2}) - h(\bar{x})| \;<\; \frac{\alpha}{2}\, h(x^{k_1}) \qquad \text{and} \qquad |f_0(x^{k_2}) - f_0(\bar{x})| \;<\; \frac{\alpha}{2}\, h(x^{k_1}). \tag{2.4}$$

Using the triangle inequality, (2.3) and (2.4), we have,

$$|h(x^{k_2}) - h(x^{k_1})| \;<\; \alpha\, h(x^{k_1}) \qquad \text{and} \qquad |f_0(x^{k_2}) - f_0(x^{k_1})| \;<\; \alpha\, h(x^{k_1}).$$

Therefore $x^{k_2} \in \mathcal{F}_{k_1+1}$, contradicting Fact 2.2 and completing the proof. □

LEMMA 2.6. *Suppose that $(x^k)_{k\in\mathbb{N}}$ has no stationary accumulation point. Then for k sufficiently large, all iterations are f_0-iterations.*

Proof. Assume by contradiction that there exists an infinity of h-iterations. Then there exists an infinite set $\mathcal{K}_1 \subset \mathbb{N}$ such that for $k \in \mathcal{K}_1$, the iteration k is an h-iteration. By Hypothesis (H1), $(x^k)_{k\in\mathcal{K}_1}$ is bounded, and hence there exists $\mathcal{K}_2 \subset \mathcal{K}_1$ and $\bar{x} \in \mathbb{R}^n$ such that $x^k \xrightarrow{\mathcal{K}_2} \bar{x}$. From the previous lemma, $\bar{x}$ must be a stationary accumulation point, contradicting the hypothesis and completing the proof. □

THEOREM 2.7. *The sequence (x^k) has a stationary accumulation point.*

Proof. Assume by contradiction that $(x^k)_{k\in\mathbb{N}}$ has no stationary accumulation point. Then from Lemma 2.6 for k large (say, $k > k_1$), all iterations are f_0-iterations, $f_0(x^k)$ decreases and hence

$$f_0(x^{k+1}) - f_0(x^k) \to 0. \tag{2.5}$$

For any $k \geq k_1$, $F_k = F_{k_1}$ by construction, and using Fact 2.4, $H_k \geq H_{k_1} > 0$.

The sequence (x^k) cannot have a feasible accumulation point, because by the hypothesis (H4), if there exists $\mathcal{K}_1 \in \mathbb{N}$ and a feasible $\bar{x} \in X$ such that $x^k \xrightarrow{\mathcal{K}_1} \bar{x}$, then for large $k \in \mathcal{K}_1$, (say $k > k_2 > k_1$)

$$f_0(x^k) - f_0(x^{k+1}) = \Omega(\sqrt{H_{k_1}}) > 0,$$

contradicting (2.5).

Now we prove the following claim: for large $k \in \mathbb{N}$,

$$h(x^{k+1}) \leq (1-\alpha)\; h(x^k). \tag{2.6}$$

Assume by contradiction that in some infinite set $\mathcal{K}_2 \subset \mathbb{N}$,

$$h(x^{k+1}) > (1-\alpha)\; h(x^k).$$

Using Fact 2.3, for $k \in \mathcal{K}_2$,

$$f_0(x^{k+1}) \leq f_0(x^k) - \alpha\; h(x^k).$$

Using (2.5), we conclude that $h(x^k) \xrightarrow{\mathcal{K}_2} 0$, which contradicts the fact that (x^k) has no feasible accumulation points.

Hence (2.6) holds and $h(x^k)$ converges linearly to zero. This again contradicts the fact that (x^k) has no feasible accumulation points, completing the proof. □

The hypothesis (H4).
This hypothesis is an assumption on each complete iteration. Although it may be difficult to check for specific algorithms, its interpretation is simple: near a non-stationary point, the optimality step dominates, and the reduction of f_0 is large. The filter slack H_k indicates how much h is allowed to increase in the tangential step and, by being tangential, it is expected that h changes with the square of the variation of x. In an efficient tangential step, f_0 will vary linearly with the variation of x, and then (H4) will be true.

Now we show how (H4) can be replaced by simpler hypotheses made separately for the feasibility and optimality steps.

Feasibility step condition.
(H5) At all iterations $k \in \mathbb{N}$, the feasibility step must satisfy

$$h(x^k) - h(z^k) = \Omega(\|z^k - x^k\|). \tag{2.7}$$

This can also be stated as

$$\|z^k - x^k\| = O(h(x^k)), \tag{2.8}$$

because $h(z^k) \geq 0$. Note that since $\nabla f_0(\cdot)$ is bounded in X, by the mean-value theorem, for all $k \in \mathbb{N}$,

$$|f_0(z^k) - f_0(x^k)| = O(\|z^k - x^k\|).$$

Using this and (2.8) we have

$$|f_0(z^k) - f_0(x^k)| = O(h(x^k)). \tag{2.9}$$

Optimality step condition.
(H6) Given a feasible non-stationary point $\bar{x} \in X$, there exists a neighborhood V of $\bar{x}$ such that for any iterate $x^k \in V$,

$$f_0(z^k) - f_0(x^{k+1}) = \Omega(\sqrt{H_k}). \tag{2.10}$$

The assumption (H5) is used by Martínez [10], and is a global condition. It means that the feasibility step must be efficient, in the sense that the direction $z^k - x^k$ must be a good descent direction for h. Martínez discusses this hypothesis and shows that it is satisfied under reasonable conditions. The assumption (H6) isolates the tangential step, and is local (associated with each given non-stationary feasible point). It has the same interpretation as the one given for (H4), but now without the influence of the feasibility step.

Remark. Note however that this condition is not completely independent of the feasibility phase, because it uses H_k, which is associated with x^k. Also, the condition is stated for $x^k \in V$, and not $z^k \in V$, but this is not important because $\|x^k - z^k\| = O(h(x^k))$: if x^k is near $\bar{x}$, then the same is true for z^k.

Before proving that (H5) and (H6) imply (H4), let us state one more hypothesis which is not needed here, but which is very reasonable and will be useful ahead. It is similar to (H5), but applied to the objective function.

(H7) Given a feasible non-stationary point $\bar{x} \in X$, there exists a neighborhood V of $\bar{x}$ such that for any iterate $x^k \in V$,

$$f_0(z^k) - f_0(x^{k+1}) = \Omega(\|x^{k+1} - z^k\|),$$

With this hypothesis, the interpretation of (H6) becomes quite simple: it can be stated as $\|z^k - x^{k+1}\| = \Omega(\sqrt{H_k})$, and will be true when the variation of h is of the order of $\|z^k - x^{k+1}\|^2$, which is quite reasonable in a tangential step. Figure 2.1 illustrates the trajectory of the pair $(f_0(z^k + \lambda d), h(z^k + \lambda d))$ as λ grows and $d = x^{k+1} - z^k$.

Finally, we prove two lemmas, extending for the whole step the properties of the tangential step near a feasible non-stationary point.

LEMMA 2.8. *(H5) and (H6) imply (H4).*

Proof. Let $\bar{x}$ be a non-stationary feasible point, and let V_1 be the neighborhood defined by (H6). Since $\|x^k - z^k\| = O(h(x^k))$, there exists a neighborhood $\tilde{V}_1 \subset V_1$ of $\bar{x}$ such that for $x^k \in \tilde{V}_1$, $z^k \in V_1$. Consider an iterate x^k in $\tilde{V}_1$. By (H5) and (H6), there are positive constants M and N such that

$$\begin{aligned} f_0(x^k) - f_0(x^{k+1}) &= f_0(x^k) - f_0(z^k) + f_0(z^k) - f_0(x^{k+1}) \\ &\geq M\sqrt{H_k} - Nh(x^k) \\ &\geq \left(M - N\sqrt{h(x^k)}\right)\sqrt{H_k}, \end{aligned}$$

using the fact that $h(x^k) \leq H_k$. By continuity of h at $\bar{x}$, there exists a neighborhood $V \subset \tilde{V}_1$ such that for any $x \in V$, $\sqrt{h(x)} \leq 0.5M/N$. For any iterate x^k in this neighborhood, $f_0(x^k) - f_0(x^{k+1}) \geq 0.5M\sqrt{H_k}$, completing the proof. □

