GLOBAL CONVERGENCE OF FILTER METHODS FOR NONLINEAR PROGRAMMING

ADEMIR A. RIBEIRO§, ELIZABETH W. KARAS§, AND CLÓVIS C. GONZAGA¶

Abstract. We present a general filter algorithm that allows a great deal of freedom in the step computation. Each iteration of the algorithm consists basically in computing a point which is not forbidden by the filter, from the current point. We prove its global convergence, assuming that the step must be efficient, in the sense that, near a feasible non-stationary point, the reduction of the objective function is “large”. We show that this condition is reasonable, by presenting two classical ways of performing the step which satisfy it. In the first one, the step is obtained by the inexact restoration method of Martínez and Pilotta. In the second, the step is computed by sequential quadratic programming.

Key words. Filter methods, nonlinear programming, global convergence

AMS subject classifications. 49M37, 65K05, 90C30

1. Introduction. We shall study the nonlinear programming problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_E(x) = 0 \\
& \quad f_I(x) \leq 0,
\end{align*}
\]

where the index sets \( E \) and \( I \) refer to the equality and inequality constraints respectively. Let the cardinality of \( E \cup I \) be \( m \), and assume that the functions \( f_i : \mathbb{R}^n \to \mathbb{R} \), \( i = 0, 1, \ldots, m \), are continuously differentiable.

A nonlinear programming algorithm must deal with two different (and possibly conflicting) criteria, related respectively to optimality and to feasibility. Optimality is measured by the objective function \( f_0 \); feasibility is typically measured by penalization of constraint violation, for instance, by the function \( h : \mathbb{R}^n \to \mathbb{R}_+ \), given by

\[
h(x) = \| f^+(x) \|,
\]

where \( \| \cdot \| \) is an arbitrary norm and \( f^+ : \mathbb{R}^n \to \mathbb{R}^m \) is defined by

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

Both criteria must be optimized and the algorithm should follow a certain balance between them at every step of the iterative process. Several algorithms for nonlinear programming have been designed in which a merit function is a tool to guarantee global convergence [3, 10, 11, 19, 25]. As an alternative to merit function, Fletcher and Leyffer [7] introduced the so-called filter to globalize sequential quadratic programming type methods. Filter methods are based on the concept of dominance, borrowed from multi-criteria optimization.

§Department of Mathematics, Federal University of Paraná, Cx. Postal 19081, 81531-980, Curitiba, PR, Brazil; e-mail: ademir@mat.ufpr.br, karas@mat.ufpr.br.

¶Department of Mathematics, Federal University of Santa Catarina, Cx. Postal 5210, 88040-970, Florianópolis, SC, Brazil; e-mail: clovis@mtm.ufsc.br. This author is supported by CNPq.
A filter algorithm defines a *forbidden region*, by memorizing pairs \((f_0(x^j), h(x^j))\), chosen conveniently from former iterations and then avoids points dominated by these by the Pareto domination rule:

\[
y \text{ is dominated by } x \text{ if, and only if, } f_0(y) \geq f_0(x) \text{ and } h(y) \geq h(x).
\]

Figure 1.1 shows a filter with four pairs, where we have simplified the notation by using \(x\) to represent the pair \((f_0(x), h(x))\). The point \(y\) is in the forbidden region and \(z\) is not.

![Filter Diagram](image)

**Fig. 1.1. A filter with four pairs.**

The filter methods were also applied to sequential linear programming (SLP). The works of Chin and Fletcher [5] and Fletcher, Leyffer and Toint [8] present global convergence proofs of the method.

For SQP-filter methods, global convergence has been proved by Fletcher, Leyffer and Toint [9], assuming that the quadratic subproblems are solved globally. Without this requirement, that is, allowing approximate solutions of the subproblems, Fletcher, Gould, Leyffer, Toint and Wächter [6] have also proved convergence to first-order critical points. Their approach uses a composite-step SQP method similar in spirit to the ones pioneered by Byrd [2] and Omojokun [21]. Another SQP-filter algorithm, using line search, was proposed by Wächter and Biegler [26], where global convergence was obtained.

In the context of interior points, Ulbrich, Ulbrich and Vicente [23] have proposed a globally convergent primal-dual interior-point filter method. However, the filter entries have components that take into account the centrality and complementarity measures arising from interior-point techniques.

The filter was also studied by Gonzaga, Karas and Vanti [12], in an algorithm that resembles the inexact restoration method of Martínez and Pilotta [18, 19]. By suitable rules for building the filter they prove stationarity of all qualified accumulation points.

The good performance of these methods [7, 15, 16] has motivated their use in other problems, like nonlinear systems of equations [13], unconstrained optimization [14] and nonsmooth convex constrained optimization [17]. This last work, by Karas, Ribeiro, Sagastizábal and Solodov, combines the ideas of the proximal bundle methods [22] with the filter strategy.

Although we know that filter methods may suffer from the Maratos effect, we shall not discuss local convergence issues in this work. Some strategies can be found in [4, 24, 27] to ensure fast rate of convergence.
In this paper, we propose a general filter algorithm that does not depend on the particular method used for the step computation. The only requirement is that the points generated must be acceptable for the filter and that near a feasible non-stationary point, the reduction of the objective function be large. This efficiency condition, stated ahead as Hypothesis H3, is the main tool of the global convergence analysis. It is a weaker version of the one introduced by Gonzaga, Karas and Vanti [12] in their inexact restoration filter method. Under this hypothesis, we prove that every sequence generated by the algorithm has at least one stationary accumulation point. Furthermore, we show how to compute the step in order to fulfill this hypothesis. One way to do this is by inexact restoration, for which H3 has proven in [12]. Another way for computing the step is by a sequential quadratic programming algorithm. We prove in this work that this approach also satisfies the efficiency condition H3.

The paper is organized as follows. Our general filter algorithm and its convergence analysis are described in Section 2. In Section 3 we present the SQP method for computing the step and prove that Hypothesis H3 is satisfied.

2. The algorithm. In this section we present a general filter algorithm that allows a great deal of freedom in the step computation. Afterwards we state an assumption on the performance of the step, and prove that any sequence generated by the algorithm has a stationary accumulation point. In the next section we show that this condition is reasonable, by presenting a classical way of performing the step, satisfying this condition.

The algorithm constructs a sequence of filter sets $F_0 \subset F_1 \subset \cdots \subset F_k$, composed of pairs $(\tilde{f}_0^j, \tilde{h}_j) \in \mathbb{R}^2$. We also mention in the algorithm the sets $F_k \subset \mathbb{R}^n$, which are formally defined in each step for clarity, but are never actually constructed.

**Algorithm 2.1. General filter algorithm model**

**Given:** $x^0 \in \mathbb{R}^n$, $F_0 = \emptyset$, $F_0 = \emptyset$, $\alpha \in (0, 1)$.

$k = 0$

REPEAT

$$(\tilde{f}_0, \tilde{h}) = (f_0(x^k) - \alpha h(x^k), (1 - \alpha) h(x^k)).$$

Set $\bar{F}_k = F_k \cup \{(\tilde{f}_0, \tilde{h})\}$ and define $\bar{F}_k = F_k \cup \{x \in \mathbb{R}^n | f_0(x) \geq \tilde{f}_0, h(x) \geq \tilde{h}\}$.

**Step:**

IF $x^k$ is stationary, stop with success
ELSE, compute $x^{k+1} \notin \bar{F}_k$.

