Global convergence of a derivative-free inexact restoration filter algorithm for nonlinear programming

P.S. Ferreira[a] * *and E.W. Karas[b] and M. Sachine[b] and F.N.C. Sobral[c]

[a] *Department of Mathematics, Federal University of Technology – Paraná, Curitiba, PR, Brazil.*

[b] *Department of Mathematics, Federal University of Paraná, Curitiba, PR, Brazil.*

[c] *Department of Mathematics, State University of Maringá, Maringá, PR, Brazil.*

August 30, 2016

Abstract

In this work we present an algorithm for solving constrained optimization problems that does not make explicit use of the objective function derivatives. The algorithm mixes an inexact restoration framework with filter techniques, where the forbidden regions can be given by the flat or slanting filter rule. Each iteration is decomposed in two independent phases: a feasibility phase which reduces an infeasibility measure without evaluations of the objective function, and an optimality phase which reduces the objective function value. As the derivatives of the objective function are not available, the optimality step is computed by derivative-free trust-region internal iterations. Any technique to construct the trust-region models can be used since the gradient of the model is a reasonable approximation of the gradient of the objective function at the current point. Assuming this and classical assumptions, we prove that the full steps are efficient in the sense that near a feasible nonstationary point, the decrease in the objective function is relatively large, ensuring the global convergence results of the algorithm. Numerical experiments show the effectiveness of the proposed method.

Keywords: *Derivative-free optimization; inexact restoration; filter methods; global convergence; numerical experiments*

AMS Subject Classification: 49M37; 65K05; 90C30; 90C56

1 Introduction

In this work we discuss the global convergence of a derivative-free filter method for solving the nonlinear programming problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c_{\mathcal{E}}(x) = 0 \\ & c_{\mathcal{I}}(x) \leq 0, \end{array} \tag{1}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the functions $c_i : \mathbb{R}^n \to \mathbb{R}$, for $i \in \mathcal{E} \cup \mathcal{I}$, that define the constraints are continuously differentiable. We assume that the derivatives of the constraints are available whereas the derivatives of the objective function are not. The feasible set of the problem is given by

$$\Omega = \{x \in \mathbb{R}^n \mid c_{\mathcal{E}}(x) = 0 \text{ and } c_{\mathcal{I}}(x) \leq 0\}.$$

There are many applications of derivative-free optimization, particularly when the objective function is provided by a simulation package or a black box, these and more cases can be

**Corresponding author. Email: psferreira@utfpr.edu.br

seen in the book [18]. Such situations motivated researchers to pursue techniques for derivative-free optimization.

Several derivative-free methods have been developed for unconstrained problems [15, 19, 26, 27, 61, 64], box-constrained problems [18, 37, 44, 62, 66], linearly constrained problems [42, 45, 63], convex constrained problems [13], composite nonsmooth optimization [36].

Regarding the class of problems we address in this paper which involves equality and/or inequality constraints, many derivative-free techniques have been used. Methods based on feasible iterates [5, 43, 51] may not work well in the presence of nonlinear equality constraints or thin domains. Penalty methods [25, 48] and augmented Lagrangian methods [21, 46, 47] penalize, generally, "difficult" constraints and solve box or linear constrained subproblems. These methods evaluate function and constraints at the same points which can cause the necessity of performing unnecessary evaluations in the presence of topologically complex constraints. In contrast, derivative-free two-phase algorithms [38, 54] deal with cases where finding a more feasible point is easier than minimizing the objective function. Inexact restoration approaches [28, 52, 53] were proposed in a derivative-free context in which the progress of the algorithm is measured by a merit function [3, 8]. Since adjusting the penalty parameters of merit functions or penalty functions can be a difficult task, filters have been suggested in literature as an alternative.

Filter methods were initially proposed by Fletcher and Leyffer [30] to solve nonlinear programming problems. Chin and Fletcher [9] considered the slanting filter, which is a slight modification of the original flat filter. These methods have been combined to trust-region approaches [59, 65], SQP techniques [29, 73], inexact restoration algorithms [32, 40], interior point strategies [70] and line-search algorithms [39, 57, 72]. They also have been extended to other areas of optimization such as nonlinear equations and inequalities [23, 31, 33, 35], nonsmooth optimization [41, 58], unconstrained optimization [34, 74], complementarity problems [49, 50] and derivative-free optimization [4, 24].

In this work we propose a derivative-free inexact restoration algorithm for general constrained problems using the flat or the slanting filter. Each iteration is composed of two phases. First, a feasibility step is computed from the current point in order to obtain a restored point that reduces an infeasibility measure h. In this phase, basically any method for reducing h can be used [23, 52, 53]. Next, from the restored point, a trust-region [14, 56, 67, 75] optimality phase computes a point which is not forbidden by the filter and that reduces the objective function value. Linear or quadratic models that approximate the objective function based only on zero-order information are considered. In our analysis, the Hessians of the models must be bounded symmetric matrices and the gradients must represent properly the gradients of the objective function. Models satisfying these properties can be constructed by many derivative-free techniques, such as polynomial interpolation [11, 17, 18, 20] and support vector regression [71]. Under classical assumptions, global convergence results are obtained. Numerical results illustrate the performance of the proposed algorithm for a set of test problems from the Hock-Schittkowski collection [69].

A derivative-free inexact restoration filter algorithm has also been proposed in [24]. It does not use derivatives of the constraints, but uses only linear models for the objective function, deals with equality constraints and considers just the flat filter. It is worthwhile to mention that the use of the slanting filter in the present work allow us to prove stronger convergence results, namely all accumulation points are stationary. Moreover if some constraints do not have available derivatives or their evaluations are expensive, it is sensible to include these constraints in the objective function in an augmented Lagrangian context [1].

The paper is organized as follows. Section 2 describes the derivative-free inexact restoration filter algorithm. Section 3 shows that the step computed in each iteration is efficient, ensuring the global convergence results. Numerical experiments are discussed in Section 4. Final remarks are presented in Section 5.

2 The algorithm

This section presents a derivative-free inexact restoration filter algorithm to solve the problem (1). The sequence of points generated by the algorithm will be denoted by $(x^k)_{k\in\mathbb{N}}$. In order to prove the global convergence of the algorithm, the following standard assumptions are considered.

A1 *All the functions* f, c_i, *for* $i \in \mathcal{E} \cup \mathcal{I}$, *are continuously differentiable.*

A2 *The sequence* (x^k) *remains in a convex compact set* $X \subset \mathbb{R}^n$.

A3 *The gradient* ∇f *is Lipschitz continuous in an open set containing* X, *that is, there exists a constant* $L > 0$ *such that*

$$\|\nabla f(x) - \nabla f(y)\| \le L \|x - y\|$$

for all x *and* y *in the open set containing* X.

Although A2 is an assumption on the sequence generated by the algorithm, it can be enforced by including a bounded box into the problem constraints.

We define an infeasibility measure function $h : \mathbb{R}^n \to \mathbb{R}_+$ by

$$h(x) = \left\|c^+(x)\right\|,$$

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

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

Note that $h(x) = 0$ if and only if $x \in \mathbb{R}^n$ is a feasible point.

We say that a feasible point $\bar{x}$ is a *stationary point* for the original problem (1) when

$$\liminf_{x\to\bar{x}} \left\|\mathcal{P}_{\mathcal{L}(x)}(x - \nabla f(x)) - x\right\| = 0, \tag{2}$$

where $\mathcal{P}_{\mathcal{L}(x)}$ denotes the orthogonal projection onto the set $\mathcal{L}(x)$ defined by

$$\mathcal{L}(x) = \{x + d \in \mathbb{R}^n \mid Jc_{\mathcal{E}}(x)d = 0 \text{ and } c_{\mathcal{I}}(x) + Jc_{\mathcal{I}}(x)d \le c_{\mathcal{I}}^+(x)\},$$

where $Jc_{\mathcal{E}}(.)$ and $Jc_{\mathcal{I}}(.)$ are the Jacobian matrices of the constraints $c_{\mathcal{E}}$ and $c_{\mathcal{I}}$, respectively. At a feasible point x, the set $\mathcal{L}(x)$ is a linearization of the feasible set.

To simplify the notation we denote $f(x^k)$ and $h(x^k)$ by f^k and h^k, respectively. Each iteration k of the algorithm is composed of a feasibility phase which reduces the infeasibility measure without evaluations of the objective function and an optimality phase which reduces the objective function. These phases are independent and the coupling between them is provided by a filter F_k which is a set of pairs (f^j, h^j) from well-chosen former iterations. Given $\alpha \in (0, 1)$, the filter defines a *forbidden region* $\mathcal{F}_k = \left\{\cup\mathcal{R}_j \mid (f^j, h^j) \in F_k\right\}$ where $\mathcal{R}_j$ is given either by

$$\overline{\mathcal{R}}_j = \left\{x \in \mathbb{R}^n \;\middle|\; f(x) \ge f^j - \alpha h^j \text{ and } h(x) \ge (1 - \alpha)h^j\right\}, \tag{3}$$

as suggested originally in [30], or by

$$\grave{\mathcal{R}}_j = \left\{x \in \mathbb{R}^n \;\middle|\; f(x) + \alpha h(x) \ge f^j \text{ and } h(x) \ge (1 - \alpha)h^j\right\}, \tag{4}$$

as proposed in [10]. A filter based on the rule (3) will be referred as *flat filter* and the one based on (4) will be called *slanting filter*. Note that the slanting filter satisfies the following inclusion property

$$f^j = f^i \quad \text{and} \quad h^j > h^i \quad \Rightarrow \quad \grave{\mathcal{R}}_j \subset \grave{\mathcal{R}}_i,$$

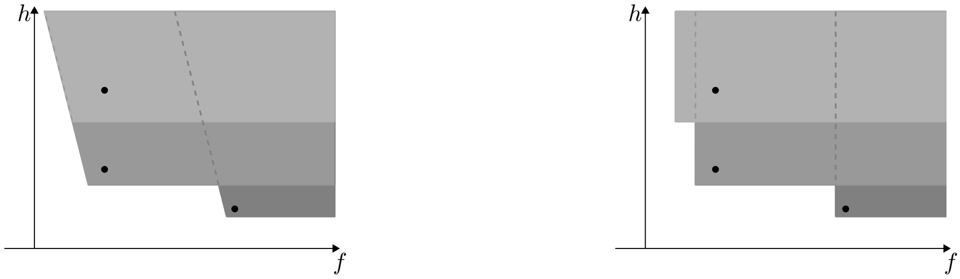


Figure 1: The inclusion property of the slanting filter, which is not satisfied for the flat filter.

as illustrates Figure 1 which also shows that this property does not hold for the flat filter. The slanting filter allows a stronger statement about the convergence of the algorithm as will be shown ahead.