LEMMA 2.9. *Assume that (H5-H7) hold. Then given a feasible non-stationary point $\bar{x} \in X$, there exists a neighborhood V of $\bar{x}$ such that for any $x^k \in V$,*

$$f_0(x^k) - f_0(x^{k+1}) = \Omega(\|x^{k+1} - x^k\|).$$

Proof. Let V_1 and V_2 be the neighborhoods of a feasible non-stationary point $\bar{x}$ provided respectively by (H6) and (H7). As in the proof of Lemma 2.8, in some neighborhood $\tilde{V}_1 \subset V_1$ of $\bar{x}$, we have

$$f_0(x^k) - f_0(x^{k+1}) = f_0(x^k) - f_0(z^k) + f_0(z^k) - f_0(x^{k+1}),$$

with $|f_0(x^k) - f_0(z^k)| = O(h(x^k))$ by (2.9) and $f_0(z^k) - f_0(x^{k+1}) = \Omega(\sqrt{H^k})$. We easily deduce from these two facts that for x^k sufficiently near $\bar{x}$, say $x^k \in V_3 \subset \tilde{V}_1$, $|f_0(x^k) - f_0(z^k)| \leq 0.5(f_0(z^k) - f_0(x^{k+1}))$. It follows that

$$f_0(x^k) - f_0(x^{k+1}) \geq 0.5(f_0(z^k) - f_0(x^{k+1})). \tag{2.11}$$

We can also write

$$\|x^k - x^{k+1}\| \leq \|x^k - z^k\| + \|z^k - x^{k+1}\|,$$

with $\|x^k - z^k\| = O(h(x^k))$ by (H5) and $\|z^k - x^{k+1}\| = \Omega(f_0(z^k) - f_0(x^{k+1}))$, by the Lipschitz continuity of f_0. Again by the same reasoning as in the proof of last lemma, for $x^k \in \tilde{V}_2 \subset V_2$, $z^k \in V_2$ and we obtain from (H6) $\|z^k - x^{k+1}\| = \Omega(\sqrt{H_k})$. Like above, we deduce that for x^k sufficiently near $\bar{x}$, say $x^k \in V_4 \subset \tilde{V}_2$, $\|x^k - z^k\| \leq \|z^k - x^{k+1}\|$, and hence

$$\|x^k - x^{k+1}\| \leq 2\,\|z^k - x^{k+1}\|.$$

Finally, merging this expression with (2.11) in the neighborhood $V = V_3 \cap V_4$, we obtain

$$\begin{aligned} f_0(x^k) - f_0(x^{k+1}) &\geq 0.5(f_0(z^k) - f_0(x^{k+1})) \\ &= \Omega(\|z^k - x^{k+1}\|) \\ &= \Omega(\|x^k - x^{k+1}\|), \end{aligned}$$

completing the proof. □

3. Internal algorithms. In this section we discuss the internal steps used in each iteration of the main algorithm. We assume that Algorithm 2.1 has generated infinite sequences (x^k) and (z^k), and that Hypotheses H1-H3 are satisfied.

Feasibility step algorithm. The purpose of the feasibility phase is to find a point z^k such that $h(z^k) < (1-\alpha)h(x^k))$ and $z^k \notin \bar{\mathcal{F}}_k$. The procedure used in this phase could in principle be any iterative algorithm for decreasing h, and finite termination should be achieved because as we have seen above all filter entries $(f_0^j, h^j) \in F_k$ have $h^j > 0$.

The feasibility step studied by Martínez [10], satisfies assumption (H5) and applies directly to our case. Thus we shall not describe the feasibility procedure in detail in this paper.

Note that the feasibility algorithm may fail, if $h(\cdot)$ has an infeasible stationary point. In this case, the method stops with insuccess.

Optimality step algorithm. The optimality step must find x^{k+1} in the linearized set $L(z^k)$ such that $f_0(x^{k+1}) \leq f_0(z^k)$, and so that $x^{k+1} \notin \bar{\mathcal{F}}_k$. We shall describe a very general trust region method for this, and then show that the resulting step satisfies the assumptions (H6) and (H7).

The main tool for the analysis (not necessarily for the construction) of such algorithms is the projected Cauchy direction described in the introduction.

The projected gradient method. A very simple (but impractical) method for the tangential step is the following: from z^k, compute the projected Cauchy direction $d_c(z^k) = P_{L(z^k)}(z^k - \nabla f_0(z^k)) - z^k$, and perform an Armijo search along $z^k + \lambda d_c(z^k)$, $\lambda \in [0,1]$. The search must avoid forbidden points, which can be achieved by using in the Armijo search the objective $\theta(\lambda) = f_0(z^k + \lambda d_c(z^k))$ if $z^k + \lambda d_c(z^k) \notin \bar{\mathcal{F}}_k$, $\theta(\lambda) = +\infty$ otherwise.

We shall not prove the efficiency of this tangential step, because it is a particular case of the trust region iteration to be described from now on. The main requirement on the trust region step will be that it produces a point at least as good as the so called "Cauchy point", which lies on this projected gradient direction.

The quadratic model. Given $z^k \in X$ generated by Algorithm 2.1 in the feasibility phase, the trust region algorithm associates to z^k a quadratic model of f_0,

$$x \in \mathbb{R}^n \mapsto m_k(x) = f_0(z^k) + \nabla f_0(z^k)^T(x - z^k) + \frac{1}{2}(x - z^k)^T B_k (x - z^k), \tag{3.1}$$

where B_k is an $n \times n$ symmetric matrix. This matrix may be an approximation of $\nabla^2 f_0(z^k)$, or any other matrix, provided that the hypothesis (H8) below is verified. Usually, B_k will be an approximation of the Hessian of some Lagrangian function, and then m_k deviates from a straightforward model of f_0 by incorporating the curvature along the manyfold of the constraints. Although this may be essencial in the design of efficient algorithms, this discussion is out of the scope of this paper.

(H8) There exists $\beta > 0$ such that the quadratic model (3.1) satisfies $\|B_k\| \leq \beta$ for all $k \in \mathbb{N}$.

The trust region step uses a radius $\Delta > 0$ and computes a step $d(z^k, \Delta) \in \mathbb{R}^n$ such that $\|d(z^k, \Delta)\| \leq \Delta$. We define the *predicted reduction* produced by the step $d(z^k, \Delta)$ as

$$pred(z^k, \Delta) = m_k(z^k) - m_k(z^k + d(z^k, \Delta)), \tag{3.2}$$

and the *actual reduction* as

$$ared(z^k, \Delta) = f_0(z^k) - f_0(z^k + d(z^k, \Delta)). \tag{3.3}$$

LEMMA 3.1. *Consider $z^k \in X$ and $d(z^k, \Delta) \in \mathbb{R}^n$ generated by the trust region algorithm. Then*

$$ared(z^k, \Delta) = pred(z^k, \Delta) + o(z^k, \Delta), \tag{3.4}$$

where

$$\lim_{\Delta \to 0^+} \frac{o(z^k, \Delta)}{\Delta} = 0$$

uniformly in $z^k \in X$.

Proof. From (3.2)

$$\begin{aligned} -pred(z^k, \Delta) &= \nabla f_0(z^k)^T d(z^k, \Delta) + \tfrac{1}{2} d(z^k, \Delta)^T B_k d(z^k, \Delta) \\ &= \nabla f_0(z^k)^T d(z^k, \Delta) + O(\Delta^2) \end{aligned}$$

because $\|d(z^k, \Delta)\| \leq \Delta$ and $\|B_k\| \leq \beta$. From (3.3) and (1.3),

$$-ared(z^k, \Delta) = \nabla f_0(z^k)^T d(z^k, \Delta) + o(z^k, \Delta),$$

where

$$\lim_{\Delta \to 0^+} \frac{o(z^k, \Delta)}{\Delta} = 0.$$

This limit is uniform in $z^k \in X$ because $\nabla f_0(\cdot)$ is Lipschitz continuous in X. Hence

$$ared(z^k, \Delta) = pred(z^k, \Delta) + O(\Delta^2) - o(z^k, \Delta),$$

completing the proof. ☐

In the optimality step algorithm which we will discuss below we made the following choices which simplify the treatment:

(1) Each trust region computation starts with a radius $\Delta \geq \Delta_{\min}$ where $\Delta_{\min} > 0$ is fixed. The choice of Δ is irrelevant for the theory, and it usually comes from the former iteration.

(2) A step $d(z^k, \Delta)$ is only accepted if the sufficient decrease condition is satisfied:

$$ared(z^k, \Delta) > \eta \, pred(z^k, \Delta) \tag{3.5}$$

for a given $\eta \in (0, 1)$.