**Filter update:**

IF $f_0(x^{k+1}) < f_0(x^k)$,

$$F_{k+1} = F_k, \quad \mathcal{F}_{k+1} = \mathcal{F}_k$$

(f0-iteration: the new entry is discarded)

ELSE,

$$F_{k+1} = \bar{F}_k, \quad \mathcal{F}_{k+1} = \bar{F}_k$$

(h-iteration: the new entry becomes permanent)

$k = k + 1$.

At the beginning of each iteration, the pair $(\tilde{f}_0, \tilde{h})$ 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 $f_0$.

Note that the forbidden region was slightly modified by subtracting the expression $\alpha h(x^k)$ from both filter pair components. This prevents the acceptance of trial pairs $(f_0, h)$ arbitrarily close to old iterates $(f_0(x^j), h(x^j))$. Figure 2.1 illustrates the effect of this modification which adds a small
margin around the border of region already defined by the Pareto domination rule. Now, all points above the dashed line are forbidden.

In order to obtain our global convergence result, we shall assume that the algorithm generates an infinite sequence \((x^k)_{k \in \mathbb{N}}\) in \(\mathbb{R}^n\) and the following hypotheses are satisfied.

**H1.** All the functions \(f_i(\cdot), i = 0, 1, \ldots, m,\) are twice continuously differentiable.

**H2.** The sequence \((x^k)_{k \in \mathbb{N}}\) remains in a convex compact domain \(X \subset \mathbb{R}^n\).

**H3.** Given a feasible non-stationary point \(\bar{x} \in X\), there exist \(M > 0\) and a neighborhood \(V\) of \(\bar{x}\) such that for any iterate \(x^k \in V\),

\[
f_0(x^k) - f_0(x^{k+1}) \geq Mv_k,
\]

where \(v_k = \min \left\{ 1, \min \left\{ \tilde{h}^j \mid (\tilde{f}_0^j, \tilde{h}^j) \in F_k \right\} \right\} \).

The first ones are standard hypotheses, while the hypothesis H3 requires some discussion. With this hypothesis we are assuming that the step must be efficient, in the sense that, near a feasible non-stationary point, the reduction of the objective function is “large”. This assumption is a weaker version of the decrease condition introduced by Gonzaga, Karas and Vanti [12] who used

\[
H_k = \min \left\{ 1, \min \left\{ \tilde{h}^j \mid (\tilde{f}_0^j, \tilde{h}^j) \in F_k, \tilde{f}_0^j \leq f_0(x^k) \right\} \right\}
\]

instead of \(v_k\). Figure 2.2 shows the variables \(v_k\) and \(H_k\).

We start with some relations which follow directly from the hypotheses and the construction of the algorithm. In the next section we shall state methods which satisfy Hypothesis H3.

**Lemma 2.2.** Given \(k \in \mathbb{N}\), the following statements hold:

(i) \(F_k \subset F_{k+1}\) (Inclusion property).

(ii) \(x^{k+p} \notin F_{k+1}\) for all \(p \geq 1\).

(iii) At least one of the two conditions below occurs:

1. \(f_0(x^{k+1}) < f_0(x^k) - \alpha h(x^k)\),
2. \(h(x^{k+1}) < (1 - \alpha)h(x^k)\).

(iv) \(\tilde{h} > 0\) for all \((\tilde{f}_0, \tilde{h}) \in F_k\).
Proof. The first statement follows directly from the filter update criterion. By (i) and the
definition of $\mathcal{F}$, we have $\mathcal{F}_{k+1} \subset \mathcal{F}_{k+p} \subset \mathcal{F}_{k+p-1}$. Then the second statement follows from $x^{k+p} \notin \mathcal{F}_{k+p-1}$. The third one is clear, since $x^{k+1} \notin \mathcal{F}_k$. Finally, note that if $h(x^k) = 0$, using (iii), we obtain
\[
f_0(x^{k+1}) < f_0(x^k) - \alpha h(x^k) = f_0(x^k),
\]
that is, the iteration $k$ is an $f_0$-iteration. Thus the pair $(\tilde{f}_0, \tilde{h})$ can be added to the filter only if $h(x^k) > 0$, or equivalently if $\tilde{h} > 0$. This completes the proof. \(\square\)

For the purpose of our analysis, we shall consider
\begin{equation}
K_a = \{ k \in \mathbb{N} \mid (f_0(x^k) - \alpha h(x^k), (1 - \alpha)h(x^k)) \text{ is added to the filter} \};
\end{equation}
the set of indices of $h$-iterations. First, we analyze what happens when this set is infinite.

**Lemma 2.3.** If the set $K_a$ is infinite, then
\[
h(x^k) \xrightarrow{K_a} 0.
\]

Proof. Given $k$, we denote $(f_0(x^k), h(x^k))$ by $(f_0^k, h^k)$. Assume by contradiction that, for some $\delta > 0$, the set
\[
K = \{ k \in K_a \mid h(x^k) \geq \delta \}
\]
is infinite. The continuity of $(f_0, h)$, implied by H1, and the compactness assumption H2 ensure that there exists a convergent subsequence $(f_0^k, h^k)_{k \in K_1}, K_1 \subset K$. Therefore, since $\alpha \in (0, 1)$, we can take indices $j, k \in K_1, j < k$ such that
\[
\| (f_0^k, h^k) - (f_0^j, h^j) \| < \alpha \delta \leq \alpha h(x^j).
\]
This means that $x^k \in \mathcal{F}_j = \mathcal{F}_{j+1}$, contradicting Lemma 2.2(ii) and completing the proof. \(\square\)
We now prove that the objective function decreases along the iterations, whenever the iterates stay near a non-stationary point.

Lemma 2.4. Let \( \bar{x} \in X \) be a non-stationary point. Then there exist \( \tilde{k} \in \mathbb{N} \) and a neighborhood \( V \) of \( \bar{x} \) such that whenever \( k > \tilde{k} \) and \( x^k \in V \), the iteration \( k \) is an \( f_0 \)-iteration, that is, \( k \notin K_a \).

Proof. If \( \bar{x} \) is a feasible point, then by Hypothesis H3 there exist \( M > 0 \) and a neighborhood \( V \) of \( \bar{x} \) such that for all \( x^k \in V \),

\[
f_0(x^k) - f_0(x^{k+1}) \geq Mv_k.
\]

Using Lemma 2.2(iv), we conclude that \( v_k > 0 \), consequently \( f_0(x^{k+1}) < f_0(x^k) \) and \( k \) is an \( f_0 \)-iteration.

Now, assume that \( \bar{x} \) is infeasible and suppose by contradiction that there exists an infinite set \( K \subset K_a \) such that \( x^k \xrightarrow{K} \bar{x} \). Since \( h \) is continuous, we have \( h(x^k) \xrightarrow{K} h(\bar{x}) \). On the other hand, Lemma 2.3 ensures that \( h(x^k) \xrightarrow{K} 0 \). Thus \( h(\bar{x}) = 0 \), contradicting that \( \bar{x} \) is infeasible and completing the proof. \( \square \)

Our global convergence result is presented in the next theorem.

Theorem 2.5. The sequence \( (x^k)_{k \in \mathbb{N}} \) has a stationary accumulation point.

Proof. Let \( K_a \) be the set defined in (2.2). If \( K_a \) is infinite, then by H2 there exist \( K_1 \subset K_a \) and \( \bar{x} \in X \) such that \( x^k \xrightarrow{K_1} \bar{x} \). From Lemma 2.4, \( \bar{x} \) must be stationary.

