A pattern search and implicit filtering algorithm for solving linearly constrained minimization problems with noisy objective functions*

M. A. Diniz-Ehrhardt† D. G. Ferreira† S. A. Santos†

December 5, 2017

Abstract

PSIFA –*Pattern Search and Implicit Filtering Algorithm*– is a derivative-free algorithm that has been designed for linearly constrained problems with noise in the objective function. It combines some elements of the pattern search approach of Lewis and Torczon (2000) with ideas from the method of implicit filtering of Kelley (2011) enhanced with a further analysis of the current face and a simple extrapolation strategy for updating the step length. The feasible set is explored by PSIFA without any particular assumption about its description, being the equality constraints handled in their original formulation. Besides, compact bounds for the variables are not mandatory. The global convergence analysis is presented, encompassing the degenerate case, under mild assumptions. Numerical experiments with linearly constrained problems from the literature were performed. Additionally, problems with the feasible set defined by polyhedral 3D-cones with several degrees of degeneration at the solution were addressed, including noisy functions that are not covered by the theoretical hypotheses. To put PSIFA in perspective, comparative tests have been prepared, with encouraging results.

Keywords: Derivative-free optimization; linearly constrained minimization; noisy optimization; global convergence; degenerate constraints; numerical experiments.

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

*Partially supported by Brazilian Funding Agencies *Fundação de Amparo à Pesquisa do Estado de São Paulo* - FAPESP (grants 2013/12964-4, 2013/05475-7, 2013/07375-0) and *Conselho Nacional de Desenvolvimento Científico e Tecnológico* - CNPq (grant 302915/2016-8).

†Institute of Mathematics, University of Campinas, Rua Sergio Buarque de Holanda, 651, 13083-859, Campinas, São Paulo, Brazil

1 Introduction

Over the last few decades, derivative-free nonlinear optimization methods have been widely addressed. The progress in this area may be explained by the increasing number of applications in which the derivatives of the objective function and/or of the constraints are not available, or in which the values of the objective function are unreliable, or do not exist at some of the points. Even finite-difference derivative approximations cannot always be applied, especially if the objective function is affected by noise. However, we cannot expect that derivative-free methods are competitive, in terms of performance, with derivative-based methods. Indeed, in approaches that do not use derivatives, the information whether the search direction is a descent direction, or not, cannot be exploited. Nevertheless, derivative-free methods are supplied with proper mechanisms to obtain good theoretical and computational results (see [5] and references therein).

In this study, we propose a derivative-free algorithm for solving linearly constrained optimization problems, in which the objective function has noise. Our proposal is based on the linearly constrained pattern search method from Lewis and Torczon [25], addressed in Ferreira's master dissertation [13], in which some modifications were introduced to the pattern search implementation and to the step length updating scheme, with convergence results similar to those of [25], and with encouraging numerical results.

The linearly constrained pattern search method had its origin in the *Generalized Pattern Search* (GPS) method, introduced by Torczon [31], originally for the unconstrained case. Next it was extended for bound constrained problems by Lewis and Torczon [24], culminating with the version that has addressed our problem of interest [25] (see also Audet and Dennis [2] and Lewis, Shepherd and Torczon [23]). Linearly constrained GPS algorithms obtain a feasible sequence of iterates, requiring simple decrease of the objective function at each iteration. The set of search directions must contain a generator for the cone of feasible directions, which must take into account the boundary of the feasible region close to the current iterate. In [24], the authors prove the existence of this generating set, including the nondegenerate as well as the degenerate case, and propose an algorithm for building the pattern in the nondegenerate case. The degenerate feasible set is addressed by Kolda, Lewis and Torczon [21] and by Lewis, Shepherd and Torczon [23], where the search directions are computed based on the extreme rays of a cone with a degenerate vertex at the origin. In [24], assuming that the objective function is continuously differentiable, and under some hypotheses that include integer search directions and a rational constraint matrix, Lewis and Torczon proved that the sequence of iterates produced by the pattern search method has at least a feasible limit point that is a Karush-Kuhn-Tucker (KKT) point. The GPS method is effective for many practical applications, specially when the objective function evaluation is expensive. The performance of the method is satisfactory even if the objective function is nonsmooth, as analyzed by Custódio, Dennis and Vicente [6]. Another related reference is the work of Price and Coope [29]. The authors have addressed unconstrained and linearly constrained problems (via simple barrier) by means of a class of frame-based derivative-free methods. Their convergence analysis rests upon Clarke's nonsmooth calculus and related results.

As we intend to handle some kind of noise in the objective function, besides pattern search elements, we also consider ideas from implicit filtering methods, which dates back to the works of Stoneking et. al [30] and Gilmore and Kelley [16]. These methods are present in Kelley's book [19], and have also been tackled by Finkel and Kelley [14], among others. In his more recent book [20], Kelley describes the IMFIL algorithm, that stands for implicit filtering, a derivative-free method for solving bound constrained optimization problems. Its main purpose is to minimize functions which are noisy, nonsmooth, possibly discontinuous, and that may not be defined at all points in the design space.