(3) The trust region computation solves approximately the problem

$$\begin{array}{ll} \text{minimize} & m_k(x) \\ \text{subject to} & x \in L(z^k) \\ & \|x - z^k\| \le \Delta, \end{array} \tag{3.6}$$

where $\|\cdot\|$ is any norm in $\mathbb{R}^n$.

Now we explain what we mean by "solving approximately". Given $z \in X$ and the set $L(z)$, the projected gradient direction is defined by

$$d_c(z) = P_{L(z)}(z - \nabla f_0(z)) - z. \tag{3.7}$$

Define

$$\varphi(z) = -\,\nabla f_0(z)^T \,\frac{d_c(z)}{\|d_c(z)\|}.$$

Then φ is the descent rate of f_0 along d_c. As usual, we denote $d_c^k = d_c(z^k)$, $\varphi^k = \varphi(z^k)$. As we saw in the Introduction, $\varphi(z) > 0$ whenever z is a feasible non-stationary point. point.

Now we use known results about the minimization of $m_k(\cdot)$ along a direction – see the discussion on the Cauchy point in [14]. Defining the generalized Cauchy point as the minimizer of $m_k(\cdot)$ along d_c, in the trust region $\{x \in \mathbb{R}^n \mid \|x - z^k\| \le \Delta\}$,

$$x_c = \operatorname{argmin}\left\{m_k(x) \mid \|x - z^k\| \le \Delta, x = z^k + \lambda d_c^k, \lambda \in [0,1]\right\},$$

we know that

$$m_k(z^k) - m_k(x_c) \ge \frac{\xi\varphi^k}{2}\,\min\left\{\frac{\varphi^k}{\|B_k\|},\ \|d_c^k\|,\ \Delta\right\},$$

where ξ depends on the norms used. Using the Hypothesis (H8), this can be rewritten as

$$m_k(z^k) - m_k(x_c) \ge \frac{\xi\varphi^k}{2}\,\min\left\{\frac{\varphi^k}{\beta},\ \|d_c^k\|,\ \Delta\right\}. \tag{3.8}$$

We accept as an approximate solution of (3.6), any feasible solution for this problem satisfying (3.8).

After stating the trust region step we shall study its properties.

Algorithm 3.2. *Optimality step.*

Data: $\eta \in (0,1)$, $\Delta_{\min} > 0$, $z^k \notin \bar{\mathcal{F}}_k$, $\Delta = \Delta^0 \ge \Delta_{\min}$.

REPEAT

Compute $d = d(z^k, \Delta)$ such that $\|d\| \le \Delta$, $z^k + d \in L(z^k)$ and

$$pred(z^k,\Delta) \ge \frac{\xi\varphi^k}{2}\min\left\{\frac{\varphi^k}{\|B_k\|}, \|d_c^k\|, \Delta\right\}.$$

Set $ared(z^k,\Delta) = f_0(z^k) - f_0(z^k + d)$.

if $z^k + d \notin \bar{\mathcal{F}}_k$ and $ared(z^k,\Delta) > \eta\, pred(z^k,\Delta)$

set $x^{k+1} = z^k + d$, $\Delta_k = \Delta$, and exit with success

else $\Delta = \Delta/2$.

Our task now is proving that this algorithm satisfies the assumptions (H6,H7) made for the optimality step.

LEMMA 3.3. *For any $z \in X$, $d \in \mathbb{R}^n$ such that $(z+d) \in L(z)$,*

$$|h(z+d) - h(z)| = O(\|d\|^2).$$

Proof. From (1.3), for any $z \in X$, $d \in \mathbb{R}^n$, $i = 0, 1, \ldots, m$

$$f_i(z+d) - f_i(z) \le \nabla f_i(z)^T d + O(\|d\|^2).$$

Since $(z+d) \in L(z)$, by definition of $L(z)$ given in (1.4), we have for $i = 1, \ldots, m$

$$f_i(z) + \nabla f_i(z)^T d \le f_i^+(z),$$

and hence

$$f_i(z+d) \le f_i^+(z) + O(\|d\|^2).$$

We must prove that

$$f_i^+(z+d) \le f_i^+(z) + O(\|d\|^2). \tag{3.9}$$

If $f_i(z+d) < 0$ and $i \in \mathcal{I}$ this is true because the right hand side is positive. Otherwise $f_i^+(z+d) = f_i(z+d)$. Using (3.9) in the norm definition we set

$$\|f^+(z+d)\| = \|f^+(z)\| + O(\|d\|^2),$$

completing the proof. □

Now we study the optimality step near a non-stationary feasible point $\bar{x} \in X$. The first lemma says that if we ignore the filter, then the trust region step is large near $\bar{x}$.

LEMMA 3.4. *Let $\bar{x} \in X$ be a feasible non-stationary point satisfying a M-F condition. Then there exist a neighborhood $\tilde{V}$ of $\bar{x}$, $\tilde{\Delta} \in (0, \Delta_{\min})$ and a constant $\tilde{c} > 0$ such that for any $z^k \in \tilde{V}$,*

(i) for any $\Delta > 0$, $pred(z^k, \Delta) \ge \tilde{c}\ \min\{\Delta, \tilde{\Delta}\}$,

(ii) for any $\Delta \in (0, \tilde{\Delta})$, $ared(z^k, \Delta) > \eta\, pred(z^k, \Delta) \ge \eta \tilde{c}\, \Delta$.

Proof. From the generalized Cauchy decrease condition (3.8), which is satisfied by construction at each iteration,

$$pred(z^k, \Delta) \ge \frac{\xi \varphi(z^k)}{2} \min\left\{\frac{\varphi(z^k)}{\beta},\ \|d_c^k\|,\ \Delta\right\}.$$

From Lemma 1.1 we deduce that $z \mapsto \|d_c(z)\|$ and $z \mapsto \varphi(z) = -\nabla f_0(z)^T \dfrac{d_c(z)}{\|d_c(z)\|}$ are continuous at $\bar{x}$. Hence there exists a neighborhood $\tilde{V}$ of $\bar{x}$ such that for $z^k \in \tilde{V}$, $\varphi(z^k) \ge \varphi(\bar{x})/2$ and $\|d_c(z^k)\| \ge \|d_c(\bar{x})\|/2$. Thus, for $z^k \in \tilde{V}$,

$$pred(z^k, \Delta) \ge \frac{\xi \varphi(\bar{x})}{4} \min\left\{\frac{\varphi(\bar{x})}{2\beta},\ \frac{\|d_c(\bar{x})\|}{2},\ \Delta\right\}.$$

This can be written as $pred(z^k, \Delta) \ge \tilde{c}\ \min\{\Delta_1, \Delta\}$, proving (i).

From Lemma 3.1, for any $k \in \mathbb{N}$ and $\Delta > 0$,

$$\begin{array}{rcl} ared(z^k, \Delta) & = & pred(z^k, \Delta) + o(z^k, \Delta) \\ & = & \eta\, pred(z^k, \Delta) + (1-\eta)\, pred(z^k, \Delta) + o(z^k, \Delta), \end{array}$$

where $\lim_{\Delta \to 0^+} \dfrac{o(z^k, \Delta)}{\Delta} = 0$ uniformly in z^k. For $\Delta \leq \Delta_1$, $pred(z^k, \Delta) \geq \tilde{c}\, \Delta$ and then

$$ared(z^k, \Delta) \geq \eta\, pred(z^k, \Delta) + (1-\eta)\tilde{c}\, \Delta + o(z^k, \Delta).$$

For Δ sufficiently small, say, $\Delta \leq \tilde{\Delta} \leq \Delta_1$, $(1-\eta)\tilde{c}\, \Delta + o(z^k, \Delta) \geq 0$, completing the proof. □

Algorithm 3.2 in iteration k starts with $\Delta^0 \geq \Delta_{\min}$ and iterates by setting $\Delta^j = 2^{-j}\Delta^0$, $j = 0, 1, \cdots$ and computing for each Δ^j the step $d(z^k, \Delta^j)$. Whenever $\Delta^j < \tilde{\Delta}$, the condition $ared(z^k, \Delta^j) > \eta\, pred(z^k, \Delta^j)$ is satisfied: the radius Δ^j can only be rejected if $z^k + d(z^k, \Delta^j) \in \bar{\mathcal{F}}_k$.