On the other hand, if \( K_a \) is finite, there exists \( k_0 \in \mathbb{N} \) such that every iteration \( k \geq k_0 \) is an \( f_0 \)-iteration. Thus \( (f_0(x^k))_{k \geq k_0} \) is decreasing and by H1 and H2,

\[
f_0(x^k) - f_0(x^{k+1}) \to 0.
\]

Moreover, by construction, \( F_k = F_{k_0} \) for all \( k \geq k_0 \). Therefore, the sequence \( (v_k)_{k \in \mathbb{N}} \), defined in Hypothesis H3, satisfies

\[
v_k = v_{k_0} > 0
\]

for all \( k \geq k_0 \). If the set

\[
K_2 = \{ k \in \mathbb{N} \mid \alpha h(x^k) < f_0(x^k) - f_0(x^{k+1}) \}
\]

is infinite, using (2.3), we conclude that \( h(x^k) \xrightarrow{K_2} 0 \). Otherwise, Lemma 2.2(iii) ensures that there exists \( k_1 \in \mathbb{N} \) such that \( h(x^{k+1}) < (1 - \alpha)h(x^k) \) for all \( k \geq k_1 \), which in turn implies that \( h(x^k) \to 0 \). Anyway, \( (x^k)_{k \in \mathbb{N}} \) has a feasible accumulation point \( \bar{x} \). Now we prove that this point is stationary. Let \( K \) be a set of indices such that \( x^k \xrightarrow{K} \bar{x} \) and assume by contradiction that \( \bar{x} \) is non-stationary. By Hypothesis H3, there exist \( k_2 \in \mathbb{N} \) and \( M > 0 \) such that

\[
f_0(x^k) - f_0(x^{k+1}) \geq Mv_k
\]

for all \( k \in K, k \geq k_2 \). This together with (2.4) contradicts (2.3), completing the proof. \( \square \)

As we have seen above, the hypothesis H3 is crucial for the convergence analysis. It is a very strong assumption and we must show that there exist methods satisfying this condition. One of them is the inexact restoration method of Martínez and Pilotta [19]. Gonzaga, Karas and Vanti [12] have proved in their inexact restoration filter method a condition that implies our hypothesis.

We now discuss another way of performing the step, satisfying H3. It uses sequential quadratic programming and decomposes the step into its normal and tangential components.
3. Sequential quadratic programming. In this section we present an SQP method based on that proposed by Fletcher, Gould, Leyffer, Toint and Wächter [6], which computes the overall step in two phases. First, a feasibility phase aims at reducing the infeasibility measure $h$, satisfying a linear approximation of the constraints. Then an optimality phase computes a trial point reducing a quadratic model of the objective function in the linearization of the feasible set. We prove that this approach satisfies Hypothesis H3.

The step computation. Given the current iterate $x^k$ and a trust-region radius $\Delta > 0$, we compute the step by solving the quadratic subproblem

$$\begin{align*}
(QP_k) \quad \text{minimize} & \quad m_k(x^k + d) \\
\text{subject to} & \quad x^k + d \in L(x^k) \\
& \quad \|d\| \leq \Delta,
\end{align*}$$

where

$$m_k(x^k + d) = f_0(x^k) + \nabla f_0(x^k)^T d + \frac{1}{2} d^T B_k d,$$

with $B_k$ symmetric, and

$$L(x^k) = \{ x^k + d \in \mathbb{R}^n \mid f_E(x^k) + A_E(x^k)d = 0, \ f_I(x^k) + A_I(x^k)d \leq 0 \}.$$  

The matrix $B_k$ may be chosen as an approximation of the Hessian of some Lagrangian function or any other symmetric matrix, provided that it remains uniformly bounded. See the hypothesis H6 below.

The solution of $(QP_k)$ yields a trial point $x^k + d_\Delta$, that will be evaluated by the filter. To be accepted as the new iterate, this point must not be forbidden.

In fact, we will see the step $d_\Delta$ as the sum of two components, a feasibility step $n_k$ and a tangential (optimality) step $t_\Delta$. We now discuss each one of these steps.

Feasibility step and compatibility of $(QP_k)$. The feasibility step $n_k$ must satisfy the constraints of $(QP_k)$ and has the purpose of reducing the infeasibility measure $h$. This can be done, for example, by

$$n_k = P_{L(x^k)}(x^k) - x^k,$$

where $P_{L(x^k)}(\cdot)$ is the projection onto the set $L(x)$. However, we do not use this particular choice, but we shall assume a certain efficiency in this phase, given by the following hypothesis.

**H4.** There exist constants $\delta_h > 0$ and $c_n > 0$ such that for all $k \geq 0$ with $h(x^k) \leq \delta_h$, a step $n_k$ can be computed, satisfying

$$\|n_k\| \leq c_n h(x^k).$$

This assumption means that the feasibility step must be reasonably scaled with respect to the constraints. In particular, $n_k = 0$ whenever $x^k$ is feasible. This hypothesis is discussed by Martínez [18], who presents a feasibility algorithm which satisfies it.
A. A. RIBEIRO, E. W. KARAS AND C. C. GONZAGA

The step \( n^k \) is only useful if it is not too close to the trust-region boundary because, otherwise, the tangential step is unlikely to produce a sufficient decrease in the model \( m_k \). We say that the subproblem \((QP_k)\) is compatible when \( \mathcal{L}(x^k) \neq \emptyset \) and

\[
\|n^k\| \leq \xi \Delta, 
\]

where \( \xi \in (0, 1) \) is a constant.

In our analysis, we shall consider

\[
z^k = x^k + n^k
\]

the point obtained in the feasibility phase. Note that, from (3.1) and (3.4), we have

\[
m_k(z^k) = m_k(x^k + n^k) = f_0(x^k) + \nabla f_0(x^k)^T n^k + \frac{1}{2} n^k^T B_k n^k.
\]

Tangential step. If the subproblem \((QP_k)\) is compatible, we anticipate a satisfactory decrease in the model when performing a tangential step \( t \Delta \), approximate solution of the quadratic problem

\[
\begin{align*}
\text{minimize} & \quad (\nabla f_0(x^k) + B_k n^k)^T t + \frac{1}{2} t^T B_k t \\
\text{subject to} & \quad A_E(x^k) t = 0 \\
& \quad f_I(x^k) + A_I(x^k)(n^k + t) \leq 0 \\
& \quad \|n^k + t\| \leq \Delta.
\end{align*}
\]

This problem is equivalent to \((QP_k)\) with \( d = n^k + t \).

Given the current iterate \( x^k \) and a trust-region radius \( \Delta > 0 \), if \((QP_k)\) is compatible, the trial point is

\[
x^k + d_\Delta = z^k + t_\Delta,
\]

where \( z^k = x^k + n^k \) is the point which comes from the feasibility phase and \( t_\Delta \) is the tangential step.