From the current point x^k, the feasibility phase computes a restored point $z^k \notin \widehat{\mathcal{F}}_k$ satisfying

$$h(z^k) < (1-\alpha)h(x^k) \qquad \text{and} \qquad \left\| z^k - x^k \right\| \leq \beta h(x^k), \tag{5}$$

where $\widehat{\mathcal{F}}_k = \mathcal{F}_k \cup \mathcal{R}_k$ and $\beta > 0$. The procedure used in this phase could, in principle, be any iterative algorithm for decreasing h, and finite termination should be achieved because all filter entries $(f^j, h^j) \in F_k$ have $h^j > 0$ (see [59, Lemma 2.1]). As the feasibility step studied by Martínez [52] can be applied directly to our case, 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 without success.

Once z^k is computed, the optimality phase must find a point $x^{k+1} = z^k + d^k \notin \widehat{\mathcal{F}}_k$, with $z^k + d_k \in \mathcal{L}(z^k)$, such that $f(x^{k+1}) < f(z^k)$. Within the optimality phase, we will perform internal trust-region iterations k_j for $j \in \mathbb{N}$, with radius Δ_{k_j}. The quadratic model $m_{k_j} : \mathbb{R}^n \to \mathbb{R}$ of f around the restored point z^k is defined by

$$m_{k_j}(x) = f(z^k) + (x - z^k)^T g_{k_j} + \frac{1}{2}(x - z^k)^T B_{k_j}(x - z^k),$$

where $g_{k_j} \in \mathbb{R}^n$ and $B_{k_j} \in \mathbb{R}^{n \times n}$ is a symmetric matrix. The model is updated at the beginning of each iteration k_j and its quality is controlled by a second radius $\delta_{k_j} \in (0, \Delta_{k_j}]$. To ensure a good approximation for f near the restored point z^k, we require that each quadratic model m_{k_j} satisfies the following condition:

A4 *There exist constants $\gamma > 0$ and $\sigma > 0$ such that*

$$\left\| B_{k_j} \right\| \leq \gamma \quad \text{and} \quad \left\| g_{k_j} - \nabla f(z^k) \right\| \leq \sigma \delta_{k_j}$$

for all $k \in \mathbb{N}$, $j \in \mathbb{N}$ and $\delta_{k_j} > 0$.

Observe that once z^k is fixed, this assumption is responsible for updating the models in each internal iteration k_j, since $g_{k_j} = g(z^k, \delta_{k_j})$ and consequently $m_{k_j}(\,.\,) = m(\,.\,, z^k, \delta_{k_j})$. This dependence of the models with each radius δ_{k_j} occurs only in the gradients. The Hessians can always be the same, in particular null for linear models. There are algorithms able to find models with such properties without computing $\nabla f(z^k)$, for instance [18, Chapter 6]. Any technique that fulfills Assumption A4 can be used in the optimality step, although in literature the most usual procedure is polynomial interpolation [18, 26, 68].

Each internal iteration k_j should compute a step d^{k_j} such that $\|d^{k_j}\| \leq \Delta_{k_j}$ and the point $z^k + d^{k_j} \in \mathcal{L}(z^k)$ provides a sufficient reduction in the objective function value. We define the *actual reduction* provided by this step by

$$\text{ared}_{k_j} = f(z^k) - f\left(z^k + d^{k_j}\right) \tag{6}$$

and the *predicted reduction* by

$$\text{pred}_{k_j} = m_{k_j}(z^k) - m_{k_j}\left(z^k + d^{k_j}\right). \tag{7}$$

The step d^{k_j} can be any approximate solution of the trust-region subproblem

$$\begin{array}{ll} \text{minimize} & m_{k_j}(z^k + d) \\ \text{subject to} & z^k + d \in \mathcal{L}(z^k) \\ & \|d\| \leq \Delta_{k_j}, \end{array} \tag{8}$$

since

$$\text{pred}_{k_j} \geq \xi \pi_{k_j}(z^k) \min\left\{\frac{\pi_{k_j}(z^k)}{1 + \|B_{k_j}\|}, \Delta_{k_j}\right\}, \tag{9}$$

where $\xi > 0$ is a constant independent of k and j and

$$\pi_{k_j}(x) = \left\|\mathcal{P}_{\mathcal{L}(x)}(x - \nabla m_{k_j}(x)) - x\right\|$$

is the measure of stationarity at a point x for the problem of minimizing m_{k_j} over the set $\mathcal{L}(x)$. To satisfy (9), the approximate solution d^{k_j} has only to achieve a reduction that is at least some fixed fraction ξ of the reduction achieved by the Cauchy point [18, Theorem 10.1].

The algorithm may be stated in the following form.

Algorithm 1

Given: $x^0 \in \mathbb{R}^n$, $\alpha \in (0,1), \eta \in (0,1)$, $\varepsilon > 0, \beta > 0, \Delta_{\min} > 0$, $\mu > 0$
Set $F_0 = \emptyset$, $\mathcal{F}_0 = \emptyset$, $k = 0$
REPEAT
  Define $\widehat{F}_k = F_k \cup \left\{\left(f^k, h^k\right)\right\}$ and
  $\widehat{\mathcal{F}}_k = \mathcal{F}_k \cup \mathcal{R}_k$, where either $\mathcal{R}_k = \overline{\mathcal{R}}_k$ or $\mathcal{R}_k = \grave{\mathcal{R}}_k$ as given in (3) and (4)
  Computing the step
    FEASIBILITY PHASE
      If $h(x^k) = 0$, then $z^k = x^k$
      Else, compute $z^k \notin \widehat{\mathcal{F}}_k$ satisfying (5)
        If impossible, then stop without success
    OPTIMALITY PHASE
      Set $j = 0$, choose $\Delta_{k_0} \geq \Delta_{\min}$ and set $\delta_{k_0} = \Delta_{k_0}$
      REPEAT
        Construct the model m_{k_j}
        If $\delta_{k_j} > \mu \pi_{k_j}(z^k)$, then
          If $\delta_{k_j} \leq \varepsilon$ and $h(z^k) = 0$, then stop the algorithm with success
          Else, set $d^{k_j} = 0$, $\delta_{k_{j+1}} = \delta_{k_j}/2$ and choose $\Delta_{k_{j+1}} \in [\delta_{k_{j+1}}, \Delta_{k_j}]$
        Else
          Compute d^{k_j} as an approximate solution of (8)
          If $z^k + d^{k_j} \notin \widehat{\mathcal{F}}_k$ and $\text{ared}_{k_j} > \eta\ \text{pred}_{k_j}$
            Set $x^{k+1} = z^k + d^{k_j}$, $d^k = d^{k_j}$, $\Delta_k = \Delta_{k_j}$, $\delta_k = \delta_{k_j}$
            and stop the optimality phase with success
          Else, set $\delta_{k_{j+1}} = \delta_{k_j}/2$ and $\Delta_{k_{j+1}} = \Delta_{k_j}/2$
        $j = j + 1$
      END
  Filter update
    If $f(x^{k+1}) < f(x^k)$, then
      $F_{k+1} = F_k, \quad \mathcal{F}_{k+1} = \mathcal{F}_k$ ($f-$iteration)
    Else,

$$F_{k+1} = \widehat{F}_k, \quad \mathcal{F}_{k+1} = \widehat{\mathcal{F}}_k \qquad (h\text{−iteration})$$

$k = k + 1$

END

At the beginning of each iteration, the pair (f^k, h^k) is temporarily introduced in the filter. This pair helps to define the forbidden region $\mathcal{R}_k = \overline{\mathcal{R}}_k$ as given in (3) or $\mathcal{R}_k = \grave{\mathcal{R}}_k$ as given in (4) depending on the considered filter rule. After the iteration is completed, the pair (f^k, h^k) will become permanent in the filter only if the iteration does not produce a decrease in f, that is, if k is an h-iteration. When k is an f-iteration, the new entry (f^k, h^k) is discarded and the filter is not updated. Note that if x^k is feasible then any point x that is not forbidden must satisfy $f(x) < f(x^k)$. The pairs forbidden by the filter and the permanent F_k and temporary $\widehat{F}_k$ filters can be seen in the Figure 2.

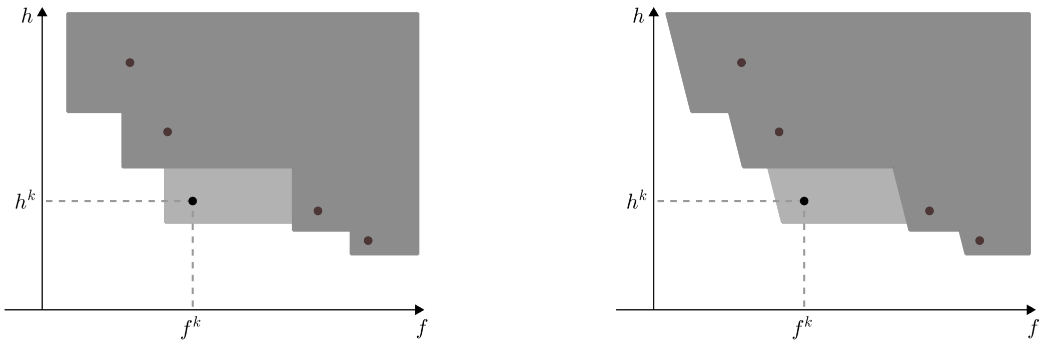


Figure 2: Permanent and temporary filters at the beginning of the iteration k.

3 Global convergence

The task of this section is to ensure the global convergence of the Algorithm 1, i.e. to show that the sequence generated by the algorithm has a stationary accumulation point.

Although Algorithm 1 decomposes each iteration in two phases, it essentially consists of calculating a point x^{k+1} not forbidden by the filter from the current point x^k. Furthermore, the construction and updating rule of the filter are made by the same way considered in the general filter algorithm [59, Alg. 1]. So, our algorithm fits in the general filter algorithm from that paper, which considers classical hypotheses and the following efficiency condition in order to prove the global convergence.

Efficiency condition on the step. Given a feasible nonstationary point $\bar{x} \in X$, there exist $M > 0$ and a neighborhood V of $\bar{x}$ such that for any iterate $x^k \in V$,

$$f(x^k) - f(x^{k+1}) \geq M v_k,$$

where v_k is the filter height given by

$$v_k = \min\left\{1, \min\left\{(1-\alpha)h^j \mid (f^j, h^j) \in F_k\right\}\right\}.$$

In [65], a filter slack is defined as

$$H_k = \min\left\{1, \min\left\{(1-\alpha)h^j \;\middle|\; (f^j, h^j) \in F_k \text{ and } f^j \leq f^k\right\}\right\}.$$

Note that $H_k \geq v_k$ as shown in Figure 3.

So, our task now is to prove that our algorithm satisfies this efficiency condition, inheriting the global convergence results from [59]. First, we show that if the algorithm has successful finite termination then an approximate stationary point is obtained.

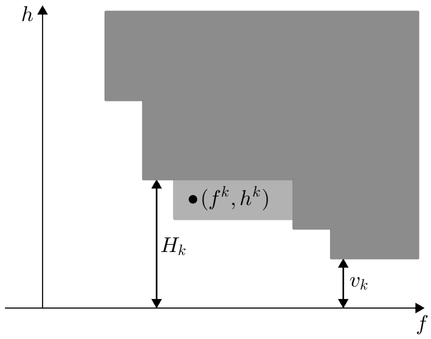


Figure 3: The difference between the filter slack H_k and the filter height v_k.