By construction, $z^k \notin \bar{\mathcal{F}}_k$. Since $\bar{\mathcal{F}}_k$ is a closed set, for Δ^j sufficiently small, $z^k + d(z^k, \Delta^j) \notin \bar{\mathcal{F}}_k$. This shows that the algorithm always terminates.

LEMMA 3.5. *Let $\bar{x} \in X$ be a feasible non-stationary point satisfying a M-F condition, and assume that* (2.7) *holds. Then there exists a neighborhood V of $\bar{x}$ such that for $x^k \in V$,*

$$f_0(z^k) - f_0(x^{k+1}) = \Omega(\sqrt{H_k}), \tag{3.10}$$

$$f_0(z^k) - f_0(x^{k+1}) = \Omega(\|x^{k+1} - z^k\|), \tag{3.11}$$

where $x^{k+1} = z^k + d(z^k, \Delta)$ is computed by Algorithm 3.2.

Proof. By a usual argument, it is enough to prove that for any subsequence $(x^k)_{k \in \mathcal{K}}$ converging to $\bar{x}$, (3.10) and (3.11) are true for large $k \in \mathcal{K}$.

Assume that $x^k \xrightarrow{\mathcal{K}} \bar{x}$, where $\mathcal{K} \subset \mathbb{N}$. It follows that $z^k \xrightarrow{\mathcal{K}} \bar{x}$, because by (2.8) $\|x^k - z^k\| = O(h(x^k)) \xrightarrow{\mathcal{K}} 0$.

Let $\tilde{V} \subset X$ and $\tilde{\Delta} > 0$ be the neighborhood of $\bar{x}$ and radius given by Lemma 3.4. For large $k \in \mathcal{K}$, say $k \in \mathcal{K}_1 \subset \mathcal{K}$, $z^k \in \tilde{V}$. Let us now consider an iteration $k \in \mathcal{K}_1$, and denote $(\tilde{f}_0, \tilde{h}) = (f_0(x^k) - \alpha h(x^k), (1-\alpha)h(x^k))$ the temporary entry in the filter.

The Algorithm 3.2 starts with a radius $\Delta^0 \geq \Delta_{\min}$ and computes $d(z^k, \Delta^j)$, $\Delta^j = 2^{-j}\Delta^0$, $j = 0, 1, \cdots$, until $z^k + d(z^k, \Delta^j) \notin \bar{\mathcal{F}}_k$ and $ared(z^k, \Delta^j) > \eta\, pred(z^k, \Delta^j)$. Then $\Delta_k = \Delta^j$. Let us define $\hat{\Delta}$ as the first Δ^j such that

$$ared(z^k, \Delta^j) > \eta\, pred(z^k, \Delta^j) \quad \text{and} \tag{3.12}$$

$$z^k + d(z^k, \Delta^j) \notin \bar{\mathcal{F}}_k \quad \text{or} \quad f_0(z^k + d(z^k, \Delta^j)) \geq \tilde{f}_0. \tag{3.13}$$

Let us denote $\hat{d} = d(z^k, \hat{\Delta})$ and $\hat{x} = z^k + \hat{d}$. Note that $\hat{\Delta} \geq \Delta_k$, and $\hat{\Delta} > \Delta_k$ happens only when $f_0(\hat{x}) \geq \tilde{f}_0$. We shall derive properties of this step $\hat{d}$, and then prove that this situation cannot occur when x^k is sufficiently near $\bar{x}$.

We shall first prove that $\hat{x}$ satisfies the bounds in the lemma.
Choose $\bar{\Delta} \leq \tilde{\Delta}/2$.

(i) First, the easy case: assume that $\hat{\Delta} \geq \bar{\Delta}$. Then by Lemma 3.4,

$$pred(z^k, \hat{\Delta}) \geq \tilde{c} \min\{\hat{\Delta}, \tilde{\Delta}\} \geq \tilde{c}\bar{\Delta}.$$

By definition of $\hat{\Delta}$, (3.12) holds, and hence

$$f_0(z^k) - f_0(\hat{x}) \geq \eta\tilde{c}\bar{\Delta} = \Omega(1).$$

It follows trivially that $f_0(z^k) - f_0(\hat{x}) = \Omega(\sqrt{H_k})$ and $f_0(z^k) - f_0(\hat{x}) = \Omega(\|x^k - \hat{x}\|)$, because in both cases the right hand side is bounded in X.

(*ii*) Now, assume that $\hat{\Delta} < \bar{\Delta}$. Then the radius $2\hat{\Delta} < 2\bar{\Delta} \leq \tilde{\Delta} < \Delta_{\min}$ does not satisfy (3.13) (and was rejected by Algorithm 3.2). By Lemma 3.4,

$$ared(z^k, d(z^k, 2\hat{\Delta})) > \eta\, pred(z^k, d(z^k, 2\hat{\Delta})),$$

and it follows from (3.13) that $z^k + d(z^k, 2\hat{\Delta}) \in \bar{\mathcal{F}}$ and $f_0(z^k + d(z^k, 2\hat{\Delta})) < \tilde{f}_0$. By definition of H_k, we must have $h(z^k + d(z^k, 2\hat{\Delta})) \geq H_k$.

By construction, $h(z^k) < (1-\alpha)h(x^k) \leq (1-\alpha)H_k$. Thus,

$$h(z^k + d(z^k, 2\hat{\Delta})) - h(z^k) \geq \alpha H_k.$$

By Lemma 3.3,

$$h(z^k + d(z^k, 2\hat{\Delta})) - h(z^k) = O(\|d(z^k, 2\hat{\Delta})\|^2) = O(\hat{\Delta}^2), \tag{3.14}$$

because $\|d(z^k, 2\hat{\Delta})\| \leq 2\hat{\Delta}$. Merging these two results, we obtain $\alpha H_k \leq O(\hat{\Delta}^2)$, or

$$\hat{\Delta} = \Omega(\sqrt{H_k}). \tag{3.15}$$

Using again Lemma 3.4 with $\hat{\Delta} < \bar{\Delta} < \tilde{\Delta}$,

$$f_0(z^k) - f_0(\hat{x}) \geq \eta\tilde{c}\,\Omega(\sqrt{H_k}) = \Omega(\sqrt{H_k}), \tag{3.16}$$

$$f_0(z^k) - f_0(\hat{x}) \geq \eta\tilde{c}\hat{\Delta} = \Omega(\hat{\Delta}). \tag{3.17}$$

So, the step $\hat{d}$ satisfies the conditions in the Lemma.

To finish the proof, we must show that for large $k \in \mathcal{K}_2$, $f_0(\hat{x}) < \tilde{f}_0$, which implies $\hat{x} \notin \bar{\mathcal{F}}_k$, and thus $\hat{x} = x^{k+1}$.

From (3.6) and (2.9), there are positive constants M and N such that

$$\begin{aligned} f_0(\hat{x}) &\leq f_0(z^k) - M\sqrt{H_k} \\ f_0(z^k) &\leq f_0(x^k) + Nh(x^k). \end{aligned}$$

Adding these expressions, we get $f_0(\hat{x}) \leq f_0(x^k) - M\sqrt{H_k} + Nh(x^k)$. It is immediate to check that for $k \in \mathcal{K}_2$ such that $\sqrt{h(x^k)} < M/(N+\alpha)$ (say, $k \in \mathcal{K}_3$), $f_0(\hat{x}) < f_0(x^k) - \alpha h(x^k) = \tilde{f}_0$, completing the proof. ∎

4. Improvements. In this section we present three improvements to our treatment. First, we improve the convergence analysis, by showing that under hypotheses (H5,H6) the objective function values always converge, thus limiting very much the possibility of reaching non-stationary accumulation points, specially when (H7) is added. Then we show how a small change in the master algorithm totally precludes the possibility of generating non-stationary accumulation points. Third, we discuss a simplified optimality step, using the Jacobian matrices already calculated in the feasibility phase instead of the ones for z^k.

4.1. Convergence of the objective function values. We shall continue the analysis of sequences (x^k) generated by the algorithm made in Sec. 2. We start by showing that f_0 cannot grow much in a single iteration.

LEMMA 4.1. *Assume that hypotheses* $(H5, H6)$ *hold. Then there exists a constant* $M > 0$ *such that in any iteration* k*,*

$$f_0(x^{k+1}) \leq f_0(x^k) + Mh(x^k).$$

Proof. Note that $f_0(\cdot)$ can only grow in an h-iteration. From (2.9), there exists a constant $M > 0$ such that in any iteration k, $f_0(z^k) \leq f_0(x^k) + Mh(x^k)$. By construction, $f(x^{k+1}) \leq f(z^k)$, completing the proof. □

Now we show that f_0 cannot grow much in a sequence of iterations.