Restoration procedure. If the subproblem \((QP_k)\) is not compatible, the algorithm calls a restoration procedure, whose aim is to obtain a point \( x^{k+1} \notin \mathcal{F}_k \) with \( h(x^{k+1}) < h(x^k) \), where the function \( h \) is the infeasibility measure defined by (1.1). This can be done by taking steps of some algorithm for solving the problem

\[
\begin{align*}
\text{minimize} & \quad h(x) \\
\text{subject to} & \quad x \in \mathbb{R}^n.
\end{align*}
\]

We can now summarize the above discussion in the following algorithm for the step computation. After stating the algorithm we shall make some comments about its features.
Algorithm 3.1. Computation of $x^{k+1} \notin \bar{\mathcal{F}}_k$

Data: $x^k \in \mathbb{R}^n$, the current filter $\bar{\mathcal{F}}_k$, $0 < \Delta_{\min} < \Delta_{\max}$, $\Delta \in [\Delta_{\min}, \Delta_{\max}]$ and $c_p, \xi, \eta, \gamma \in (0, 1)$.

if $\mathcal{L}(x^k) = \emptyset$, use the restoration procedure to obtain $x^{k+1} \notin \bar{\mathcal{F}}_k$.

else

compute a feasibility step $n^k$ such that $x^k + n^k \in \mathcal{L}(x^k)$

repeat (while the point $x^{k+1}$ is not obtained)

if $\|n^k\| > \xi \Delta$

use the restoration procedure to obtain $x^{k+1} \notin \bar{\mathcal{F}}_k$.

else,

compute the tangential step $t_\Delta$ as above and define $d_\Delta = n^k + t_\Delta$

set $ared = f_0(x^k) - f_0(x^k + d_\Delta)$ and $pred = m_k(x^k) - m_k(x^k + d_\Delta)$

IF $\{x^k + d_\Delta \in \bar{\mathcal{F}}_k\}$ OR $\{pred \geq c_p(h(x^k))^2$ AND $ared < \eta pred\}$

$\Delta = \gamma \Delta$

ELSE

$x^{k+1} = x^k + d_\Delta$

determine $B_{k+1}$ symmetric

$\Delta_k = \Delta$


Algorithm 3.1 was inspired in the SQP-filter algorithm proposed by Fletcher, Gould, Leyffer, Toint and Wächter [6]. However, there exist some differences between them, which we now point out. The first one is that here the step computation is made separately from the main filter algorithm, presented in Section 2. This simplifies the study of the step properties and leaves the convergence analysis of the main algorithm in a clean framework. Another difference is in the trust-region radius. Algorithm 3.1 starts with a radius $\Delta \in [\Delta_{\min}, \Delta_{\max}]$, where $\Delta_{\min}, \Delta_{\max} > 0$ are constants. This procedure is not used in [6], making the convergence proofs involved. To overcome some difficulties they impose a condition like

$$\|n^k\| \leq c \Delta^{1+\mu},$$

where $c > 0$ and $\mu \in (0, 1)$, to accept the normal step and to proceed with the tangential step. In our algorithm, this condition is replaced by (3.3), that is,

$$\|n^k\| \leq \xi \Delta,$$

where $\xi \in (0, 1)$ is a constant. This requirement is usual in the composite-step approaches that we are considering.

We mention that the choice of a minimum radius $\Delta_{\min}$ may cause practical disadvantages, like the rejection of many trial points before the progress of the algorithm. On the other hand, it simplifies the analysis and enhances the chance of taking a pure Newton step.

Remarks. At iteration $k$, we denote by $d_\Delta$ the trial step obtained with the trust-region radius $\Delta \geq \Delta_k$. The point $x^{k+1}$ can be computed in two different ways: by means of a restoration procedure or by $x^{k+1} = x^k + d_\Delta$. We also have two possibilities for rejecting the trial step $d_\Delta$:

$$x^k + d_\Delta \in \bar{\mathcal{F}}_k$$
or
\begin{equation}
\text{pred} \geq c_p \left( h(x^k) \right)^2 \quad \text{and} \quad \text{ared} < \eta \text{pred}.
\end{equation}

In both cases the trust-region radius is reduced and a new step is computed. Thus, in order to accept the step $d_\Delta$, it is not enough to pass the filter criterion. It also must ensure a sufficient decrease in the objective function whenever the predicted reduction is more significant than the constraint violation. In particular, if all iterates are feasible, the first inequality in (3.7) will be always true, because $n^k = 0$ in this case. Furthermore, if $\text{ared} \geq \eta \text{pred}$, then $x^k + d_\Delta \notin \mathcal{F}_k$. So, the step acceptance criterion reduces to $\text{ared} \geq \eta \text{pred}$ and the algorithm may be viewed as a classical unconstrained trust-region method.

We now prove that Hypothesis H3 is satisfied if Algorithm 2.1 is applied to problem $(P)$ and the step is obtained by Algorithm 3.1. For that, we shall introduce a function used as a stationarity measure. Given $x, z \in X$ and the set $\mathcal{L}(x)$ defined in (3.2), we denote
\begin{equation}
\varphi(x, z) = \begin{cases} 
- \nabla f_0(x)^T \frac{d^c(x, z)}{\|d^c(x, z)\|} & \text{if } d^c(x, z) \neq 0, \\
0 & \text{otherwise},
\end{cases}
\end{equation}
the stationarity measure. According to [12] we have, at a feasible point $\bar{x}$, that the KKT conditions are equivalent to $d^c(\bar{x}, \bar{x}) = 0$. Furthermore, if $\bar{x}$ is non-stationary, then $\varphi(\bar{x}, \bar{x}) > 0$.

The projected gradient direction given above is based on a direction introduced by Martínez and Svaiter [20] to define a new optimality condition, called $\text{AGP property}$ (Approximate Gradient Projection), that implies, and is strictly stronger than, Fritz-John optimality conditions. Unlike the KKT conditions, it is satisfied by local minimizers of nonlinear programming problems, independently of constraint qualifications.

Note. Let us give an interpretation for the direction $d^c(x, z)$ when $z \in \mathcal{L}(x)$ (which is the case in the algorithm). It is an approximation to
\begin{equation}
d_B(z) = P_{\mathcal{L}(x)} (z - \nabla f_0(x)) - z.
\end{equation}
This is the projected Cauchy direction defined by Bertsekas [1] for the minimization of $f_0(\cdot)$ in $\mathcal{L}(x)$ and $d_B(z) = 0$ implies that $z$ is stationary for this problem. If, in addition, $z$ is feasible for $(P)$ it is also stationary for $(P)$. The direction $d^c(x, z)$ may be a good descent direction for $(P)$ if
\begin{equation}
\frac{\nabla f_0(x)}{\|\nabla f_0(x)\|} \approx \frac{\nabla f_0(z)}{\|\nabla f_0(z)\|},
\end{equation}
but otherwise it may be meaningless (possibly null). If $d^c(x, z) \neq 0$, we consider $d^c_i = \frac{d^c(x, z)}{\|d^c(x, z)\|}$.

To continue our analysis we define the $\text{generalized Cauchy step}$ given by
\begin{equation}
t^c = \begin{cases} 
\arg\min_{\lambda \geq 0} \left\{ m_k(z^k + \lambda d^c_i) \mid \|z^k + \lambda d^c_i - x^k\| \leq \Delta \right\} & \text{if } d^c(x^k, z^k) \neq 0, \\
0 & \text{otherwise},
\end{cases}
\end{equation}

\end{document}
and we assume the following hypotheses related to Algorithm 3.1.

**H5.** If the subproblem \((QP_k)\) is compatible, then the model decrease at the tangential step \(t_\Delta\) satisfies

\[
m_k(z^k) - m_k(z^k + t_\Delta) \geq m_k(z^k) - m_k(z^k + t^c) .
\]

**H6.** The matrices \(B_k\) are uniformly bounded, that is, there exists a constant \(\beta > 0\) such that

\[
\|B_k\| \leq \beta \quad \text{for all} \quad k \geq 0.
\]

The assumption H5 says that the tangential step must be at least as good as the generalized Cauchy step \(t^c\). We also consider a very standard condition on the Hessians \(B_k\), stated in Hypothesis H6.

We start our task by evaluating the infeasibility measure before and after the trial step.