Lemma 3.1 *If the algorithm stopped with success, then a stationary point was obtained with sufficient accuracy.*

Proof. Suppose that the algorithm stopped with z^k at the internal iteration k_j. By the triangle inequality and the contraction property of projections we have

$$\begin{aligned}\left\|\mathcal{P}_{\mathcal{L}(z^k)}(z^k - \nabla f(z^k)) - z^k\right\| &\leq \\ \leq \left\|\mathcal{P}_{\mathcal{L}(z^k)}(z^k - g_{k_j}) - z^k\right\| &+ \left\|\mathcal{P}_{\mathcal{L}(z^k)}(z^k - \nabla f(z^k)) - \mathcal{P}_{\mathcal{L}(z^k)}(z^k - g_{k_j})\right\| \\ \leq \left\|\mathcal{P}_{\mathcal{L}(z^k)}(z^k - g_{k_j}) - z^k\right\| &+ \left\|\nabla f(z^k) - g_{k_j}\right\|.\end{aligned}$$

Using the definition of π_{k_j}, Assumption A4 and the successful finite termination criterion we obtain

$$\left\|\mathcal{P}_{\mathcal{L}(z^k)}(z^k - \nabla f(z^k)) - z^k\right\| \leq \pi_{k_j}(z^k) + \sigma\delta_{k_j} \leq \left(\frac{1}{\mu} + \sigma\right)\varepsilon,$$

which implies that z^k is an approximate stationary point. □

From now on we assume that the algorithm has generated infinite sequences (x^k) and (z^k) and that Assumptions A1-A4 are satisfied. The following lemma states that the number of internal iterations that just construct the models but do not compute trust-region steps is finite.

Lemma 3.2 *Consider z^k a nonstationary point. Then the set*

$$\mathscr{J}_k = \{j \in \mathbb{N} \mid \delta_{k_j} > \mu\pi_{k_j}(z^k)\} \tag{10}$$

is finite.

Proof. Suppose by contradiction that $\mathscr{J}_k$ is infinite. By the algorithm we have $\delta_{k_j} = \left(\frac{1}{2}\right)^j \delta_{k_0} > 0$ and consequently

$$\delta_{k_j} \longrightarrow 0. \tag{11}$$

Thus, by the definition of $\mathscr{J}_k$,

$$\pi_{k_j}(z^k) \xrightarrow{j \in \mathscr{J}_k} 0. \tag{12}$$

On the other hand, since z^k is nonstationary, by (2) we have $\left\|\mathcal{P}_{\mathcal{L}(z^k)}(z^k - \nabla f(z^k)) - z^k\right\| = \bar{c} > 0$. By (11), there exists $j_0 \in \mathbb{N}$ such that for all $j > j_0$, $\delta_{k_j} < \dfrac{\bar{c}}{2\sigma}$. Using these facts, the contraction property of projections and Assumption A4 we have for all $j > j_0$

$$\begin{aligned}\pi_{k_j}(z^k) &= \left\|\mathcal{P}_{\mathcal{L}(z^k)}(z^k - g_{k_j}) - \mathcal{P}_{\mathcal{L}(z^k)}(z^k - \nabla f(z^k)) + \mathcal{P}_{\mathcal{L}(z^k)}(z^k - \nabla f(z^k)) - z^k\right\| \\ &\geq -\left\|g_{k_j} - \nabla f(z^k)\right\| + \left\|\mathcal{P}_{\mathcal{L}(z^k)}(z^k - \nabla f(z^k)) - z^k\right\| > \bar{c} - \sigma\delta_{k_j} > \frac{\bar{c}}{2} > 0,\end{aligned}$$

which contradicts (12). $\square$

The requirement that $\delta_{k_j} \leq \mu \pi_{k_j}(z^k)$ in the trust-region steps is inherently related to the fact that the models should be reasonable approximations of the objective function in a neighborhood of the current point. Note that when $\pi_{k_j}(z^k)$ is small, the current point is probably close to a solution of the subproblem (8). On the other hand, if the radius δ_{k_j} is large, we cannot guarantee that the objective function is well represented by the model.

Now we present an auxiliary result.

Lemma 3.3 *[32, Lemma 3.3] There exists a constant $C_1 > 0$ such that, for any $z \in X$ and $z + d \in \mathcal{L}(z)$,*

$$|h(z+d) - h(z)| \leq C_1 \|d\|^2 .$$

From now on we also assume the following classical constraint qualification [56].

A5 *Every feasible point $\overline{x}$ satisfies the Mangasarian-Fromovitz constraint qualification, i.e., there exists a vector $d \in \mathbb{R}^n$ such that $\nabla c_i(\overline{x})^t d = 0$ and $\nabla c_j(\overline{x})^t d < 0$, for all $i \in \mathcal{E}$ and $j \in \{\mathcal{I} \mid c_j(\overline{x}) = 0\}$ and the set of equality constraint gradients $\{\nabla c_i(\overline{x}) \mid i \in \mathcal{E}\}$ is linearly independent.*

The next result ensures the continuity of some auxiliary functions.

Lemma 3.4 *[32, Lemmas A.1 and A.2] Consider a point $\overline{x}$ that satisfies the Mangasarian-Fromovitz constraint qualification and $p : \mathbb{R}^n \to \mathbb{R}^n$ a continuous function at $\overline{x}$. Then the point to set map $\mathcal{L}(.)$ and the function $\mathcal{P}_{\mathcal{L}(.)}(p(.))$ are continuous at $\overline{x}$.*

An auxiliary result involving the measure of stationarity is presented below.

Lemma 3.5 *Let $\overline{x} \in X$ be a nonstationary feasible point, $\overline{c} = \frac{1}{4}\left\|\mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla f(\overline{x})) - \overline{x}\right\| > 0$ and $\sigma > 0$ given by A4. Then there exists a neighborhood V_1 of $\overline{x}$ such that for any $z^k \in V_1$ and $j \in \mathbb{N}$*

$$\pi_{k_j}(z^k) > -\sigma \delta_{k_j} + \frac{\overline{c}}{2}.$$

Proof. Let $\overline{x}$ be a nonstationary feasible point. Consider $\widetilde{V}_1 = B\left(\overline{x}, \frac{\overline{c}}{2}\min\{1, 1/\gamma\}\right)$ where $\gamma > 0$ is given by A4. Then, for any $x \in \widetilde{V}_1$ and $k, j \in \mathbb{N}$,

$$\left\|\nabla m_{k_j}(\overline{x}) - \nabla m_{k_j}(x)\right\| = \left\|B_{k_j}(\overline{x} - x)\right\| < \frac{\overline{c}}{2}. \tag{13}$$

Given $v \in \mathbb{R}^n$, by Lemma 3.4, $\mathcal{P}_{\mathcal{L}(.)}(v)$ is continuous at $\overline{x}$ and there exists a neighborhood $\widetilde{V}_2$ of $\overline{x}$ such that for all $z^k \in \widetilde{V}_2$,

$$\left\|\mathcal{P}_{\mathcal{L}(z^k)}(v) - \mathcal{P}_{\mathcal{L}(\overline{x})}(v)\right\| < \frac{\overline{c}}{2}. \tag{14}$$

Since $\left\|\mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla f(.)) - \overline{x}\right\|$ is continuous at $\overline{x}$, there exists a neighborhood $\widetilde{V}_3$ of $\overline{x}$ such that for any $z^k \in \widetilde{V}_3$

$$\left\|\mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla f(z^k)) - \overline{x}\right\| > \frac{3}{4}\left\|\mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla f(\overline{x})) - \overline{x}\right\| = 3\overline{c}. \tag{15}$$

Consider $V_1 = \widetilde{V}_1 \cap \widetilde{V}_2 \cap \widetilde{V}_3$. Using the definition of π_{k_j}, the triangle inequality and the contraction property of projections, it follows that for any $z^k \in V_1$

$$
\begin{aligned}
&\left|\pi_{k_j}(z^k) - \pi_{k_j}(\overline{x})\right| \leq \\
&\quad \leq \left\|\mathcal{P}_{\mathcal{L}(z^k)}(z^k - \nabla m_{k_j}(z^k)) - z^k - \mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla m_{k_j}(\overline{x})) + \overline{x}\right\| \\
&\quad \leq \left\|\mathcal{P}_{\mathcal{L}(z^k)}(z^k - \nabla m_{k_j}(z^k)) - \mathcal{P}_{\mathcal{L}(z^k)}(\overline{x} - \nabla m_{k_j}(\overline{x}))\ \right\| + \\
&\quad\quad + \left\|\mathcal{P}_{\mathcal{L}(z^k)}(\overline{x} - \nabla m_{k_j}(\overline{x})) - \mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla m_{k_j}(\overline{x}))\right\| + \left\|\overline{x} - z^k\right\| \\
&\quad \leq \left\|\nabla m_{k_j}(\overline{x}) - \nabla m_{k_j}(z^k)\right\| + \left\|\mathcal{P}_{\mathcal{L}(z^k)}(\overline{x} - \nabla m_{k_j}(\overline{x})) - \mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla m_{k_j}(\overline{x}))\right\| + \\
&\quad\quad + 2\left\|\overline{x} - z^k\right\|.
\end{aligned}
$$

From this, (13), (14) and definition of $\widetilde{V}_1$,

$$\pi_{k_j}(z^k) > \pi_{k_j}(\overline{x}) - 2\overline{c}. \tag{16}$$

On the other hand, using the definition of π_{k_j} again, the triangle inequality and the contraction property of projections, we have that

$$
\begin{aligned}
\pi_{k_j}(\overline{x}) &= \left\|\mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - g_{k_j} - B_{k_j}(\overline{x} - z^k)) - \mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla f(z^k)) + \mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla f(z^k)) - \overline{x}\right\| \\
&\geq - \left\|\mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - g_{k_j} - B_{k_j}(\overline{x} - z^k)) - \mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla f(z^k))\right\| + \left\|\mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla f(z^k)) - \overline{x}\right\| \\
&\geq - \left\|g_{k_j} + B_{k_j}(\overline{x} - z^k) - \nabla f(z^k)\right\| + \left\|\mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla f(z^k)) - \overline{x}\right\| \\
&\geq - \left\|g_{k_j} - \nabla f(z^k)\right\| - \left\|B_{k_j}(\overline{x} - z^k)\right\| + \left\|\mathcal{P}_{\mathcal{L}(\overline{x})}(\overline{x} - \nabla f(z^k)) - \overline{x}\right\|.
\end{aligned}
$$

Thus, using Assumption A4, (15) and the fact that $z^k \in V_1 \subset \widetilde{V}_1$, we get

$$\pi_{k_j}(\overline{x}) > -\sigma\delta_{k_j} - \gamma\left\|\overline{x} - z^k\right\| + 3\overline{c} \geq -\sigma\delta_{k_j} + \frac{5}{2}\overline{c}.$$

Using this in (16) we obtain the desired result. □

As an immediate result of this lemma we prove that the measure of stationarity around a nonstationary feasible point is bounded below by a positive constant.