LEMMA 4.2. *Assume that hypotheses* $(H5, H6)$ *hold. Consider a finite sequence of iterations* $I = \{\bar{k}, \bar{k}+1, \dots, K\}$ *such that for* $k \in I$, $f^k \equiv f_0(x^k) \geq f_0(x^{\bar{k}})$[1]*. Then*

$$f^K \leq f^{\bar{k}} + \frac{M}{\alpha} h(x^{\bar{k}}).$$

Proof.

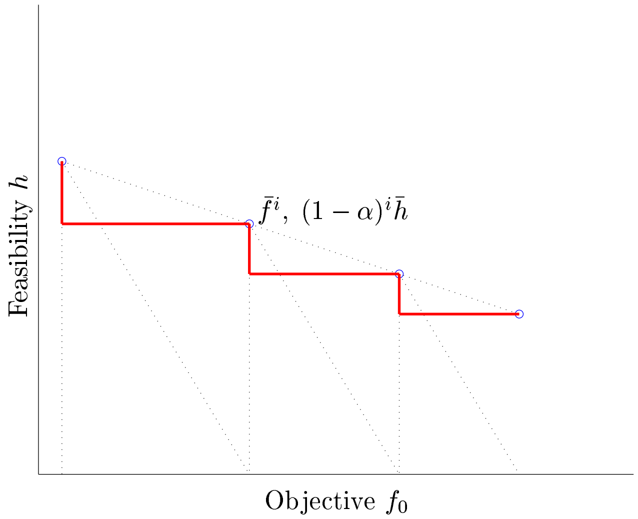

FIG. 4.1. *Geometry of the values used in the proof of Lemma 4.2*

Let us denote $f^k = f_0(x^k)$, $h^k = h(x^k)$, for $k \in I$ and $\bar{h} = h^{\bar{k}}$. Let us also define the following values, depicted in Fig 4.1:

$$\begin{aligned}
\bar{f}_0 &= f^{\bar{k}} \\
\bar{f}_1 &= \bar{f}_0 + M\bar{h} \\
\bar{f}_2 &= \bar{f}_1 + M(1-\alpha)\bar{h} = \bar{f}_0 + [1 + (1-\alpha)]M\bar{h} \\
\bar{f}_j &= \bar{f}_0 + (\sum_{i=0}^{j-1} (1-\alpha)^i) M\bar{h} \leq \bar{f}_0 + \frac{M}{\alpha}\bar{h}.
\end{aligned}$$

[1]Note that (f^k) is not necessarily increasing, but $\bar{k}$ is an h-iteration.

We show the following: the sequence $(f^k)_{k\in I}$ has at least one element in each interval $[\bar{f}_j, \bar{f}_{j+1}]$, $j = 0, 1, \ldots$, until its last element. Consequently f^K will be smaller than $\bar{f}_0 + M\bar{h}/\alpha$.
– First interval: the iteration $\bar{k}$ is an h-iteration. The pair $(\bar{f}_0 - \alpha\bar{h}, (1-\alpha)\bar{h})$ enters the permanent filter, and hence $h^k \leq (1-\alpha)\bar{h}$ for $k = \bar{k}+1, \ldots, K$ and $f^{\bar{k}+1} \leq \bar{f}^0 + M\bar{h} = \bar{f}^1$ by Lemma 4.1.
– Second interval: Let k_1 be the largest $k \in I$ such that $f^k \leq \bar{f}_1$ (several h-iterations and f_0-iterations may have occurred between $\bar{k}$ and k_1). Now k_1 is an h-iteration, and like for the first interval,

$$(f^{k_1} - \alpha h^1, (1-\alpha)h^1) \leq (\bar{f}^1, (1-\alpha)^2\bar{h})$$

enters the filter. Hence $h^k \leq (1-\alpha)^2\bar{h}$ for $k = k_1+1, \ldots, K$ and $\bar{f}_1 \leq f^{k_1+1} \leq \bar{f}^{k_1} + M(1-\alpha)\bar{h} \leq \bar{f}^2$ by Lemma 4.1.
Following the same process, we detect an h-iteration k_2, the last in the second interval, and f^{k_2+1} will be in the third interval, and so on until f^K is reached. Then $f^K \leq \bar{f}_0 + M\bar{h}/\alpha$ completing the proof. ☐

We can now prove the main result in this analysis:

THEOREM 4.3. *Assume that hypotheses $(H5, H6)$ hold. Then the sequence $(f_0(x^k))$ converges.*

Proof. Let us denote $f^k \equiv f_0(x^k)$ for $k \in \mathbb{N}$. The sequence (f^k) is bounded by hypothesis. We shall use the following fact, which is a simple exercise in sequences: Given a sequence (f^k) such that $\limsup(f^k) > \liminf(f^k) + \delta$, $\delta > 0$, it is possible to extract two subsequences $(f^k)_{k\in\mathcal{K}}$ and $(f^{k+j_k})_{k\in\mathcal{K}}$, $\mathcal{K} \subset \mathbb{N}$ such that for any $k \in \mathcal{K}$,

$$\begin{aligned} f^{k+j_k} &\geq f^k + \delta \\ f^{k+r} &\geq f^k \text{ for } r = 1, \ldots, j_k. \end{aligned}$$

In fact, to prove this fact it is enough to take a subsequence convergent to $\limsup(f^k)$ and associate with each index (say, l) the last index $l - j_l$ such that $f^{l-j_l} \leq f^l - \delta$, if it exists. For large l, the construction will always be well-defined.

Assume by contradiction that $\limsup(f^k) > \liminf(f^k) + \delta$, for some $\delta > 0$, and let the subsequence $(f^k)_{k\in\mathcal{K}}$ be given by the construction above. Then from Lemma 4.2, we conclude that for all $k \in \mathcal{K}$, the iteration k is an h-iteration and

$$f^k + \delta \leq f^{k+j_k} \leq f^k + \frac{M}{\alpha}h(x^k). \tag{4.1}$$

Taking subsequences if necessary, assume that $(x^k)_{k\in\mathcal{K}}$ converges to a point $\bar{x}$. Then $\bar{x}$ must be stationary by Lemma 2.5, and consequently $h(x^k) \xrightarrow{\mathcal{K}_1} 0$. This contradicts (4.1), completing the proof. ☐

Now we incorporate the hypothesis (H7), and show that near a feasible non-stationary point the objective function always changes by a large amount, precluding the possibility of feasible non-stationary accumulation points.

LEMMA 4.4. *Assume that hypotheses (H5-H7) hold. Let $\bar{x} \in X$ be a feasible non-stationary point. Then there exist a neighborhood V of $\bar{x}$ and $\delta > 0$ such that whenever $x^k \in V$, there exists $l_k \in \mathbb{N}$ such that*

$$f_0(x^{k+l_k}) \leq f_0(x^k) - \delta. \tag{4.2}$$

Proof. From (H4) and Lemma 2.9, there exist a neighborhood V_1 of $\bar{x}$ and constants $\beta_1, \beta_2 > 0$ such that for all $x^k \in V_1$,

$$f_0(x^k) - f_0(x^{k+1}) \geq \beta_1 \, \|x^{k+1} - x^k\|, \tag{4.3}$$

$$f_0(x^k) - f_0(x^{k+1}) \geq \beta_2 \, \sqrt{H_k}, \tag{4.4}$$

and the iteration k is an f_0-iteration.

Consider $\epsilon > 0$ such that $\mathcal{B}_\epsilon(\bar{x}) = \{x \in \mathbb{R}^n \mid \|x - \bar{x}\| < \epsilon\} \subset V_1$, and define $V = \mathcal{B}_{\epsilon/2}(\bar{x})$.