**Lemma 3.2.** Suppose that Hypotheses H1 and H2 hold. There exists a constant \(c_h > 0\) such that for any \(x^k \in X\) and \(\Delta > 0\) so that the problem \((QP_k)\) is compatible,

\[
h(x^k) \leq c_h \Delta \quad \text{and} \quad h(x^k + d_\Delta) \leq c_h \Delta^2,
\]

where \(d_\Delta\) is the trial step obtained by Algorithm 3.1.

**Proof.** It follows from Hypotheses H1 and H2 that there exists a constant \(c_h > 0\) such that

\[
(3.10) \quad \|\nabla f_i(x)\| \leq c_h \quad \text{and} \quad \|\nabla^2 f_i(x)\| \leq c_h
\]

for all \(x \in X\) and \(i = 1, \ldots, m\). Consider \(x^k \in X\) and \(\Delta > 0\) so that the problem \((QP_k)\) is compatible. Thus the feasibility step \(n^k\) and the trial step \(d_\Delta\) were computed by Algorithm 3.1. Taking, without loss of generality, the norm \(\|\cdot\|_\infty\) in (1.1) and using the fact that \(x^k + n^k \in L(x^k)\), we conclude that

\[
h(x^k) = |f_i(x^k)| = | - \nabla f_i(x^k)^T n^k|
\]

for some \(i \in E\), or

\[
h(x^k) = f_i(x^k) \leq -\nabla f_i(x^k)^T n^k
\]

for some \(i \in I\). Hence, from the Cauchy-Schwarz inequality, (3.10) and the trust-region boundedness of \(n^k\), we obtain

\[
h(x^k) \leq \|\nabla f_i(x^k)\| \|n^k\| \leq c_h \Delta,
\]

proving the first claim in the lemma.

To prove the other inequality, note that by Taylor’s theorem and the fact that \(x^k + d_\Delta \in L(x^k)\),

\[
f_i(x^k + d_\Delta) = \frac{1}{2} d_\Delta^T \nabla^2 f_i(x^k + \theta_i d_\Delta) d_\Delta,
\]

for \(i \in E\), and

\[
f_i(x^k + d_\Delta) \leq \frac{1}{2} d_\Delta^T \nabla^2 f_i(x^k + \theta_i d_\Delta) d_\Delta,
\]
for $i \in I$, where $\theta_i \in (0, 1)$. Because the trust-region radius is bounded, we may assume without loss of generality that the trial points also remain in the compact set $X$. Thus, from (3.10), the Cauchy-Schwarz inequality and since $\|d_\Delta\| \leq \Delta$,
\[
h(x^k + d_\Delta) \leq c_h \Delta^2,
\]
completing the proof. □

We next assess the model and the objective function growth in the feasibility step computed by Algorithm 3.1.

**Lemma 3.3.** Suppose that Hypotheses H1, H2, H4 and H6 hold. Given a feasible point $\bar{x} \in X$, there exist $N > 0$ and a neighborhood $V_1$ of $\bar{x}$ such that if $x^k \in V_1$ and $z^k = x^k + n^k$, then
\[
(i) \quad |m_k(x^k) - m_k(z^k)| \leq N h(x^k). \\
(ii) \quad |f_0(x^k) - f_0(z^k)| \leq N h(x^k).
\]

**Proof.** Let $\delta_h$ and $c_n$ be the constants given by H4, and $\beta$ given by H6. Consider the constant $c_g = \max \{\|\nabla f_0(x)\| \mid x \in X\}$, whose existence is ensured by H1 and H2. Since $h(\bar{x}) = 0$ and $h$ is continuous, there exists a neighborhood $V_1$ of $\bar{x}$ such that if $x^k \in V_1$, then
\[
(3.11) \quad h(x^k) \leq \delta_h \quad \text{and} \quad \frac{1}{2} \beta c_n^2 (h(x^k))^2 \leq c_g c_n h(x^k).
\]
By (3.5), we have
\[
m_k(z^k) = m_k(x^k) + \nabla f_0(x^k)^T n^k + \frac{1}{2} n^T B_k n^k.
\]
Using the Cauchy-Schwarz inequality, H4 and H6, we obtain
\[
|m_k(x^k) - m_k(z^k)| \leq c_g \|n^k\| + \frac{1}{2} \beta \|n^k\|^2 \\
\phantom{|m_k(x^k) - m_k(z^k)|} \leq c_g c_n h(x^k) + \frac{1}{2} \beta c_n^2 (h(x^k))^2.
\]
From the second inequality in (3.11), it follows that
\[
|m_k(x^k) - m_k(z^k)| \leq 2 c_g c_n h(x^k).
\]
On the other hand, by Hypotheses H1 and H2, there exists a constant $L > 0$ such that
\[
|f_0(x^k) - f_0(z^k)| \leq L \|x^k - z^k\|.
\]
This together with H4 yields
\[
|f_0(x^k) - f_0(z^k)| \leq L c_n h(x^k).
\]
Taking $N = \max \{2 c_g c_n, L c_n\}$, we complete the proof. □
Lemma 3.4. Suppose that Hypotheses H1, H2, H4-H6 hold. Let \( \bar{x} \in X \) be a feasible non-stationary point and \( \bar{\eta} \in (0, 1) \). Consider the neighborhood \( V_1 \) and the constant \( \Delta_{\text{min}} \) given by Lemma 3.3 and Algorithm 3.1, respectively. Then there exist constants \( \Delta, \bar{\Delta} \in (0, \Delta_{\text{min}}] \), \( \bar{\epsilon} > 0 \) and a neighborhood \( V_2 \subset V_1 \) of \( \bar{x} \) such that whenever \( x^k \in V_2 \), \( z^k = x^k + n^k \) and a tangential step \( t_\Delta \) is obtained by the algorithm, we have:

(i) \( m_k(z^k) - m_k(z^k + t_\Delta) \geq \bar{\epsilon} \Delta' \) for all \( \Delta, \Delta' \) such that \( \Delta' \leq \min \{ \Delta, \Delta_{\text{min}} \} \).

(ii) \( f_0(z^k) - f_0(z^k + t_\Delta) \geq \bar{\eta} \left( m_k(z^k) - m_k(z^k + t_\Delta) \right) \) for all \( \Delta \in (0, \Delta_{\text{min}}) \).

Proof. Let \( \Delta > 0 \) and \( \lambda_{\Delta'} = (1 - \xi) \Delta' \), where \( \xi \) is given by (3.3) and \( \Delta' \leq \Delta \). First, note that the vector \( \bar{d}_1 \), defined in H5, satisfies \( \| \bar{d}_1 \| = 1 \). Consequently,

\[
\| z^k + \lambda_{\Delta'} \bar{d}_1 - x^k \| = \| n^k + \lambda_{\Delta'} \bar{d}_1 \| \leq \| n^k \| + \lambda_{\Delta'} \leq \xi \Delta + (1 - \xi) \Delta' \leq \Delta.
\]

Using the assumption on the Cauchy point H5, we obtain

\[
m_k(z^k) - m_k(z^k + t_\Delta) \geq m_k(z^k) - m_k(z^k + t') \geq m_k(z^k) - m_k(z^k + \lambda_{\Delta'} \bar{d}_1).
\]

Developing the quadratic model (3.1) in the right hand side, we conclude that

\[
m_k(z^k) - m_k(z^k + t_\Delta) \geq \lambda_{\Delta'} \left( -\nabla f_0(x^k)^T \bar{d}_1 - n^k T B_k \bar{d}_1 - \frac{1}{2} \lambda_{\Delta'} \bar{d}_1^T B_k \bar{d}_1 \right).
\]