Corollary 3.6 *Consider a nonstationary feasible point $\overline{x} \in X$ and the neighborhood V_1 given by Lemma 3.5. Then there exist constants $\widehat{\delta} \in (0, \Delta_{\min})$ and $C_2 > 0$, such that for any $z^k \in V_1$ and*

$$\text{either} \qquad j \notin \mathscr{J}_k \qquad \text{or} \qquad j \in \mathbb{N} \text{ with } \delta_{k_j} \leq \widehat{\delta},$$

we have

$$\pi_{k_j}(z^k) > C_2.$$

Proof. Let $z^k \in V_1$. Consider first that $j \notin \mathscr{J}_k$. By Lemma 3.5 and the definition of $\mathscr{J}_k$ given in (10), it follows that $\pi_{k_j}(z^k) > -\sigma\mu\pi_{k_j}(z^k) + \overline{c}/2$. Then

$$\pi_{k_j}(z^k) > \frac{\overline{c}}{2(1 + \sigma\mu)}.$$

Consider now $\widehat{\delta} = \min\left\{\frac{\overline{c}}{4\sigma}, \frac{\Delta_{\min}}{2}\right\} > 0$ and $j \in \mathbb{N}$ such that $\delta_{k_j} \leq \widehat{\delta}$. By Lemma 3.5

$$\pi_{k_j}(z^k) > -\sigma\widehat{\delta} + \frac{\overline{c}}{2} \geq \frac{\overline{c}}{4}.$$

Setting $C_2 = \min\left\{\frac{1}{2(1 + \sigma\mu)}, \frac{1}{4}\right\}\overline{c} > 0$, we complete the proof. □

The next lemma states that, near a nonstationary feasible point, every point obtained with a sufficiently small radius will be accepted by the trust-region criterion.

Lemma 3.7 *Consider a nonstationary feasible point* $\overline{x} \in X$ *and the neighborhood* V_1 *given by Lemma 3.5. Then there exist constants* $\overline{\Delta} \in (0, \Delta_{\min})$ *and* $C_3 > 0$ *such that for any* $z^k \in V_1$ *and* $j \notin \mathscr{J}_k$*,*

$$pred_{k_j} > C_3 \ min\left\{\Delta_{k_j}, \overline{\Delta}\right\} \tag{17}$$

and

$$ared_{k_j} > \eta \ pred_{k_j} > \eta C_3 \Delta_{k_j}, \qquad for \ \ \Delta_{k_j} \in (0, \overline{\Delta}]. \tag{18}$$

Proof. Consider $z^k \in V_1$ and $j \notin \mathscr{J}_k$. So, by Corollary 3.6, $\pi_{k_j}(z^k) > C_2$. Using this and Assumption A4 in (9), it follows that

$$\text{pred}_{k_j} > \xi C_2 \text{min}\left\{\frac{C_2}{1+\gamma}, \Delta_{k_j}\right\}.$$

Thus, setting $\widetilde{\Delta} = \min\left\{\frac{\Delta_{\min}}{2}, \frac{C_2}{1+\gamma}\right\}$ and $C_3 = \xi C_2 > 0$ we have

$$\text{pred}_{k_j} > C_3 \ \min\left\{\Delta_{k_j}, \widetilde{\Delta}\right\}. \tag{19}$$

On the other hand, by Assumption A1 and the mean value theorem there exist $\theta \in (0,1)$ such that

$$f(z^k) - f(z^k + d^{k_j}) = -\nabla f(z^k + \theta d^{k_j})^T d^{k_j}.$$

Using the definitions of the reductions ared and pred and the above result, we have

$$\begin{aligned} \left|\text{ared}_{k_j} - \text{pred}_{k_j}\right| &= \left|-\nabla f(z^k + \theta d^{k_j})^T d^{k_j} + {g_{k_j}}^T d^{k_j} + \frac{1}{2} {d^{k_j}}^T B_{k_j} d^{k_j}\right| \\ &= \left|\left(\nabla f(z^k) - \nabla f(z^k + \theta d^{k_j}) + g_{k_j} - \nabla f(z^k) + \frac{1}{2} B_{k_j} d^{k_j}\right)^T d^{k_j}\right|. \end{aligned}$$

Using this, the triangle and the Cauchy-Schwarz inequalities and the Assumption A3 we obtain

$$\begin{aligned} \left|\text{ared}_{k_j} - \text{pred}_{k_j}\right| &\le \left(\left\|\nabla f(z^k + \theta d^{k_j}) - \nabla f(z^k)\right\| + \left\|\nabla f(z^k) - g_{k_j}\right\|\right) \left\|d^{k_j}\right\| + \\ &\quad + \frac{1}{2}\left\|B_{k_j}\right\| \left\|d^{k_j}\right\|^2 \\ &\le \left(L\theta \left\|d^{k_j}\right\| + \left\|\nabla f(z^k) - g_{k_j}\right\| + \frac{1}{2}\left\|B_{k_j}\right\| \left\|d^{k_j}\right\|\right) \left\|d^{k_j}\right\|. \end{aligned}$$

Therefore, by Assumption A4 and the facts that $\left\|d^{k_j}\right\| \le \Delta_{k_j}$ and $\Delta_{k_j} \ge \delta_{k_j}$

$$\left|\text{ared}_{k_j} - \text{pred}_{k_j}\right| \le \left(L\theta + \sigma + \frac{1}{2}\gamma\right) \Delta_{k_j}^2. \tag{20}$$

Define $\overline{\Delta} \in (0, \Delta_{\min})$ as

$$\overline{\Delta} = \min\left\{\widetilde{\Delta}, \frac{C_2(1-\eta)}{(L\theta + \sigma + \frac{1}{2}\gamma)}\right\} > 0.$$

Using the fact that $\overline{\Delta} \le \widetilde{\Delta}$ in (19), we get (17).

Now consider $j \notin \mathscr{J}_k$ with $\Delta_{k_j} \in \left(0, \overline{\Delta}\right]$. Thus, by (17), (20) and the definition of $\overline{\Delta}$, we have

$$\left|\frac{\text{ared}_{k_j}}{\text{pred}_{k_j}} - 1\right| = \left|\frac{\text{ared}_{k_j} - \text{pred}_{k_j}}{\text{pred}_{k_j}}\right| < \frac{(L\theta + \sigma + \frac{1}{2}\gamma)}{C_2} \Delta_{k_j} \le 1 - \eta.$$

In this way, for any $z^k \in V_1$ and $\Delta_{k_j} \in \left(0, \overline{\Delta}\right]$, we obtain that $\mathrm{ared}_{k_j} > \eta \mathrm{pred}_{k_j}$. Using this and (17) we get the second inequality of (18), which concludes the proof. □

The next lemma proves that, near a nonstationary feasible point, the refusal of an optimality step is due to a large increase of the infeasibility. For that, note that by A1, A2 and the mean value theorem there exists a constant $\widehat{L} > 0$ such that $\left|f(z^k) - f(x^k)\right| \leq \widehat{L}\left\|z^k - x^k\right\|$. Using this and (5), we have that there exists a constant $C_4 > 0$ such that

$$\left|f(z^k) - f(x^k)\right| \leq C_4 h(x^k). \tag{21}$$

Lemma 3.8 *Let $\overline{x} \in X$ be a nonstationary feasible point. Consider the constants given by Lemmas 3.3 and 3.7 and Corollary 3.6 and the neighborhood V_1 given by Lemma 3.5. Set $C_5 = \frac{1}{2}\min\left\{\frac{1}{\sqrt{C_1}}, \mu C_2\right\} > 0$ and $\Delta' \in (0, \Delta_{\min})$ given by*

$$\Delta' = \min\left\{\widehat{\delta}, \overline{\Delta}, \frac{\eta C_3}{8\alpha C_1}, \eta C_3\left(\frac{1-\alpha}{(C_5)^2} + C_1\alpha + \frac{C_4}{\alpha (C_5)^2}\right)^{-1}\right\}. \tag{22}$$

Then there exists a neighborhood $V_2 \subset V_1$ of $\overline{x}$ such that for any $x^k \in V_2$ and $\Delta_{k_j} \in (\Delta'/2, \Delta']$ with $j \notin \mathscr{I}_k$ we have $z^k \in V_1$ and

$$f(z^k + d^{k_j}) + \alpha h(x^k) < f(x^k) \qquad \text{for the flat filter} \tag{23}$$

and

$$f(z^k + d^{k_j}) + \alpha h(z^k + d^{k_j}) < f(x^k) \qquad \text{for the slanting filter.} \tag{24}$$

Moreover, if d^{k_j} is refused by the algorithm, then

$$h(z^k + d^{k_j}) \geq H_k.$$

Proof. Using the second inequality given in (5), we have

$$\|z^k - \overline{x}\| \leq \|z^k - x^k\| + \|x^k - \overline{x}\| \leq \beta h(x^k) + \|x^k - \overline{x}\|.$$

Since $\overline{x}$ is feasible and h is a continuous function, there exists a neighborhood $V_2 \subset V_1$ of $\overline{x}$ such that if $x^k \in V_2$, then $z^k \in V_1$ and $h(x^k)$ is sufficiently small, i.e.,

$$h(x^k) \leq \frac{\eta C_3 \Delta'}{2}\min\left\{\frac{1}{\alpha + C_4}, \frac{1}{2C_4}, \frac{1}{4\alpha(1-\alpha)}\right\}. \tag{25}$$

Consider $x^k \in V_2$, $z^k \in V_1$ and $j \notin \mathscr{I}_k$ with $\Delta_{k_j} \in \left(\frac{\Delta'}{2}, \Delta'\right] \subset \left(0, \overline{\Delta}\right]$. Since $\frac{\Delta'}{2} < \Delta_{k_j} \leq \overline{\Delta}$, by Lemma 3.7,

$$f(z^k) - f(z^k + d_{k_j}) > \eta C_3 \Delta_{k_j} > \frac{\eta C_3 \Delta'}{2}.$$

Thus, using this and (21) we obtain that

$$f(x^k) - f(z^k + d_{k_j}) = f(x^k) - f(z^k) + f(z^k) - f(z^k + d_{k_j}) > -C_4 h(x^k) + \frac{\eta C_3 \Delta'}{2}. \tag{26}$$

Using (25) it follows that

$$f(x^k) - f(z^k + d_{k_j}) > -C_4 h(x^k) + (\alpha + C_4) h(x^k) = \alpha h(x^k)$$

which implies the result (23) to the flat filter.