Let $k \in \mathbb{N}$ be such that $x^k \in V$. While x^{k+i}, $i = 1, 2, \ldots$ remain in $\mathcal{B}_\epsilon(\bar{x})$, the iterations $(k+i)$ are f_0-iterations, and the filter does not change, i.e.,

$$F_{k+i} = F_k \qquad \text{and} \qquad \mathcal{F}_{k+i} = \mathcal{F}_k \qquad \text{for } i = 1, 2, \ldots.$$

Consequently, from (4.4) f_0 decreases by at least the constant amount $\beta_2\sqrt{H_k}$. Hence, there exists a finite $l_k \in \mathbb{N}$ such that $x^{k+l_k} \notin \mathcal{B}_\epsilon(\bar{x})$, $x^{k+i} \in \mathcal{B}_\epsilon(\bar{x})$, for $i = 0, 1, \ldots, l_k - 1$. We have

$$\|x^{k+l_k} - x^k\| \geq \frac{\epsilon}{2} \tag{4.5}$$

because $x^k \in \mathcal{B}_{\epsilon/2}(\bar{x})$. Using (4.3), (4.5) and the triangle inequality,

$$\begin{aligned}
f_0(x^k) - f_0(x^{k+l_k}) &= \sum_{i=0}^{l_k-1} f_0(x^{k+i}) - f_0(x^{k+i+1}) \\
&\geq \beta_1 \sum_{i=0}^{l_k-1} \|x^{k+i+1} - x^{k+i}\| \\
&\geq \beta_1 \, \|x^k - x^{k+l_k}\| \\
&\geq \beta_1 \epsilon/2,
\end{aligned}$$

completing the proof. ∎

It is now trivial to prove (and this will be done in moment) that feasible non-stationary point cannot be accumulation points of the sequence. The presence of non-stationary accumulation points is then reduced to a single seemingly unreasonable possibility: there must exist an infeasible accumulation point, reached by large jumps from points arbitrarily near a stationary solution, and the objective values must converge. We now show how a simple change in the algorithm precludes this possibility.

4.2. The modified algorithm. The only change is in the criterion used to introduce a point in the filter, which now becomes:

Filter update: Given $\varepsilon > 0$
  if $f_0(x^{k+1}) < f_0(x^k) - \min\{(h(x^k))^2, \varepsilon\}$ then
    $F_{k+1} = F_k$, $\mathcal{F}_{k+1} = \mathcal{F}_k$ (f_0-iteration)
  else
    $F_{k+1} = \bar{F}_k$, $\mathcal{F}_{k+1} = \bar{\mathcal{F}}_k$ (h-iteration)

This implies that potentially more points will be introduced in the filter.

Near feasible non-stationary points, the criterion for entering the filter becomes

$$f_0(x^k) - f_0(x^{k+1}) \leq (h(x^k))^2 = o(h(x^k)) = o(H_k).$$

The term $o(H_k)$ vanishes when added to $f_0(x^k) - f_0(x^{k+1}) = \Omega(\sqrt{H_k})$, and hence Lemma 2.5 remains true. Lemma 2.6 and Theorem 2.7 also remain true; it is immediate to check that the same proofs apply to the modified algorithm. Hence all the results of Sec. 2 remain valid, and the sequence has a stationary accumulation point.

THEOREM 4.5. *Any accumulation point of* (x^k) *is stationary.*

Proof. By contradiction, assume that $x^k \xrightarrow{\mathcal{K}} \bar{x}$, $\bar{x}$ non-stationary. From Lemma 2.5, we know that for large $k \in \mathcal{K}$ all iterations are f_0 iterations.

If $h(\bar{x}) > 0$ then for large $k \in \mathcal{K}$, $h(x^k) > h(\bar{x})/2$ and hence $f_0(x^{k+1}) \le f_0(x^k) - \delta_1$, with $\delta_1 = \min\{\varepsilon, (h(\bar{x}))^2/4\} > 0$.

If $h(\bar{x}) = 0$ then Lemma 4.4 ensures that for large $k \in \mathcal{K}$, there exist $\delta_2 > 0$ and $j_k \in \mathbb{N}$ such that

$$f_0(x^{k+j_k}) \le f_0(x^k) - \delta_2.$$

In any case, we construct a subsequence $(x^{k+j_k})_{k \in \mathcal{K}}$ such that for $k \in \mathcal{K}$

$$f_0(x^{k+j_k}) \le f_0(x^k) - \delta,$$

where $\delta = \min\{\delta_1, \delta_2\} > 0$.

It follows that the sequence $f_0(x^k)$ is not a Cauchy sequence, contradicting Theorem 4.3 and completing the proof. □

4.3. The simplified tangential step. In our algorithm the feasibility and tangential steps are independent. This means that the Jacobians $A_{\mathcal{E}}$ and $A_{\mathcal{I}}$ must be calculated both at x^k and z^k. In most algorithms based on feasibility and optimality steps, the tangential step uses at z^k the linear model computed at x^k, reducing the computations.

This makes sense if x^k is near z^k, and if the feasibility algorithm has taken only one step to reach z^k from x^k. If multiple steps were used, then the tangential step can be simplified by approximating the Jacobians by the last ones computed in the feasibility procedure.

We shall now change the tangential step and use the following maps, which associate with each $(z, x) \in \mathbb{R}^{2n}$ the set

$$L(z,x) = \{y \in \mathbb{R}^n \mid A_{\mathcal{E}}(x)\,(y - z) = 0,\ f_{\mathcal{I}}(z) + A_{\mathcal{I}}(x)\,(y - z) \le f_{\mathcal{I}}^{+}(z)\} \tag{4.6}$$

and the point

$$d_c(z,x) = P_{L(z,x)}(z - \nabla f_0(z)) - z. \tag{4.7}$$

So, $L(z^k, x^k)$ is the same as $L(z^k)$ given by (1.4), with $A_{\mathcal{E}}(z^k)$, $A_{\mathcal{I}}(z^k)$ replaced by $A_{\mathcal{E}}(x^k)$, $A_{\mathcal{I}}(x^k)$. Similarly, the projected gradient direction is now projected into $L(z^k, x^k)$.

When $x^k \xrightarrow{\mathcal{K}} \bar{x}$, $\bar{x} \in X$ feasible, $\mathcal{K} \subset \mathbb{N}$, it is also true that $z^k \xrightarrow{\mathcal{K}} \bar{x}$, because by (2.8) $\|x^k - z^k\| = O(h(x^k))$. So, we need the continuity of (4.6) and (4.7) at a pair $(\bar{x}, \bar{x})$: this is guaranteed by straightforward changes in the proof of Lemma A.1.

The only change in the treatment is in Lemma 3.3, which now becomes:

LEMMA 4.6. *For any* $z, x \in X$*, such that* $\|z - x\| = O(h(x))$ *and* $d \in \mathbb{R}^n$ *such that* $(z + d) \in L(z, x)$*,*

$$|h(z + d) - h(z)| = h(x)\,O(\|d\|) + O(\|d\|^2).$$

Proof. From (1.3), for any $z \in X$, $d \in \mathbb{R}^n$, $i = 0, 1, \ldots, m$

$$f_i(z+d) - f_i(z) \leq \nabla f_i(z)^T d + O(\|d\|^2).$$

By the Lipschitz condition on ∇f_i, $i = 0, 1, \ldots, m$,

$$\|\nabla f_i(z) - \nabla f_i(x)\| = O(\|z - x\|) = O(h(x))$$

and

$$\begin{aligned} \nabla f_i(z)^T d &= \nabla f_i(x)^T d + O(h(x))\, \|d\| \\ &= \nabla f_i(x)^T d + h(x)\, O(\|d\|). \end{aligned}$$

Now, proceeding as in the proof of Lemma 3.3, we get

$$\|f^+(z+d)\| = \|f^+(z)\| + h(x)\, O(\|d\|) + O(\|d\|^2),$$

completing the proof. ∎

Lemma 3.4 is not affected by the changes, and we only need to change the proof of Lemma 3.5.

The only place where Lemma 3.3 is used in the proof is in the expressions (3.14) and (3.15). We will show now that with a good choice of the value $\bar{\Delta}$ (defined in the beginning of the proof), (3.15) is still valid.

Let us modify (3.14), using Lemma 4.4:

$$\alpha H_k \leq h(z^k + d(z^k, 2\hat{\Delta})) - h(z^k) = h(x^k)\, O(\|d(z^k, 2\hat{\Delta})\|) + O(\|d(z^k, 2\hat{\Delta})\|^2).$$

Since $d(z^k, 2\hat{\Delta}) \leq 2\hat{\Delta}$ and $h(x^k) < H_k$,

$$H_k \leq c H_k \hat{\Delta} + O(\hat{\Delta}^2),$$

where $c > 0$ is a constant dependent only on $\bar{x}$. For a choice of $\bar{\Delta}$ so that $c\bar{\Delta} < 1/2$, we obtain

$$\frac{1}{2} H_k = O(\hat{\Delta}^2),$$

which implies (3.15).