By (3.9), \( \varphi(x^k, z^k) = -\nabla f_0(x^k)^T \bar{d}_1 \) and by H6, \( \| B_k \| \leq \beta \). Hence

\[
(3.12) \quad m_k(z^k) - m_k(z^k + t_\Delta) \geq \lambda_{\Delta'} \left( \varphi(x^k, z^k) - \| n^k \| \beta - \frac{1}{2} \lambda_{\Delta'} \beta \right).
\]

Since \( \bar{x} \) is feasible non-stationary, the continuous function \( \varphi \) satisfies \( \varphi(\bar{x}, \bar{x}) > 0 \). Using the fact that \( \| n^k \| \leq c_n h(x^k) \) by H4, we conclude that there exist a neighborhood \( V_2 \) of \( \bar{x} \) and \( \Delta_0 \in (0, \Delta_{\text{min}}] \) such that for any \( x^k \in V_2 \) and \( \Delta' \in (0, \Delta_0] \),

\[
\varphi(x^k, z^k) \geq \frac{1}{2} \varphi(\bar{x}, \bar{x}) \quad \text{and} \quad \| n^k \| \beta + \frac{1}{2} \lambda_{\Delta_{\text{min}}} \beta \leq \frac{1}{4} \varphi(\bar{x}, \bar{x}).
\]

Thus, by (3.12), we obtain for \( \Delta' \leq \min \{ \Delta, \Delta_0 \} \),

\[
m_k(z^k) - m_k(z^k + t_\Delta) \geq \frac{1}{4} \lambda_{\Delta_{\text{min}}} \varphi(\bar{x}, \bar{x}) = \frac{1}{4} (1 - \xi) \varphi(\bar{x}, \bar{x}) \Delta'.
\]

This proves (i) for any \( \Delta_{\text{min}} \leq \Delta_0 \) and \( \bar{c} = \frac{1}{4} (1 - \xi) \varphi(\bar{x}, \bar{x}) \).

To prove (ii), note that by the mean value theorem,

\[
ared_{z^k} \overset{\text{def}}{=} f_0(z^k) - f_0(z^k + t_\Delta) = -\nabla f_0(z^k + \theta t_\Delta)^T t_\Delta,
\]
for some $\theta \in (0, 1)$. On the other hand,

$$\text{pred}_{z^k} \overset{\text{def}}{=} m_k(z^k) - m_k(z^k + t_\Delta) = -\nabla f_0(x^k)^T t_\Delta - t_\Delta^T B_k n^k - \frac{1}{2} t_\Delta^T B_k t_\Delta.$$ 

By H1 and H2, we can apply the mean value inequality to $\nabla f_0$ to conclude that there exists a constant $L > 0$ such that

$$\|\nabla f_0(z^k) - \nabla f_0(z^k + \theta t_\Delta)\| \leq L \|z^k - x^k + \theta t_\Delta\|,$$

so, using the facts that $L > 0$, we obtain

$$|\text{pred}_{z^k} - \text{pred}_{z^k}| \leq L \|z^k - x^k + \theta t_\Delta\| \|t_\Delta\| + \beta \|n^k\| \|t_\Delta\| + \frac{1}{2} \beta \|t_\Delta\|^2$$

We can restrict the neighborhood $V_2$, if necessary, and take $\Delta_p \leq \Delta_0$ such that for any $x^k \in V_2$ and $\Delta \in (0, \Delta_p]$,

$$\frac{(L + \beta) \|n^k\|}{\bar{c}} \leq \frac{1 - \bar{\eta}}{2} \quad \text{and} \quad \frac{(L + \frac{1}{2} \beta) \Delta}{\bar{c}} \leq \frac{1 - \bar{\eta}}{2}.$$

Consequently, using $(i)$ with $\Delta' = \Delta$,

$$\frac{|\text{pred}_{z^k} - \text{pred}_{z^k}|}{\text{pred}_{z^k}} \leq \frac{(L + \beta) \|n^k\| \Delta + (L + \frac{1}{2} \beta) \Delta^2}{\bar{c} \Delta} \leq 1 - \bar{\eta},$$

completing the proof. \qed

In the next lemma we extend for the whole step the properties of the tangential step near a feasible non-stationary point.

**Lemma 3.5.** Suppose that Hypotheses H1, H2, H4-H6 hold. Let $\bar{x} \in X$ be a feasible non-stationary point and $0 < \eta < 1$. Consider the constant $\gamma$ given in Algorithm 3.1, the neighborhood $V_2$ and the constant $\Delta_p$ given in Lemma 3.4. Then there exists a neighborhood $V_3 \subset V_2$ of $\bar{x}$ such that whenever $x^k \in V_3$, $z^k = x^k + n^k$ and a tangential trial step $t_\Delta$ is obtained by the algorithm, we have for all $\Delta \in [\gamma^2 \Delta_p, \Delta_p]$,

1. $m_k(x^k) - m_k(z^k + t_\Delta) \geq \frac{1}{2} \bar{c} \Delta$,
2. $f_0(x^k) - f_0(z^k + t_\Delta) \geq \eta (m_k(x^k) - m_k(z^k + t_\Delta))$.

Proof. Let $\bar{\eta} \in (\eta, 1)$ and $\tau = \frac{\bar{\eta} - \eta}{\bar{\eta} + \eta}$. Consider the constants $N$ and $\bar{c}$ given by Lemmas 3.3 and 3.4, respectively, and $V_3 \subset V_2$ a neighborhood of $\bar{x}$ such that for all $x \in V_3$,

$$N h(x) \leq \min \left\{ \frac{1}{2} \bar{c} \gamma^2 \Delta_p, \tau \bar{\eta} \bar{c} \gamma^2 \Delta_p \right\}.$$

Hence, if $x^k \in V_3$ and $\Delta \in [\gamma^2 \Delta_p, \Delta_p]$, we can apply Lemma 3.3 to conclude that

$$|m_k(x^k) - m_k(z^k)| \leq N h(x^k) \leq \frac{1}{2} \bar{c} \gamma^2 \Delta_p \leq \frac{1}{2} \bar{c} \Delta.$$
It follows from this and Lemma 3.4(i) with $\Delta' = \Delta$ that
\[ m_k(x^k) - m_k(z^k + t\Delta) = m_k(x^k) - m_k(z^k) + m_k(z^k) - m_k(z^k + t\Delta) \geq \frac{1}{2}c\Delta. \]
proving (i).