By (25) and (26) we have

$$f(x^k) - f(z^k + d_{k_j}) > -C_4 \frac{\eta C_3 \Delta'}{4 C_4} + \frac{\eta C_3 \Delta'}{2} = \frac{\eta C_3 \Delta'}{4}. \tag{27}$$

On the other hand, by Lemma 3.3 we have $|h(z^k+d_{k_j})-h(z^k)| \le C_1\Delta_{k_j}^2$ and by the mechanism of Algorithm 1, $h(z^k) < (1-\alpha)h(x^k)$. Therefore, by these and the fact that $\Delta_{k_j} \le \Delta'$, we have

$$h(z^k+d_{k_j}) \le h(z^k) + C_1\Delta_{k_j}^2 < (1-\alpha)h(x^k) + C_1(\Delta')^2.$$

Multiplying by $\alpha > 0$ and using (22) and (25), we obtain

$$\alpha h(z^k+d_{k_j}) < \alpha(1-\alpha)\frac{\eta C_3\Delta'}{8\alpha(1-\alpha)} + C_1\alpha(\Delta')^2 = \frac{\eta C_3\Delta'}{8} + C_1\alpha(\Delta')^2 < \frac{\eta C_3}{4}\Delta'.$$

Therefore, combining this with (27), we have

$$f(x^k) - f(z^k+d_{k_j}) > \alpha h(z^k+d_{k_j})$$

which proves (24) to the slanting filter.

To complete the proof, suppose that the point $z^k+d_{k_j}$ was refused by the algorithm. Since $\Delta_{k_j} \le \overline{\Delta}$ and $j \notin \mathscr{J}_k$, the Lemma 3.7 ensures that this point was accepted by the trust-region criterion. Thus, it was refused by the filter criterion, i.e.,

$$z^k + d_{k_j} \in \widehat{\mathcal{F}}_k.$$

Therefore, since (23) or (24) holds by the flat or the slanting filter, respectively, we have, by the definitions of filter and of H_k, that $h(z^k+d^{k_j}) \ge H_k$ and the proof is complete. □

The next result presents that, near a nonstationary feasible point the optimization phase provides a sufficient decrease in the objective function.

Theorem 3.9 *Given a nonstationary feasible point $\overline{x} \in X$, there exist constants $C_6, C_7 > 0$ and a neighborhood V_3 of $\overline{x}$ such that for any iterate $x^k \in V_3$, the point x^{k+1} obtained by Algorithm 1 satisfies*

$$f(z^k) - f(x^{k+1}) \ge C_6\sqrt{H_k} \tag{28}$$

and

$$f(z^k) - f(x^{k+1}) \ge C_7\left\|z^k - x^{k+1}\right\|. \tag{29}$$

Proof. As the sequence (x^k) is infinite, by the mechanism of the algorithm

$$f(z^k) - f(x^{k+1}) = \text{ared}_k > \eta\text{pred}_k, \tag{30}$$

where $\text{ared}_k = \text{ared}_{k_j}$ and $\text{pred}_k = \text{pred}_{k_j}$ with $j \in \mathbb{N}$ such that $d^k = d^{k_j}$. By Lemma 3.3 we have that for any $\Delta_{k_j} > 0$ there exists a constant $C_1 > 0$ such that

$$\left|h(z^k+d^{k_j}) - h(z^k)\right| \le C_1\Delta_{k_j}^2. \tag{31}$$

Consider $\overline{\Delta} > 0$ given by Lemma 3.7, $\Delta' \le \overline{\Delta}$ and V_2 given by Lemma 3.8. Let $V_3 \subset V_2$ be the neighborhood of $\overline{x}$ such that $h(x^k) < 1$ and $z^k \in V_2$, for any $x^k \in V_3$. This neighborhood is well defined since $h(\overline{x}) = 0$, $h(.) \ge 0$ is a continuous function and (5) holds.

Consider $x^k \in V_3$ and consequently $z^k \in V_2$. By the feasibility phase, the definition of H_k and the fact that $h(x^k) < 1$, we have

$$h(z^k) < (1-\alpha)h(x^k) < (1-\alpha)H_k. \tag{32}$$

Denote j^* the index of the successful internal iteration, i.e., $\Delta_k = \Delta_{k_{j^*}}$. Note that by the mechanism of the algorithm, $j^* \notin \mathscr{J}_k$. We shall consider some cases as summarized on Table 1.

First case: suppose that $\Delta_k > \frac{\Delta'}{2}$. By the fact that $j^* \notin \mathscr{J}_k$, Lemma 3.7 and the fact $\frac{\Delta'}{2} < \min\{\Delta_k, \overline{\Delta}\}$ we have that $\text{pred}_k > C_3\frac{\Delta'}{2}$. Applying this in (30), we obtain

$$f(z^k) - f(x^{k+1}) > \frac{1}{2}\eta C_3\Delta'. \tag{33}$$

<table>
<tr><td>First case:
$\Delta_k > \frac{\Delta'}{2}$</td><td colspan="3">Second case:
$\Delta_k \leq \frac{\Delta'}{2}$</td></tr>
<tr><td></td><td>$h(z^k + d_{k_j}) \geq H_k$,
$\forall\ \Delta_{k_j} \leq (\Delta'/2)$ with $j \notin \mathscr{J}_k$</td><td colspan="2">$\exists j \notin \mathscr{J}_k$ with $\Delta_{k_j} \leq (\Delta'/2)$
such that $h(z^k + d_{k_j}) < H_k$</td></tr>
<tr><td></td><td></td><td>$\widehat{j} - 1 \notin \mathscr{J}_k$</td><td>$\widehat{j} - 1 \in \mathscr{J}_k$</td></tr>
</table>

Table 1: Cases considered on the proof of Theorem 3.9.

The sequences $\left(\sqrt{H_k}\right)_{k\in\mathbb{N}}$ and $\left(\left\|z^k - x^{k+1}\right\|\right)_{k\in\mathbb{N}}$ are bounded, because for any $k \in \mathbb{N}$ we have $0 < H_k \leq 1$ and $(\Delta_k)_{k\in\mathbb{N}}$ is a positive nonincreasing sequence. Thus, there exist constants $C_6, C_7 > 0$ such that $\frac{1}{2}\eta C_3 \Delta' \geq C_6\sqrt{H_k}$ and $\frac{1}{2}\eta C_3\Delta' \geq C_7 \left\|z^k - x^{k+1}\right\|$. Applying this in (33), we get (28) and (29).

Second case: suppose that

$$\Delta_k \leq \frac{\Delta'}{2}. \tag{34}$$

Since $\overline{x}$ is nonstationary we can restrict V_3, if necessary, such that for any $z^k \in V_3$, we have that z^k is nonstationary. By Lemma 3.2 and the mechanism of the algorithm, there exists at least one $j \in \mathbb{N}$ such that $j \notin \mathscr{J}_k$. Moreover $x^{k+1} = z^k + d_{k_{j^*}}$ with $j^* \notin \mathscr{J}_k$ and the trust-region steps are computed only in internal iterations in which $j \notin \mathscr{J}_k$. Thus it is enough to ensure the result for $j \notin \mathscr{J}_k$. Let us look again at two situations.

- Suppose that the condition

$$h(z^k + d^{k_j}) \geq H_k, \tag{35}$$

holds for any $\Delta_{k_j} \leq \dfrac{\Delta'}{2}$ with $j \notin \mathscr{J}_k$. Thus, by (34) we have that (35) holds in particular for j^*. Therefore, by (32) and (35) we have

$$h(z^k + d^k) - h(z^k) > \alpha H_k.$$

Using this and (31), we obtain

$$\alpha H_k < h(z^k + d^k) - h(z^k) \leq C_1 \Delta_k^2. \tag{36}$$

From (30), Lemma 3.7 and the fact that $\Delta_k \leq \dfrac{\Delta'}{2} < \overline{\Delta}$, it follows that

$$f(z^k) - f(x^{k+1}) > \eta C_3 \Delta_k. \tag{37}$$

Using this and (36) we have

$$f(z^k) - f(x^{k+1}) > \eta C_3 \sqrt{\frac{\alpha}{C_1}}\sqrt{H_k}.$$

Thus, setting $C_6 = \eta C_3 \sqrt{\dfrac{\alpha}{C_1}} > 0$, we obtain (28). On the other hand, using the fact that $\left\|z^k - x^{k+1}\right\| \leq \Delta_k$ in (37) and considering $C_7 = \eta C_3 > 0$, we get (29).

- Now assume that there exists some $j \notin \mathscr{J}_k$ with $\Delta_{k_j} \leq \dfrac{\Delta'}{2}$ such that (35) does not hold. Consider $\widehat{j} \notin \mathscr{J}_k$ the first index with $\Delta_{k_j} \leq \dfrac{\Delta'}{2}$ such that (35) fails. Denote $\widehat{\Delta} = \Delta_{k_{\widehat{j}}}$ and $\widehat{x} = z^k + d^{k_{\widehat{j}}}$.

First let us bound the radius $\widehat{\Delta}$ with respect to the filter slack H_k. Note that since $2\widehat{\Delta} \leq \Delta' \leq \overline{\Delta} < \Delta_{\min}$ and $\Delta_{k_0} > \Delta_{\min}$, we have that the radius $2\widehat{\Delta}$ of the internal iteration $\widehat{j}-1$ was refused by the algorithm.

∘ Suppose that $\widehat{j}-1 \notin \mathscr{J}_k$. In this case, $\Delta_{k_{\widehat{j}-1}} = 2\widehat{\Delta}$. We claim that (35) holds in $2\widehat{\Delta}$. In fact, if $2\widehat{\Delta} \leq \frac{\Delta'}{2}$, the statement follows immediately by the definition of $\widehat{\Delta}$ and the fact that $\widehat{j}-1 \notin \mathscr{J}_k$. On the other hand, if $2\widehat{\Delta} \in (\frac{\Delta'}{2}, \Delta']$, the claim is due to Lemma 3.8. Thus, by (31), (32) and (35), we have

$$4C_1\widehat{\Delta}^2 \geq h(z^k + d^{k_{\widehat{j}-1}}) - h(z^k) > H_k - (1-\alpha)H_k = \alpha H_k.$$

Therefore

$$\widehat{\Delta} > \frac{1}{2\sqrt{C_1}}\sqrt{\alpha H_k}. \tag{38}$$

∘ Suppose now that $\widehat{j}-1 \in \mathscr{J}_k$. Thus, by the definitions of $\widehat{\Delta}$ and $\mathscr{J}_k$ and the fact that $\sqrt{H_k} \leq 1$ and $\alpha \in (0,1)$,

$$2\widehat{\Delta} = \Delta_{k_{\widehat{j}-1}} \geq \delta_{k_{\widehat{j}-1}} > \mu\pi_{k_{\widehat{j}-1}}(z^k) > \mu\sqrt{\alpha H_k}\pi_{k_{\widehat{j}-1}}(z^k).$$