From this point on, the proof identical, proving that the simplified algorithm has the same convergence properties as the original one.

5. A graphical example. In this section we present a graphical example of the mechanics of the algorithms. Consider the bidimensional problem

$$\begin{array}{ll} \text{minimize} & x_2 \\ \text{subject to} & f(x) = x_2 + (2 + x_1) cos(x_1) = 0. \end{array}$$

Figure 5.1 (as well as all figures to follow) shows the level curves of $h(x) = |f(x)|$ and a local minimizer. The figure at the right shows the pairs $(f(x), h(x))$.

Using a merit function $\psi(x) = f(x) + \nu h(x)$ with $\nu = 0.5$, the figure shows the forbidden points associated with the point $(1, -4)$, i.e., the points x such that $\psi(x) \geq \psi((1, -4))$. Notice that the local optimizer is forbidden for this value of ν. This happens because ν is too small, smaller than the KKT multiplier at the optimum.

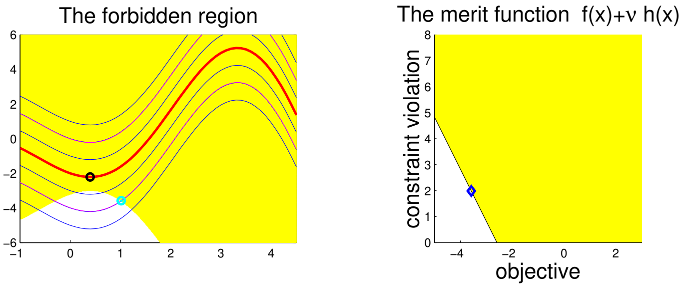

FIG. 5.1. *The merit function forbids the local optimizer.*

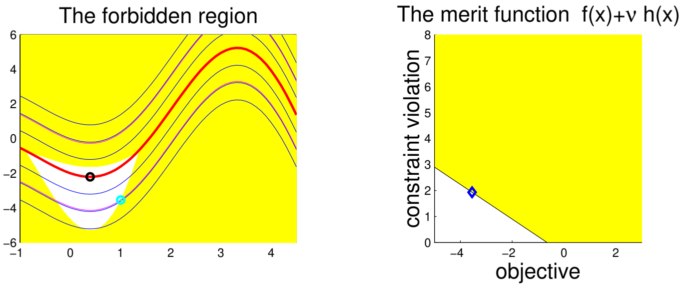

FIG. 5.2. *Now the optimizer is not forbidden.*

Figure 5.2 shows the same situation for $\nu = 1.5$, and now the local optimizer is never forbidden. This is actually true for any value of $\nu \geq 1$, the value of the optimal multiplier.

The following figures show some iterations of the filter method, programmed in Matlab and using internal algorithms which are intentionally imprecise but satisfy all the hypotheses. Figure 5.3 shows the first iteration. On the left, the temporarily forbidden region associated with the first iterate, and a feasibility step followed by a tangential step. The figure at the right shows the filter: now $F_0 = \emptyset$, and $\bar{F_0}$ contains only the point $(f_0(x^0) - \alpha h(x^0), (1-\alpha)h(x^0))$. The pairs resulting from the feasibility and tangential steps are also shown. For the tangential step, we show the pairs corresponding to $z^k + \lambda(x^{k+1} - z^k)$, $\lambda \in [0,1]$.

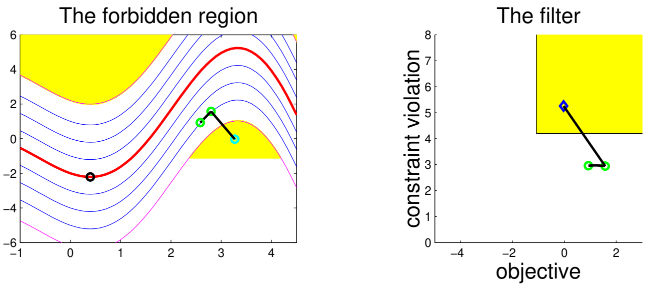

FIG. 5.3. *First iteration of a filter method.*

The first iteration was an h-iteration, because $f_0(x^1) > f_0(x^0)$. So, $(f_0(x^0), h(x^0))$ becomes a permanent entry in the filter. Figure 5.4 shows the second iteration, where the permanently forbidden points and pairs are in the darker region.

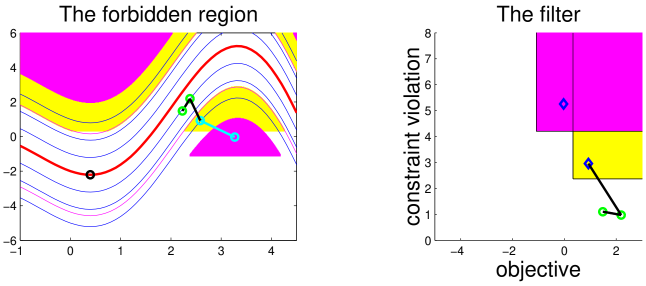

FIG. 5.4. *Temporarily and permanently forbidden points after the first iteration.*

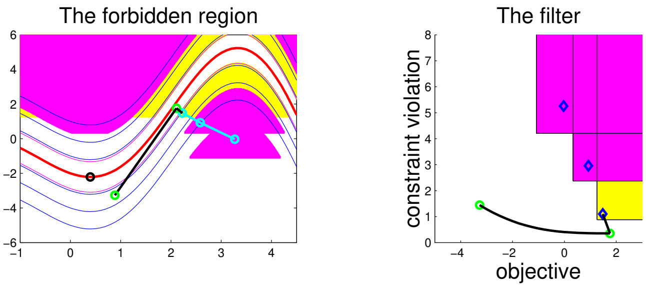

FIG. 5.5. *Third iteration: an f_0-iteration.*

The second iteration was also an h-iteration, and the permanent filter has two points. The third iteration is an f_0-iteration.

After one more h-iteration (iteration 4), filter entries dominated by the new one can be eliminated from the filter. The last figure shows the fifth iteration.

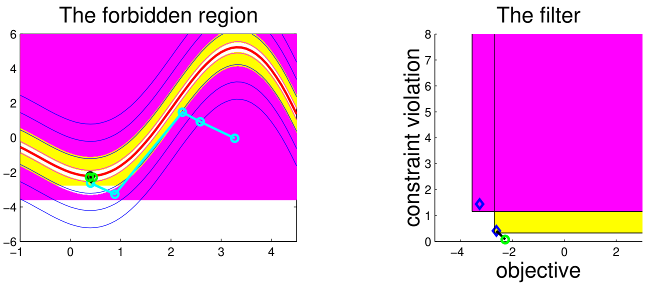

FIG. 5.6. *Fifth iteration: two filter elements were eliminated.*

Appendix. Continuity properties of maps.

Let $L : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ be the map defined in (1.4),

$$z \in \mathbb{R}^n \mapsto L(z) = \{x \in \mathbb{R}^n \mid A_{\mathcal{E}}(z)(x - z) = 0, f_{\mathcal{I}}(z) + A_{\mathcal{I}}(z)(x - z) \le f_{\mathcal{I}}^{+}(z)\}$$

where $z \mapsto A_{\mathcal{E}}(z)$ and $z \mapsto A_{\mathcal{I}}(z)$ are continuous.

We say that $x \in \mathbb{R}^n$ is an interior point of $L(z)$ if $x \in L(z)$ and $f_{\mathcal{I}}(z) + A_{\mathcal{I}}(z)(x - z) < f_{\mathcal{I}}^+(z)$.

LEMMA A.1. *Consider $\bar{z} \in \mathbb{R}^n$ such that $A_{\mathcal{E}}(\bar{z})$ has linearly independent rows and $L(\bar{z})$ has an interior point (i.e, the M-F qualification condition is satisfied at $\bar{z}$). Then the point to set map $L(\cdot)$ is continuous at $\bar{z}$.*

Proof. Consider a sequence (z^k) such that $z^k \to \bar{z}$, and the sets $L(z^k)$.

(1) Upper semi-continuity: Let $x^k \in L(z^k)$, $k \in \mathbb{N}$, be such that $x^k \to \bar{x}$. Using the continuity of all functions involved in the definition of $L(\cdot)$, the fact that $\bar{x} \in L(\bar{z})$ is straightforward.

(2) Lower semi-continuity: Consider an arbitrary point $\bar{x} \in L(\bar{z})$. We must exhibit a sequence $x^k \in L(z^k)$, $k \in \mathbb{N}$, such that $x^k \to \bar{x}$.