(ii) Applying again Lemmas 3.3 and 3.4 together with (3.13), we obtain
\[ |f_0(x^k) - \tau h(z^k)| \leq Nh(x^k) \leq \tau \tilde{\eta}c\gamma^2 \Delta_p \leq \tau \tilde{\eta}c\Delta \leq \tau (f_0(z^k) - f_0(z^k + t\Delta)) \]
and
\[ m_k(x^k) - m_k(z^k) \leq Nh(x^k) \leq \tau \tilde{\gamma}^2 \Delta_p \leq \tau \Delta \leq \tau (m_k(z^k) - m_k(z^k + t\Delta)). \]
Consequently
\begin{equation}
(3.14) \quad f_0(x^k) - f_0(z^k + t\Delta) = f_0(x^k) - f_0(z^k) + f_0(z^k) - f_0(z^k + t\Delta) \geq (1 - \tau)(f_0(z^k) - f_0(z^k + t\Delta))
\end{equation}
and
\begin{equation}
(3.15) \quad m_k(x^k) - m_k(z^k + t\Delta) = m_k(x^k) - m_k(z^k) + m_k(z^k) - m_k(z^k + t\Delta) \leq (1 + \tau)(m_k(z^k) - m_k(z^k + t\Delta)).
\end{equation}
Therefore, if $x^k \in V_3$ and $\Delta \in [\gamma^2 \Delta_p, \Delta_p]$, using (3.14), (3.15) and Lemma 3.4(ii), we obtain
\[ f_0(x^k) - f_0(z^k + t\Delta) \geq (1 - \tau)\tilde{\eta}(m_k(z^k) - m_k(z^k + t\Delta)) \geq \frac{(1 - \tau)\tilde{\eta}}{(1 + \tau)}(m_k(z^k) - m_k(z^k + t\Delta)) = \eta(m_k(x^k) - m_k(z^k + t\Delta)), \]
completing the proof. ☐

In the next two results we shall use the filter slack $H_k$, defined in (2.1). First we show that, near a feasible non-stationary point, the rejection of a step is due to a large increase of the infeasibility.

**Lemma 3.6.** Suppose that Hypotheses H1, H2, H4-H6 hold. Let $\bar{x} \in X$ be a feasible non-stationary point and consider the constants $\gamma$ and $\Delta_p$ given by Algorithm 3.1 and Lemma 3.4, respectively and the neighborhood $V_3$ given by Lemma 3.5. Then there exists a neighborhood $V \subset V_3$ of $\bar{x}$ such that whenever $x^k \in V$, $z^k = x^k + t\Delta$ and a tangential trial step $t\Delta$ is obtained by the algorithm, we have
\[ h(z^k + t\Delta) \geq H_k \]
for any $\Delta \in [\gamma^2 \Delta_p, \Delta_p]$ that was rejected by Algorithm 3.1.

**Proof.** Let $\alpha$, $\eta$, $N$ and $\tilde{c}$ be the constants given by Algorithms 2.1, 3.1 and Lemmas 3.3, 3.4, respectively. Consider $V \subset V_3$ a neighborhood of $\bar{x}$ such that for all $x \in V$,
\begin{equation}
(3.16) \quad Nh(x) \leq \frac{1}{2}\tilde{c}\gamma^2 \Delta_p \quad \text{and} \quad \alpha h(x) \leq \frac{1}{2}\eta\tilde{c}\gamma^2 \Delta_p.
\end{equation}
Hence, if \( x^k \in V \) and \( \Delta \in [\gamma^2 \Delta_\rho, \Delta_\rho] \), we can apply Lemma 3.3 to obtain
\[
|m_k(x^k) - m_k(z^k)| \leq Nh(x^k) \leq \frac{1}{2} \tilde{c} \gamma^2 \Delta_\rho \leq \frac{1}{2} \tilde{c} \Delta,
\]
which together with Lemma 3.4 yields
\[
m_k(x^k) - m_k(z^k + t_\Delta) \geq \frac{1}{2} \tilde{c} \Delta \geq \frac{1}{2} \tilde{c} \gamma^2 \Delta_\rho.
\]
(3.17)

Using Lemma 3.5, (3.16) and (3.17), we obtain
\[
f_0(x^k) - f_0(x^k + d_\Delta) \geq \eta (m_k(x^k) - m_k(x^k + d_\Delta))
\]
\[
\geq \frac{1}{2} \eta \tilde{c} \gamma^2 \Delta_\rho.
\]
(3.18)

Therefore, if the trial step \( d_\Delta \) was rejected by Algorithm 3.1, then \( x^k + d_\Delta \in \bar{F}_k \) because of (3.18). We thus conclude from (3.19) that
\[
h(x^k + t_\Delta) \geq H_k,
\]
completing the proof. \( \Box \)

We now prove the main result of this section: Hypothesis H3 is satisfied. Indeed, we give a sufficient condition to ensure H3. As we saw in Theorem 2.5, this hypothesis was crucial in the convergence analysis of Section 2.

For the purpose of our analysis, we shall consider the set of restoration iterations
\[
K_r = \{ k \in \mathbb{N} \mid L(x^k) = \emptyset \text{ or } \|n^k\| > \xi \Delta_k \},
\]
where \( L(x^k) \) is defined by (3.2). We also assume the following hypothesis.

H7. Every feasible accumulation point \( \bar{x} \in X \) of \((x^k)_{k \in \mathbb{N}}\) satisfies the Mangasarian-Fromovitz constraint qualification, 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 \( A_\mathcal{I}(\bar{x})d < 0 \), where \( \mathcal{I} = \{ i \in \mathcal{I} \mid f_i(\bar{x}) = 0 \} \).

Theorem 3.7. Suppose that Algorithm 2.1 is applied to problem \((P)\), with the step computed by Algorithm 3.1, and that Hypotheses H1, H2, H4-H7 hold. Given a feasible non-stationary point \( \bar{x} \in X \), there exist \( M > 0 \) and a neighborhood \( V \) of \( \bar{x} \) such that if \( x^k \in V \), then
\[
f_0(x^k) - f_0(x^{k+1}) \geq M \sqrt{H_k}.
\]
In particular, since \( \sqrt{H_k} \geq v_k \), the hypothesis H3 is satisfied.
Proof. Let $\bar{x}$ be a feasible non-stationary point. Consider the neighborhood $V$ given by Lemma 3.6 and the constant $\Delta_r$ given by Lemma 3.4. Without loss of generality, we can assume that

$$\Delta_r \leq \frac{\gamma^2}{c_h} \min \left\{ \frac{\xi}{c_n}, \frac{\bar{c} \gamma}{2N}, \frac{\eta \bar{c}}{2\alpha} \right\},$$

where $\alpha$ is the constant given in Algorithm 2.1, $\xi$, $\gamma$, $c_p$ and $\eta$ are given in Algorithm 3.1, $c_n$ is given in Hypothesis H4 and $c_h$, $N$ and $\bar{c}$ are given by Lemmas 3.2, 3.3 and 3.4, respectively. By the constraint qualification hypothesis H7, we can assume that if $x^k \in V$, then $\mathcal{L}(x^k) \neq \emptyset$. Thus, Algorithm 3.1 starts with the radius $\Delta \geq \Delta_{\text{min}}$ and ends with $\Delta_k = \gamma^r \Delta$, where $r$ is the number of times that the radius was reduced in the algorithm. We shall consider two cases, respectively with $\Delta_k \geq \gamma^2 \Delta_r$ and $\Delta_k < \gamma^2 \Delta_r$.