Using this and the fact that $\delta_{k_{\widehat{j}-1}} \leq 2\widehat{\Delta} \leq \Delta' \leq \widehat{\delta}$ we have by Corollary 3.6 that

$$\widehat{\Delta} > \frac{\mu C_2}{2}\sqrt{\alpha H_k}. \tag{39}$$

Consider $C_5 > 0$ the constant given by Lemma 3.8. Combining the definition of C_5 with (38) and (39), we have in both cases $\widehat{j}-1 \notin \mathscr{J}_k$ and $\widehat{j}-1 \in \mathscr{J}_k$, that

$$\widehat{\Delta} > C_5\sqrt{\alpha H_k}. \tag{40}$$

On the other hand, since $\widehat{j} \notin \mathscr{J}_k$, by Lemma 3.7 and the fact that $\widehat{\Delta} < \overline{\Delta}$, it results in

$$f(z^k) - f(\widehat{x}) > \eta C_3\widehat{\Delta}. \tag{41}$$

Using (40) in (41) and considering $C_6 = \eta C_3 C_5\sqrt{\alpha} > 0$, it follows that

$$f(z^k) - f(\widehat{x}) > C_6\sqrt{H_k}. \tag{42}$$

Combining $\left\|z^k - \widehat{x}\right\| \leq \widehat{\Delta}$ with (41) and taking $C_7 = \eta C_3 > 0$ we obtain

$$f(z^k) - f(\widehat{x}) > C_7\left\|z^k - \widehat{x}\right\|.$$

Therefore, (28) and (29) hold at $\widehat{x}$. To finish the proof it is sufficient to verify that $x^{k+1} = \widehat{x}$. Since $\widehat{j} \notin \mathscr{J}_k$ and $\widehat{\Delta} < \overline{\Delta}$, the Lemma 3.7 yields $\text{ared}_{k_{\widehat{j}}} > \eta\,\text{pred}_{k_{\widehat{j}}}$. Thus, the point $\widehat{x}$ satisfies the trust-region criterion. Moreover, since (35) does not hold at $\widehat{x}$, for x^{k+1} to be equal to $\widehat{x}$ it is enough to check that

$$f(\widehat{x}) < f(x^k) - \alpha h(x^k) \tag{43}$$

for the flat filter and

$$f(\widehat{x}) < f(x^k) - \alpha h(\widehat{x})$$

for the slanting filter, because then $\widehat{x} \notin \widehat{\mathcal{F}}_k$. Let us first ensure the result to the flat filter. By (21) and (42), we have

$$f(\widehat{x}) < f(z^k) - C_6\sqrt{H_k} \leq f(x^k) + C_4 h(x^k) - C_6\sqrt{H_k}. \tag{44}$$

If $h(x^k) = 0$,

$$f(\widehat{x}) < f(x^k) - C_6\sqrt{H_k} < f(x^k) = f(x^k) - \alpha h(x^k)$$

and in this case $x^{k+1} = \widehat{x}$. Suppose now that $h(x^k) > 0$. Since $h(x^k) < H_k$, because $h(x^k) < 1$, we have by (44) that

$$f(\widehat{x}) < f(x^k) + C_4 h(x^k) - C_6\sqrt{h(x^k)} = f(x^k) + \left(C_4 - \frac{C_6}{\sqrt{h(x^k)}}\right) h(x^k). \tag{45}$$

Since h is a continuous function and $\overline{x}$ is a feasible point, we can restrict V_3, if necessary, such that for any $x^k \in V_3$, we have

$$\sqrt{h(x^k)} < \frac{C_6}{C_4 + \alpha}.$$

Thus, combining this with (45) we have (43) and $x^{k+1} = \widehat{x}$ as desired.

Let us verify that $x^{k+1} = \widehat{x}$ when the slanting filter is considered. By (31) and (32),

$$h(\widehat{x}) \leq h(z^k) + C_1\widehat{\Delta}^2 < (1-\alpha)H_k + C_1\widehat{\Delta}^2.$$

Using (40) we obtain

$$h(\widehat{x}) < \left(\frac{(1-\alpha)}{\alpha(C_5)^2} + C_1\right)\widehat{\Delta}^2. \tag{46}$$

On the other hand, combining the fact that $H_k \geq h(x^k)$ with (21) and (40) it follows that

$$\left|f(z^k) - f(x^k)\right| \leq C_4 H_k < \frac{C_4}{\alpha(C_5)^2}\widehat{\Delta}^2. \tag{47}$$

Thus, (41), (47) and the fact that $\widehat{\Delta} < \Delta'$ yield

$$\begin{aligned} f(x^k) - f(\widehat{x}) &= f(x^k) - f(z^k) + f(z^k) - f(\widehat{x}) \\ &> -\frac{C_4}{\alpha(C_5)^2}\widehat{\Delta}^2 + \eta C_3\widehat{\Delta} > \left(-\frac{C_4}{\alpha(C_5)^2} + \frac{\eta C_3}{\Delta'}\right)\widehat{\Delta}^2. \end{aligned}$$

Finally, using the definition of Δ' given by Lemma 3.8 and (46), it results that

$$f(x^k) - f(\widehat{x}) > \left(\frac{(1-\alpha)}{(C_5)^2} + C_1\alpha\right)\widehat{\Delta}^2 > \alpha h(\widehat{x})$$

and the proof is complete. □

The efficiency condition of the optimality phase proved in the last theorem is extended to the full step as shows the next result, fulfilling our task.

Lemma 3.10 *Consider a nonstationary feasible point $\overline{x} \in X$. Then, there exist a constant $C_8 > 0$ and a neighborhood V_4 of $\overline{x}$ such that for any iterate $x^k \in V_4$,*

$$f(x^k) - f(x^{k+1}) \geq C_8 v_k.$$

Proof. Consider the neighborhood V_3 and the constant C_6 given by Theorem 3.9, and C_4 given in (21). As $\overline{x}$ is feasible and, by A1, the function h is continuous, there exists a neighborhood $V_4 \subset V_3$ of $\overline{x}$ such that, for all $x^k \in V_4$, $h(x^k) \leq \min\left\{1, (C_6/(2C_4))^2\right\}$. Consider $x^k \in V_4$. Using (21), Theorem 3.9 and the fact that $h(x^k) \leq H_k$, we have

$$\begin{aligned} f(x^k) - f(x^{k+1}) &= f(x^k) - f(z^k) + f(z^k) - f(x^{k+1}) \\ &\geq (-C_4\sqrt{h(x^k)} + C_6)\sqrt{H_k} \\ &\geq \frac{C_6}{2}\sqrt{H_k} \end{aligned}$$

As $H_k \leq 1$, we have $\sqrt{H_k} \geq H_k \geq v_k$. Using this and taking $C_8 = C_6/2 > 0$, we complete the proof. □

Having the efficiency condition on the step, the global convergence of the Algorithm 1 is inherited from [59] as shown below.

Theorem 3.11 *Let (x^k) be the infinite sequence generated by the Algorithm 1. Assume that the Assumptions A1-A5 hold. Then the sequence (x^k) has a stationary accumulation point.*

Proof. Our algorithm fits in the general framework of [59, Algorithm 1]. By Lemma 3.10, the sequence (x^k) satisfies the efficiency condition on the step which coincides with the Hypothesis H3 of [59]. Consequently the global convergence result is inherited from [59, Theorem 3.5], i.e., the sequence (x^k) has a stationary accumulation point. □

The last theorem ensures that the sequence (x^k) generated by Algorithm 1 has a stationary accumulation point, independent of the filter rule adopted. However, from [40], stronger results are obtained when the slanting filter is used, as presented in the next theorem.

Theorem 3.12 *Let (x^k) be the infinite sequence generated by the Algorithm 1 with $\mathcal{R}_k = \grave{\mathcal{R}}_k$ as defined in* (4). *Assume that the Assumptions A1-A5 hold. Then any accumulation point of the sequence (x^k) is stationary.*

Proof. Our algorithm fits in the general framework of [40, Algorithm 2.1] which uses the slanting filter. By Theorem 3.9, the sequence (x^k) satisfies the Hypothesis H5 of [40]. Consequently the global convergence result is inherited from [40, Theorem 3.7], i.e., any accumulation point of the sequence (x^k) is stationary. □

4 Implementation and numerical experiments

Inexact restoration algorithms allow the use of different algorithms in the feasibility and optimality phases. Algorithm 1 was implemented with the ability of using different algorithms for: feasibility phase, building and updating the trust-region model, and solving subproblem (8).

At iteration k, as explained in [8], to find z^k in the feasibility phase we solve

$$\begin{array}{ll} \text{minimize} & \|z - x^k\|_2^2 \\ \text{subject to} & z \in \Omega \end{array} \tag{48}$$

only when $h(x^k) > 0$. Since the derivatives of the constraints are available, problem (48) is solved with Augmented Lagrangian algorithm `ALGENCAN` [1, 2]. Moreover, to ensure conditions (5), we set $\beta = 10^3$ and $\alpha = 10^{-1}$ and, to ensure $z^k \notin \widehat{\mathcal{F}}_k$, we set `ALGENCAN`'s feasibility tolerance to a value smaller than $h(x^k)$. If the restored point does not satisfy (5) or belongs to $\widehat{\mathcal{F}}_k$, then the feasibility tolerance is decreased. This procedure will certainly stop, since $h(x^k) > 0$ and $\Omega \neq \emptyset$.

The quadratic models required at optimality phase were obtained by polynomial interpolation. The construction and updating of the interpolation sets were based on the ideas proposed in [12]. The number of interpolation points was fixed to 5 if $n = 2$ and $2n + 3$ otherwise. At inner iterations of the optimality phase, these sets were constructed from scratch or updated from the previous iterations. Between outer iterations, the algorithm tries to reuse the last interpolation set of iteration $k - 1$ to construct the first model at iteration k.

To solve subproblem (8) we use `ALGENCAN` with default parameters. According to [2], a solution d^{k_j} found by `ALGENCAN` satisfies the KKT conditions of subproblem (8). Since the feasible set of (8) is convex, denoting by x^+ the point $z^k + d^{k_j}$, it is not hard to show (see [7, Chapter 9] and [6, Chapter 2]) that

$$x^+ = \mathcal{P}_{\mathcal{L}(z^k)}(x^+ - \nabla m_{k_j}(x^+)). \tag{49}$$

Condition (49) gives us a reasonable way of estimating $\pi_{k_j}(z^k)$. By (49), the definitions of x^+ and m_{k_j} and the contraction property of projections

$$\|x^+ - z^k\| = \|\mathcal{P}_{\mathcal{L}(z^k)}(x^+ - \nabla m_{k_j}(x^+)) - z^k\| \geq -\|B_{k_j}(x^+ - z^k)\| - \|x^+ - z^k\| + \pi_{k_j}(z^k).$$

Using this and Assumption A4, we have

$$\|d^{k_j}\| = \|x^+ - z^k\| \geq \frac{1}{2+\gamma}\pi_{k_j}(z^k),$$

which implies that $\pi_{k_j}(z^k)$ can be estimated by $\|d^{k_j}\|$ in the condition of the first `If` of the optimality phase. In practice, according to [12], this condition is replaced by

$$\delta_{k_j} > 2\|d^{k_j}\|$$

and the parameter μ is implicitly defined due to its dependency of γ. Note that subproblem (8) needs to be solved before this condition is tested. The algorithm declares that a solution of (1) was found if $z^k + d^{k_j} \notin \widehat{\mathcal{F}}_k$ is feasible and $\|d^{k_j}\| \leq \varepsilon$.

From now on Algorithm 1 will be referred to as `FIRD` (*Filter Inexact Restoration Derivative-free algorithm*). We considered two variants of the algorithm. The first one, called `F-FIRD`, uses the flat filter criterion given by (3). The second variant, called `S-FIRD`, considers the slanting filter criterion given by (4).