Define $x^k = P_{L(z^k)}(\bar{x})$, where $P_\Gamma(w)$ denotes the orthogonal projection of $w \in \mathbb{R}^n$ onto the closed set $\Gamma \subset \mathbb{R}^n$.

By contradiction assume that there exist an infinite set $\mathcal{K} \subset \mathbb{N}$ and $\varepsilon > 0$ such that for all $k \in \mathcal{K}$, $\|x^k - \bar{x}\| > \varepsilon$. We shall establish a contradiction by obtaining $k \in \mathcal{K}$ and a point $w^k \in L(z^k)$ such that $\|w^k - \bar{x}\| < \varepsilon$.

Let $y \in \mathbb{R}^n$ be an interior point of $L(\bar{z})$. Then for any $\lambda \in (0, 1)$,

$$y_\lambda = \lambda\, y + (1 - \lambda)\bar{x}$$

is an interior point of $L(\bar{z})$. Choose λ such that $\|y_\lambda - \bar{x}\| < \varepsilon/2$, and define w^k as the projection of y_λ onto $\{x \in \mathbb{R}^n \mid A_{\mathcal{E}}(z^k)(x - z^k) = 0\}$. For z^k sufficiently near $\bar{z}$, $A_{\mathcal{E}}(z^k)$ has linearly independent rows, and the projection is given by

$$(y_\lambda - w^k) = A_{\mathcal{E}}(z^k)^T \left(A_{\mathcal{E}}(z^k) A_{\mathcal{E}}(z^k)^T\right)^{-1} A_{\mathcal{E}}(z^k)(y_\lambda - z^k).$$

The projection is continuous at $\bar{z}$, and hence $y_\lambda - w^k \to 0$, because $A_{\mathcal{E}}(\bar{z})(y_\lambda - \bar{z}) = 0$. From the continuity of $A_{\mathcal{I}}$ and $f_{\mathcal{I}}(\bar{z}) + A_{\mathcal{I}}(\bar{z})(y_\lambda - \bar{z}) < f_{\mathcal{I}}^+(\bar{z})$ and the facts that $z^k \xrightarrow{\mathcal{K}} \bar{z}$ and $w^k \xrightarrow{\mathcal{K}} y_\lambda$, for large $k \in \mathcal{K}$

$$f_{\mathcal{I}}(z^k) + A_{\mathcal{I}}(z^k)(w^k - z^k) < f_{\mathcal{I}}^+(z^k).$$

Thus, for large $k \in \mathcal{K}$, we have $w^k \in L(z^k)$ and $\|w^k - y_\lambda\| < \varepsilon/2$. For such w^k,

$$\|\bar{x} - w^k\| \leq \|\bar{x} - y_\lambda\| + \|y_\lambda - w^k\| < \varepsilon,$$

completing the proof. ∎

LEMMA A.2. *Let the point to set map $z \in \mathbb{R}^n \mapsto L(z) \in \mathcal{P}(\mathbb{R}^n)$ and the function $z \in \mathbb{R}^n \mapsto p(z) \in \mathbb{R}^n$ be continuous at $\bar{z} \in \mathbb{R}^n$. Then $z \in \mathbb{R}^n \mapsto P_{L(z)}\ (p(z))$ is continuous at $\bar{z}$.*

Proof. Consider a sequence $z^k \to \bar{z} \in \mathbb{R}^n$, $x^k = P_{L(z^k)}\ (p(z^k))$. We must prove that $x^k \to \bar{x} = P_{L(\bar{z})}\ (p(\bar{z}))$.

From the lower semi-continuity of $L(\cdot)$, there exists a sequence $y^k \in L(z^k)$ such that $y^k \to \bar{x}$. By definition of projection,

$$\|p(z^k) - x^k\| \leq \|p(z^k) - y^k\|. \tag{A.1}$$

Hence $(p(z^k) - x^k)$ is bounded, and consequently (x^k) is bounded. Consider $\tilde{x} \in \mathbb{R}^n$ and $\mathcal{K} \subset \mathbb{N}$ such that $(x^k) \xrightarrow{\mathcal{K}} \tilde{x}$. Using the upper semi-continuity of $L(\cdot)$, $\tilde{x} \in L(\bar{z})$ and hence by definition of projection,

$$\|\tilde{x} - p(\bar{z})\| \geq \|\bar{x} - p(\bar{z})\|.$$

Taking limits in $(A.1)$ for $k \in \mathcal{K}$, $k \to \infty$,

$$\|p(\bar{z}) - \tilde{x}\| \leq \|p(\bar{z}) - \bar{x}\|.$$

It follows that $\|p(\bar{z}) - \tilde{x}\| = \|p(\bar{z}) - \bar{x}\|$, and thus $\tilde{x} = \bar{x}$ by uniqueness of the projection onto a convex set. This proves that $\bar{x}$ is the unique accumulation point of (x^k), completing the proof. $\square$

Acknowledgment. We thank Andreas Wächter for keenly finding some mistakes in our proofs and giving suggestions which improved the paper. We also thank Jose Mario Martínez for many suggestions.

REFERENCES

[1] J. Abadie and J. Carpentier, *Generalization of the Wolfe reduced-gradient method to the case of nonlinear constraints*, in Optimization, R. Fletcher, ed., Academic Press, New York, 1968, pp. 37–47.

[2] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts, 1995.

[3] R. H. Byrd, *Robust trust region methods for constrained optimization.* Third SIAM Conference on Optimization, 1987.

[4] R. H. Byrd, J. C. Gilbert, and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming*, Mathematical Programming, 89 (2000), pp. 149–185.

[5] R. H. Byrd, M. E. Hribar, and J. Nocedal, *An interior point algorithm for large-scale nonlinear programming*, SIAM Journal on Optimization, 9 (1999), pp. 877–900.

[6] M. R. Celis, J. E. Dennis, and R. A. Tapia, *A trust region strategy for nonlinear equality constrained optimization*, in Numerical Optimization 1984, P. T. Boggs, R. H. Byrd, and R. B. Schnabel, eds., SIAM, Philadelphia, 1985, pp. 71 – 82.

[7] R. Fletcher, N. Gould, S. Leyffer, P. Toint, and A. Wächter, *Global convergence of trust-region and SQP-filter algorithms for general nonlinear programming*, SIAM Journal on Optimization, (2002), to appear.

[8] R. Fletcher and S. Leyffer, *Nonlinear programming without a penalty function*, Tech. Report NA/171, Dundee University, Dept. of Mathematics, 1997. Revised 2000.

[9] F. A. M. Gomes, M. C. Maciel, and J. M. Martínez, *Nonlinear programming algorithms using trust regions and augmented Lagrangians with nonmonotone penalty parameters*, Mathematical Programming, 84 (1999), pp. 161–200.

[10] J. M. Martínez, *Inexact-restoration method with Lagrangian tangent decrease and a new merit function for nonlinear programming*, Journal of Optimization Theory and Applications, (2001). To appear.

[11] J. M. Martínez and E. A. Pilotta, *Inexact restoration algorithms for constrained optimization*, Journal of Optimization Theory and Applications, 104 (2000), pp. 135–163.

[12] ———, *Inexact restoration methods for nonlinear programming: Advances and perspectives*, in Optimization and Control with Applications, Qi, Teo, and Yang, eds., Kluwer, 2001. To appear.

[13] J. M. Martínez and B. F. Svaiter, *A practical optimality condition without constraint qualifications for nonlinear programming*, tech. report, Institute of Mathematics, University of Campinas, Brazil, 2001.

[14] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Series in Operations Research, Springer-Verlag, 1999.

[15] E. Omojokun, *Trust Region Algorithms for Optimization with Nonlinear Equality and Inequality Constraints*, PhD thesis, Dept. of Computer Science, University of Colorado, 1991.

[16] J. B. Rosen, *The gradient projection method for nonlinear programming, part 1, linear constraints*, SIAM Journal on Applied Mathematics, 8 (1960), pp. 181–217.

[17] M. J. D. Powell, *Convergence properties of a class of minimization algorithms*, in Nonlinear Programming 2, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds., Academic Press, New York, 1975, pp. 1–27.

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

[19] M. Ulbrich, S. Ulbrich and L. N. Vicente, *A globally convergent primal-dual interior-*

point filter method for nonconvex nonlinear programming, Preprint 00-11 Department of Mathematics, University of Coimbra, Portugal 2000.