First case: $\Delta_k \geq \gamma^2 \Delta_r$. In this case, using the hypothesis H4 and restricting the neighborhood $V$, if necessary, we have

$$\|n^k\| \leq c_nh(x^k) \leq \xi \gamma^2 \Delta_r \leq \xi \Delta_k.$$ 

So, Algorithm 3.1 does not enter the restoration phase during the iteration $k$, that is, $k \notin K_r$. Therefore, applying Lemma 3.4(i) with $\Delta' = \gamma^2 \Delta_r$, we obtain

$$m_k(z^k) - m_k(x^{k+1}) = m_k(z^k) - m_k(z^k + t\Delta_k) \geq \bar{c} \gamma^2 \Delta_r.$$ 

On the other hand, by Lemma 3.3(i),

$$|m_k(z^k) - m_k(z^k)| \leq Nh(x^k).$$

We can restrict again the neighborhood $V$, if necessary, so that

$$Nh(x^k) \leq \frac{1}{2} \bar{c} \gamma^2 \Delta_r, \quad c_p(h(x^k))^2 \leq \frac{1}{2} \bar{c} \gamma^2 \Delta_r \quad \text{and} \quad h(x^k) \leq 1.$$ 

By (3.22)-(3.24), we have

$$\text{pred}_k \overset{\text{def}}{=} m_k(x^k) - m_k(x^{k+1}) \geq \frac{1}{2} \bar{c} \gamma^2 \Delta_r \geq c_p(h(x^k))^2.$$ 

Then the mechanism of Algorithm 3.1 and the fact that $H_k \leq 1$ imply that

$$f_0(x^k) - f_0(x^{k+1}) \overset{\text{def}}{=} \text{ared}_k \geq \eta \text{pred}_k \geq \frac{1}{2} \eta \bar{c} \gamma^2 \Delta_r \geq \frac{1}{2} \eta \bar{c} \gamma^2 \Delta_r \sqrt{H_k}.$$ 

Second case: now, assume that $\Delta_k < \gamma^2 \Delta_r$. In this case we shall analyze two possibilities. In the first one, we suppose that $h(x^k + d\Delta) \geq H_k$ for all $\Delta \leq \gamma \Delta_r$ such that the trial step $d\Delta$ has been computed. Let $\bar{\Delta} = \frac{\Delta_k}{\gamma}$. Since $\Delta_k < \Delta_{\text{min}}$, the trial step $d\bar{\Delta} = \bar{d}$ was computed. Furthermore, $h(x^k + \bar{d}) \geq H_k$ because $\bar{\Delta} < \gamma \Delta_r$. So, using Lemma 3.2 and the definition of $H_k$, it follows that

$$c_h \Delta_k^2 = c_h \gamma^2 \bar{\Delta}^2 \geq \gamma^2 h(x^k + \bar{d}) \geq \gamma^2 H_k \geq \gamma^2 h(x^k).$$ 

From Hypothesis H4, (3.21) and (3.26), we obtain

$$\|n^k\| \leq c_nh(x^k) \leq \frac{Cn c_h}{\gamma^2} \Delta_k^2 \leq \xi \Delta_k.$$
meaning that Algorithm 3.1 does not enter the restoration phase during the iteration \(k\), that is, \(k \not\in K_r\). Therefore, by Lemma 3.4(i) with \(\Delta' = \Delta_k\), we have

\[
m_k(z^k) - m_k(x^{k+1}) = m_k(z^k) - m_k(z^k + t_{\Delta_k}) \geq \tilde{c} \Delta_k.
\]

Moreover, (3.23) remains true in this case and together with (3.21) and (3.26) yields

\[
|m_k(x^k) - m_k(z^k)| \leq Nh(x^k) \leq \frac{Nc_h}{\gamma} \Delta_k^2 \leq \frac{1}{2} \tilde{c} \Delta_k.
\]

Combining (3.27) and (3.28), we obtain

\[
\text{pred}_k = m_k(x^k) - m_k(x^{k+1}) \geq \frac{1}{2} \tilde{c} \Delta_k.
\]

By (3.21), (3.24) and (3.26),

\[
\text{pred}_k \geq \frac{c_p c_h}{\gamma^2} \Delta_k^2 \geq c_p h(x^k) \geq c_p (h(x^k))^2.
\]

Thus, the mechanism of Algorithm 3.1, (3.26) and (3.29) imply that

\[
f_0(x^k) - f_0(x^{k+1}) = \text{ared}_k \geq \eta \text{pred}_k \geq \frac{1}{2} \eta \tilde{c} \Delta_k \geq \frac{\eta \tilde{c} \gamma}{2 \sqrt{c_h}} \sqrt{H_k}.
\]

Let us see now the second possibility, that is, there exists \(\Delta \leq \gamma \Delta_p\) such that \(h(x^k + d\Delta) < H_k\). Let \(\hat{\Delta}\) be the first \(\Delta\) satisfying such a condition. We shall show that \(\hat{\Delta} = \Delta_k\). Let \(d = d_{\hat{\Delta}}\) be the trial step obtained with \(\hat{\Delta} = \frac{\Delta}{\gamma}\). We claim that

\[
h(x^k + d) \geq H_k.
\]

Indeed, if \(\Delta \leq \gamma \Delta_p\), the definition of \(\hat{\Delta}\) ensures the claim. On the other hand, if \(\hat{\Delta} > \gamma \Delta_p\), then \(\hat{\Delta} \in [\gamma^2 \Delta_p, \Delta_p]\) and, applying Lemma 3.6, we have

\[
h(x^k + d) = h(x^k + t_{\hat{\Delta}}) \geq H_k.
\]

So, the inequality (3.31) holds. As above, we can prove that

\[
c_h \hat{\Delta}^2 \geq \gamma^2 H_k \geq \gamma^2 h(x^k)
\]

and

\[
\text{pred}_{\hat{\Delta}} \overset{\text{def}}{=} m_k(x^k) - m_k(z^k + t_{\hat{\Delta}}) \geq \frac{1}{2} \tilde{c} \hat{\Delta}.
\]

Now, by the same reasoning as in the proof of Lemma 3.5(ii), using (3.32) and (3.33), we obtain

\[
\text{ared}_{\hat{\Delta}} \overset{\text{def}}{=} f_0(x^k) - f_0(z^k + t_{\hat{\Delta}}) \geq \eta \text{pred}_{\hat{\Delta}} \geq \frac{1}{2} \eta \tilde{c} \hat{\Delta},
\]
which together with (3.21) and (3.32) yields

\[ \text{ared}_\Delta > \frac{\alpha c_h}{\gamma^2} \hat{\Delta}^2 \geq \alpha h(x^k). \]

The definition of \( \hat{\Delta} \) and (3.35) ensure that \( z^k + t_\Delta \) is accepted by the filter. Therefore, using (3.34), we conclude that \( z^k + t_\Delta = x^{k+1} \). Moreover, (3.32) and (3.34) imply that

\[ f_0(x^k) - f_0(z^k + t_\Delta) \geq \frac{1}{2} \eta \hat{c} \hat{\Delta} \geq \frac{\eta \hat{c} \gamma}{2 \sqrt{c_h}} \sqrt{H_k}, \]

that is,

\[ f_0(x^k) - f_0(x^{k+1}) \geq \frac{\eta \hat{c} \gamma}{2 \sqrt{c_h}} \sqrt{H_k}. \]

Since (3.25), (3.30) and (3.36) run out all possibilities, by defining

\[ M = \min \left\{ \frac{1}{2} \eta \hat{c} \gamma^2 \Delta_p, \frac{\eta \hat{c} \gamma}{2 \sqrt{c_h}} \right\}, \]

we complete the proof. \( \square \)

4. Conclusions. In this work we have studied filter methods for nonlinear programming. These methods seem to be a successful strategy for globalizing algorithms without the use of merit functions. Since its appearance in 1997, the filter technique has been applied to many problems, including sequential linear programming (SLP), sequential quadratic programming (SQP), inexact restoration, interior-point methods, nonlinear systems of equations, unconstrained optimization and nonsmooth convex constrained optimization.

Our purpose here was to present a general globally convergent filter algorithm that leaves the step computation separate from the main algorithm. This technique cleans the convergence analysis and accepts any method for computing the step, as long as this internal algorithm is efficient in the sense that the hypothesis H3, is satisfied. For completeness, we have shown that there are methods which satisfy the referred hypothesis.

Acknowledgements. We thank the referees for their valuable comments and suggestions which very much improved this paper.

REFERENCES