To put our approach in perspective, `S-FIRD` and `F-FIRD` were compared with two derivative-free algorithms: the inexact restoration algorithm [8] and algorithm `DFO` [16]. The inexact restoration algorithm proposed in [8], here denoted by `IR`, controls the progress of the algorithm by a merit function, instead of filters. `DFO` is a well known derivative-free trust-region method which uses Newton polynomials to build the models. Both algorithms are able to use information of the derivatives of the constraints. For solving the feasibility suproblems needed by `IR` and the trust-region problems needed by `DFO`, it was used `ALGENCAN`. For solving the optimality subproblems needed by `IR`, the algorithm `GSS` was used (see [8]).

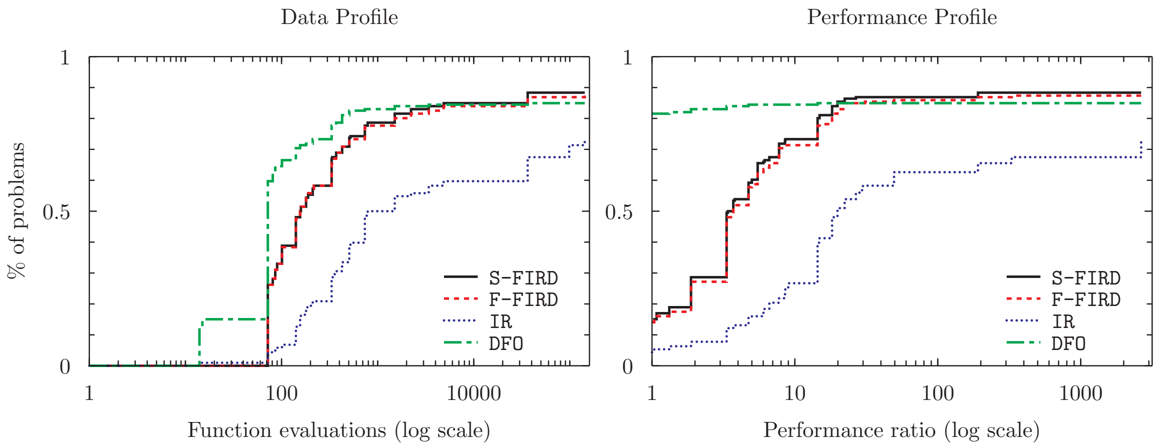
Algorithms `S-FIRD` and `F-FIRD` were implemented in Fortran 90. The numerical tests were performed on a 64-bit Intel Xeon E3-1220 v3, with 3.10 GHz of CPU and 16GB RAM, using LUbuntu operating system. The code was compiled with `gfortran` version 4.8.4. The supremum norm was always used in the implementation of `FIRD`. The feasibility tolerance used for all algorithms was 10^{-8}. The optimality tolerance for `FIRD` and `DFO` was $\varepsilon = 10^{-4}$. As suggested in [8], the optimality tolerance of `IR` was 10^{-3}. For each problem, up to 10 minutes of CPU time were allowed. The remaining parameters of `FIRD` were defined by $\eta = 0.1$ and $\Delta_{\min} = 10^{-30}$. `ALGENCAN` 3.0.0 was used for solving all nonlinear programming subproblems from the tested algorithms. In the optimality phase of `IR`, the `GSS` algorithm implemented in `HOPSPACK` 2.0 [60] was used.

The set of test problems consisted of all 206 problems from the Hock-Schittkowski collection [69] that involve at least one constraint besides box constraints for which the derivatives are available. The initial point was always the default of the collection. As suggested in [8], a problem was considered solved by an algorithm if the obtained solution $\bar{x}$ was such that

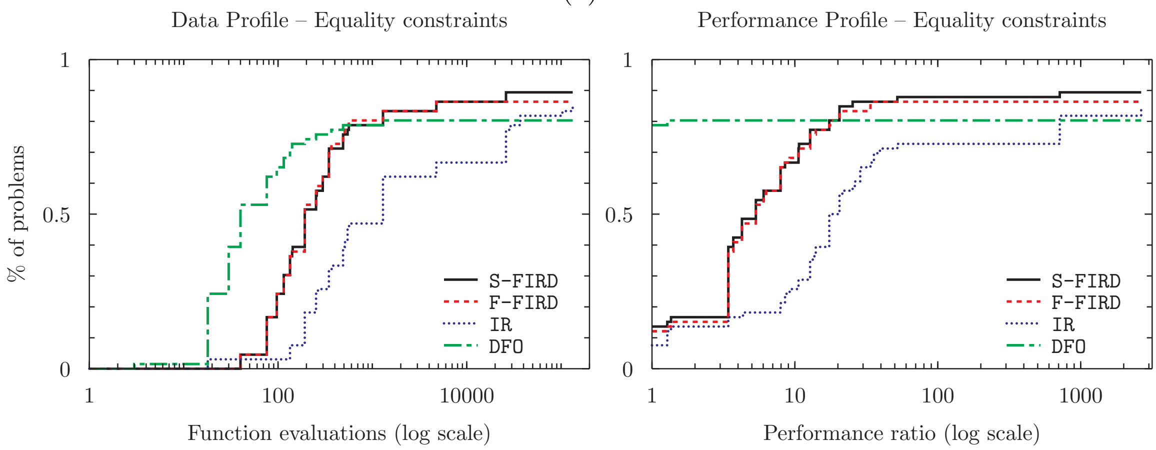
$$h(\bar{x}) \leq 10^{-8} \quad \text{and} \quad \frac{f_{\min} - f(\bar{x})}{\max\{1, f_{\min}, f(\bar{x})\}} \leq 10^{-4},$$

where $f_{\min}$ is the smallest function value found among all the strategies under comparison. Figure 4 displays the data and performance profiles [22, 55] considering the number of function evaluations.

The two profiles of Figure 4(a) are related to all 206 problems. `S-FIRD` solved 88% of the problems while `F-FIRD` solved 87%, followed by `DFO` and `IR` with 85% and 73%, respectivelly. `DFO` was the most efficient algorithm in 82% of the problems. This behavior has already been observed in the literature [54, 12]. `S-FIRD` and `F-FIRD` were the most efficient in 13% of the problems. It can be noted that the convergence properties of `FIRD` resulted in a more robust method. The direct search procedure used in `IR`'s optimality phase can explain its poor performance. Although a trust-region procedure could be used for solving the subproblems of this phase, `IR` would lose its theoretical results. Data profile shows that if an amount of approximately 5000 function evaluations is allowed, then `FIRD` and `DFO` solve the same number of problems.



(a)



(b)

Figure 4: Data and performance profiles for the compared algorithms. (a) Profiles related to the 206 constrainted problems from the Hock-Schittkowski collection. (b) Profiles related to the 66 equality constrained problems.

The two profiles in Figure 4(b) consider the subset of all 66 equality constrained problems. In this scenario, `S-FIRD`, `F-FIRD` and `IR` are the most robust algorithms, solving 59, 57 and 56 problems, respectivelly, while `DFO` solved 53. Still, `DFO` was the most efficient algorithm. The increase in the percentual difference on the robustness between inexact restoration approaches and `DFO` can be explained by the fact that inexact restoration algorithms allow infeasible points, as long as they are "good" choices. On the other hand, `DFO` projects infeasible points to build the quadratic models.

The code of `FIRD` is available to download at `https://github.com/fsobral/fird`. The complete results of the numerical experiments are available at `https://docs.ufpr.br/~ewkaras/pesquisa/publicacoes/supplemental_FKSS`.

5 Conclusions

In this work we have proposed an inexact restoration filter algorithm for nonlinear programming problems, in which the objective function derivatives are not explicitly used. Each iteration consists of two phases: a restoration phase for reducing an infeasibility measure, and an optimality phase for improving the objective function value in the linearization of the feasible direction set. These two phases are independent and the coupling between them is made by a filter, which can either be the flat filter [30] or the slanting [10] one. As the derivatives of the objective function are not available, it is in the optimality phase that the derivative-free trust-region techniques are used. The linear or quadratic trust-region models can be constructed by any technique as long as they approximate sufficiently well the objective function at the current point. We have showed that the obtained steps provides an efficiency condition using both flat and slanting filter rules. For the flat filter, we have proved that the algorithm generates a stationary accumulation point. A stronger result has been proved for the slanting filter, namely, all accumulation points are stationary. A set of numerical experiments was prepared to illustrate the practical performance of the proposed algorithm. The implementation was shown to be more robust than other methods that are capable of using derivatives of the constraints.

Acknowledgements

The authors would like to thank Sandra A. Santos for her valuable comments.

Funding

This work was partially supported by CNPq grants 477611/2013-3, 308957/2014-8, 442158/2014-9 and 458360/2014-7.

References

[1] R. Andreani, E.G. Birgin, and J.M. Martínez. On augmented Lagrangian methods with general lower-level constraints. *SIAM J. Optimiz.*, 18:1286–1309, 2007.

[2] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Math. Program.*, 111:5–32, 2008.

[3] M. B. Arouxét, N. Echebest, and E. A. Pilotta. Inexact restoration method for nonlinear optimization without derivatives. *Comput. Appl. Math.*, 290:26–43, 2015.

[4] C. Audet and J. E. Dennis Jr. A pattern search filter method for nonlinear programming without derivatives. *SIAM J. Optimiz.*, 14(4):980–1010, 2004.

[5] C. Audet and J. E. Dennis Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM J. Optimiz.*, 17:188–217, 2006.

[6] M. Bazaraa, H. Sherali, and C. Shetty. *Nonlinear Programming Theory and Algorithms.* Athena Scientific, Wiley, New York, 2rd edition, 2006.

[7] E. G. Birgin and J. M. Martínez. *Practical augmented Lagrangian Methods for constrained optimization.* SIAM Series on Fundamentals of Algorithms, SIAM, Brazil, 2014.

[8] L. F. Bueno, A. Friedlander, J. M. Martínez, and F. N. C. Sobral. Inexact restoration method for derivative-free optimization with smooth constraints. *SIAM J. Optimiz.*, 23(2):1189–1213, 2013.

[9] C. M. Chin. A local convergence theory of a filter line search method for nonlinear programming. Technical report, Numerical Optimization Report, Department of Statistics, University of Oxford, 2003.

[10] C. M. Chin and R. Fletcher. On the global convergence of an SLP-filter algorithm that takes EQP steps. *Math. Program.*, 96(1):161–177, 2003.

[11] P. G. Ciarlet and P. A. Raviart. General Lagrange and Hermite interpolation in $\mathbb{R}^n$ with applications to finite elements methods. *Arch. Ration. Mech. An.*, 46:324–330, 1972.

[12] P. D. Conejo, E. W. Karas, and L. G. Pedroso. A trust-region derivative-free algorithm for constrained optimization. *Optim. Method Softw.*, 30(6):1126–1145, 2015.

[13] P. D. Conejo, E. W. Karas, L. G. Pedroso, A. A. Ribeiro, and M. Sachine. Global convergence of trust-region algorithms for convex constrained minimization without derivatives. *Appl. Math. Comput.*, 220:324–330, 2013.

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

[15] A. R. Conn, K. Scheinberg, and Ph. L. Toint. On the convergence of derivative-free methods for unconstrained optimization. In A. Iserles and M. Buhmann, editors, *Approximation Theory and Optimization*, Tributes to M.J.D. Powell, pages 83–108. Cambridge University Press, 1997.

[16] A. R. Conn, K. Scheinberg, and Ph. L. Toint. A derivative free optimization algorithm in practice. In *Proceedings of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, St Louis, CO, USA, September 1998.

[17] A. R. Conn, K. Scheinberg, and L. N. Vicente. Geometry of interpolation sets in derivative free optimization. *Math. Program.*, 111:141–172, 2008.

[18] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization.* MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2009.

[19] A. R. Conn and Ph. L. Toint. An algorithm using quadratic interpolation for unconstrained derivative free optimization. In G. D. Pillo and F. Gianessi, editors, *Nonlinear Optimization and Applications*, pages 27–47. Plenum, 1996.

[20] P. J. Davis. *Interpolation an approximation.* Courier Corporation, United States, 2014.

[21] M. A. Diniz-Ehrhardt, J. M. Martínez, and L. G. Pedroso. Derivative-free methods for nonlinear programming with general lower-level constraints. *Comput. Appl. Math.*, 30:19–52, 2011.

[22] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213, 2002.

[23] N. Echebest, M. L. Schuverdt, and R. P. Vignau. A derivative-free method for solving box-constrained underdetermined nonlinear systems of equations. *Appl. Math. Comput.*, 219:3198–3208, 2012.

[24] N. Echebest, M. L. Schuverdt, and R. P. Vignau. An inexact restoration derivative-free filter method for nonlinear programming. *Comput. Appl. Math.*, 2015. DOI 10.1007/s40314-015-0253-0.

[25] G. Fasano, G. Liuzzi, S. Lucidi, and F. Rinaldi. A linesearch-based derivative-free approach for nonsmooth constrained optimization. *SIAM J. Optimiz.*, 24:959–992, 2014.

[26] G. Fasano, J. L. Morales, and J. Nocedal. On the geometry phase in model-based algorithms for derivative-free optimization. *Optim. Method Softw.*, 24:145–154, 2009.

[27] P. S. Ferreira, E. W. Karas, and M. Sachine. A globally convergent trust-region algorithm for unconstrained derivative-free optimization. *Comput. Appl. Math.*, 34(3):1075–1103, 2015.

[28] A. Fischer and A. Friedlander. A new line search inexact restoration approach for nonlinear programming. *Comput. Optim. Appl.*, 46:333–346, 2010.

[29] R. Fletcher, N. I. M. Gould, S. Leyffer, Ph. L. Toint, and A. Wachter. Global convergence of trust-region SQP-filter algorithm for general nonlinear programming. *SIAM J. Optimiz.*, 13:635–659, 2002.

[30] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Math. Program.*, 91:239–269, 2002.

[31] R. Fletcher and S. Leyffer. Filter-type algorithms for solving systems of algebraic equations and inequalities. In G. di Pillo and A. Murli, editors, *Advances in Optimization and Numerical Analysis*, pages 259–278. High Performance Algorithms and Software for Nonlinear Optimization, Kluwer, 2003.

[32] C. C. Gonzaga, E. W. Karas, and M. Vanti. A globally convergent filter method for nonlinear programming. *SIAM J. Optimiz.*, 14(3):646–669, 2003.

[33] N. I. M. Gould, S. Leyffer, and Ph. L. Toint. A multidimensional filter algorithm for nonlinear equations and nonlinear least-squares. *SIAM J. Optimiz.*, 15(1):17–38, 2004.

[34] N. I. M. Gould, C. Sainvitu, and Ph. L. Toint. A filter-trust-region method for unconstrained optimization. *SIAM J. Optimiz.*, 16(2):341–357, 2006.

[35] N. I. M. Gould and Ph. L. Toint. FILTRANE, a Fortran 95 filter-trust-region package for solving least-squares and nonlinear feasibility problems. *ACM T. Math. Software*, 33(1):3–25, 2007.

[36] G. N. Grapiglia, J. Yuan, and Y. Yuan. A derivative-free trust-region algorithm for composite nonsmooth optimization. *Comput. Appl. Math.*, 35:475–499, 2016.

[37] S. Gratton, Ph. L. Toint, and A. Tröltzsch. An active set trust-region method for derivative-free nonlinear bound-constrained optimization. *Optim. Method Softw.*, 26:873–894, 2011.

[38] S. Gratton and L. N. Vicente. A merit function approach for direct search. *SIAM J. Optimiz.*, 24:1980–1998, 2014.

[39] C. Gu and D. Zhu. A secant algorithm with line search filter method for nonlinear optimization. *Appl. Math. Model.*, 35(2):879–894, 2011.

[40] E. W. Karas, A. P. Oening, and A. A. Ribeiro. Global convergence of slanting filter methods for nonlinear programming. *Appl. Math. Comput.*, 200:486–500, 2008.

[41] E. W. Karas, A. A. Ribeiro, C. Sagastizábal, and M. Solodov. A bundle-filter method for nonsmooth convex constrained optimization. *Math. Program.*, 116:297–320, 2009.

[42] T. G. Kolda, R. M. Lewis, and V. Torczon. A generating set direct search augmented Lagrangian algorithm for optimization with a combination of general and linear constraints. Technical Report 5315, Sandia National Laboratories, 2006.

[43] S. Le Digabel. Algorithm 909: NOMAD: Nonlinear optimization with the MADS algorithm. *ACM Trans. Math. Softw.*, 37(4):44:1–44:15, 2011.

[44] R. M. Lewis and V. Torczon. Pattern search algorithms for bound constrained minimization. *SIAM J. Optimiz.*, 9:1082–1099, 1999.

[45] R. M. Lewis and V. Torczon. Pattern search algorithms for linearly constrained minimization. *SIAM J. Optimiz.*, 10:917–941, 2000.

[46] R. M. Lewis and V. Torczon. A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds. *SIAM J. Optimiz.*, 12:1075–1089, 2002.

[47] R. M. Lewis and V. Torczon. A direct search approach to nonlinear programming problems using an augmented Lagrangian method with explicit treatment of linear constraints. Technical Report WM-CS-2010-01, College of William and Mary, 2010.

[48] G. Liuzzi, S. Lucidi, and M. Sciandrone. Sequential penalty derivative-free methods for nonlinear constrained optimization. *SIAM J. Optimiz.*, 20(5):2614–2635, 2010.

[49] J. Long, C. Ma, and P. Nie. A new filter method for solving nonlinear complementarity problems. *Appl. Math. Comput.*, 185(1):705–718, 2007.

[50] J. Long and S. Zeng. A new Filter-Levenberg-Marquart method with disturbance for solving nonlinear complementarity problems. *Appl. Math. Comput.*, 216(2):677–688, 2010.

[51] S. Lucidi, M. Sciandrone, and P. Tseng. Objective-derivative-free methods for constrained optimization. *Math. Program.*, 92:37–59, 2002.

[52] J. M. Martínez. Inexact-restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. *J. Optim. Theory Appl.*, 111:39–58, 2001.

[53] J. M. Martínez and E. A. Pilotta. Inexact restoration algorithm for constrained optimization. *J. Optim. Theory Appl.*, 104:135–163, 2000.

[54] J. M. Martínez and F. N. C. Sobral. Constrained derivative-free optimization on thin domains. *J. Global Optim.*, 56(3):1217–1232, 2013.

[55] J. J. Moré and S. M. Wild. Benchmarking derivative-free optimization algorithms. *SIAM J. Optimiz.*, 20(1):172–191, 2009.

[56] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer Series in Operations Research and Financial Engineering. Springer-Verlag, 2nd edition, 2006.

[57] Y. Pei and D. Zhu. A trust-region algorithm combining line search filter technique for nonlinear constrained optimization. *Int. J. Comput. Math.*, 91(8):1817–1839, 2014.

[58] Y. Peng, H. Feng, and Q. Li. A filter-variable-metric method for nonsmooth convex constrained optimization. *Appl. Math. Comput.*, 208(1):119–128, 2009.

[59] G. A. Periçaro, A. A. Ribeiro, and E. W. Karas. Global convergence of a general filter algorithm based on an efficiency condition of the step. *Appl. Math. Comput.*, 219:9581–9597, 2013.

[60] T. D. Plantenga. HOPSPACK 2.0 User Manual, 2009.

[61] M. J. D. Powell. Developments of NEWUOA for minimization without derivatives. *IMA J. Numer. Anal.*, 28:649–664, 2008.

[62] M. J. D. Powell. The BOBYQA algorithm for bound constrained optimization without derivatives. Technical Report DAMTP 2009/NA06, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, 2009.

[63] M. J. D. Powell. On derivative-free optimization with linear constraints. In *21st ISMP*, Berlin, Germany, 2012.

[64] M. J. D. Powell. On the convergence of trust region algorithms for unconstrained minimization without derivatives. *Comput. Optim. Appl.*, 53:527–555, 2012.

[65] A. A. Ribeiro, E. W. Karas, and C. C. Gonzaga. Global convergence of filter methods for nonlinear programming. *SIAM J. Optimiz.*, 19(3):1231–1249, 2008.

[66] L. M. Rios and N. V. Sahinidis. Derivative-free optimization: a review of algorithms and comparison of software implementations. *J. Global Optim.*, 56:1247–1293, 2013.

[67] S. A. Santos. Trust-region-based methods for nonlinear programming: recent advances and perspectives. *Pesquisa Operacional*, 34:447–462, 2014.

[68] K. Scheinberg and Ph. L. Toint. Self-correcting geometry in model-based algorithms for derivative-free unconstrained optimization. *SIAM J. Optimiz.*, 20(6):3512–3532, 2010.

[69] K. Schittkowski. An updated set of 306 test problems for nonlinear programming with validated optimal solutions. Department of Computer Science, University of Bayreuth. Report, 2008.

[70] M. Ulbrich, S. Ulbrich, and L. N. Vicente. A globally convergent primal-dual interior-point filter method for nonlinear programming. *Math. Program.*, 100(2):379–410, 2004.

[71] A. Verdério, E. W. Karas, L. G. Pedroso, and K. Scheinberg. On the construction of quadratic models for derivative-free trust-region algorithms. Technical Report 01, Federal University of Paraná, 2015.

[72] A. Wächter and L. T. Biegler. Line search filter methods for nonlinear programming: motivation and global convergence. *SIAM J. Optimiz.*, 16(1):1–31, 2005.

[73] X. Wang, Z. Zhu, S. Zuo, and Q. Huang. An SQP-filter method for inequality constrained optimization and its global convergence. *Appl. Math. Comput.*, 217(24):10224–10230, 2011.

[74] Z. Yang and W. Sun. A filter-trust-region method for LC^1 unconstrained optimization and its global convergence. *Anal. Theor. Appl.*, 24(1):55–66, 2008.

[75] Y. X. Yuan. Recent advances in trust region algorithms. *Math. Program.*, 151(1):249–281, 2015.