The implicit filtering method is also a sampling method, such as the pattern search algorithm.

However, it explores the sample, namely the coordinate search points (see, e.g. Hooke and Jeeves [18]), for computing the stencil gradient, which might represent the gradient of the objective function. The stencil gradient is used to build a quasi-Newton model Hessian, pursuing the fast local convergence inherent to interpolation methods based on models of smooth functions, in which first and/or second derivatives information are used explicitly. IMFIL, despite being designed for solving bound constrained problems, can also be applied to general constrained problems. Nevertheless, only the bound constraints are directly handled by the algorithm. More general constraints, the so-called *hidden constraints*, are incorporated in the objective function by means of an extreme barrier (cf. Audet and Dennis [3]).

Here we propose the algorithm PSIFA –*Pattern Search Implicit Filtering Algorithm*– based on ideas from the pattern search method combined with the implicit filtering approach. We inherit the search along the pattern, that allows us to take advantage of the resulting structure of the linear constraints, assuring the feasibility of the iterates. Furthermore, we use the pattern directions to build the corresponding stencil gradient, and to perform quasi-Newton BFGS steps, named after Broyden, Fletcher, Goldfarb, and Shanno [15], trying to improve even further a successful potential iterate. As several previously proposed derivative-free approaches (e.g. [9, 12, 26], to name a few), within PSIFA, the step length is expanded in case of indisputable progress. An additional particular feature of PSIFA is to admit a linearly dependent set of directions to build the generators of the pattern, so that the degenerate case is handled within the same framework of the non-degenerate one.

To recover the good convergence results from [25], however, some new features had to be considered. In [25], the authors require that the constraint matrix is rational and that the search directions have integer components. The latter assumption might not hold, since PSIFA can generate quasi-Newton based directions and, consequently, the iterates will not be restricted to a rational lattice. Inspired by elements from Diniz-Ehrhardt, Martínez and Raydan [10] and from Lucidi, Sciandrone and Tseng [27], we have included a nonmonotone sufficient decrease condition, which was combined with some other assumptions for obtaining an original global convergence result.

This paper is organized as follows: Section 2 was reserved to preliminary information concerning the problem and the notation. In Section 3, we present the PSIFA algorithm, highligthing its pattern search ingredients, the implicit filtering elements, as well as the corresponding enhancements. The convergence results of PSIFA are stated in Section 4. Numerical experiments are described and analyzed in Section 5. Conclusions regarding the theoretical and computational results, and also our work in progress are presented in Section 6.

The following notation is adopted. M^+ denotes the Moore-Penrose pseudo inverse of the matrix M; $|\mathcal{I}|$ denotes the cardinality of the set $\mathcal{I}$ and $\mathbb{R}_{++} = \{x \in \mathbb{R} \mid x > 0\}$.

2 Background

The problem of interest is

$$\begin{array}{ll} \min & f(x) \equiv f_s(x) + \Phi(x) \\ \text{s.t.} & \ell \leq Ax \leq u, \end{array} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{m\times n}$ and $\ell, u \in \mathbb{R}^m$. Some of the bounds might be infinite, and equality constraints are identified by the set of indices $\mathcal{E} = \{i \in \{1, \ldots, m\} \mid \ell_i = u_i\}$, whereas the genuine inequalities, i.e. those for which $\ell_i < u_i$, are addressed by the set $\mathcal{I} = \{1, \ldots, m\} \setminus \mathcal{E}$. The objective function in (1) is composed of a smooth function f_s and a noisy perturbation Φ. Denoting the feasible set by

$$\Omega = \{x \in \mathbb{R}^n \mid \ell \leq Ax \leq u\},$$

and the i-th row of matrix A by the row vector $a_{i\bullet}$, let, for $i \in \mathcal{I}$,

$$\partial\Omega_{\ell_i}(\epsilon) = \{x \in \Omega \mid 0 \leq a_{i\bullet}x - \ell_i \leq \epsilon\} \quad \text{and} \quad \partial\Omega_{u_i}(\epsilon) = \{x \in \Omega \mid 0 \leq u_i - a_{i\bullet}x \leq \epsilon\}$$

be the sets that define *feasible stripes* of width ϵ along the boundary of the i-th lower and upper inequality constraints, respectively.

The sets of the indices of the ϵ-*active constraints* at a point $x \in \Omega$ are denoted [25] by

$$\mathcal{I}_\ell(x,\epsilon) = \{i \in \mathcal{I} \mid x \in \partial\Omega_{\ell_i}(\epsilon)\} \quad \text{and} \quad \mathcal{I}_u(x,\epsilon) = \{i \in \mathcal{I} \mid x \in \partial\Omega_{u_i}(\epsilon)\}.$$

Given a feasible point x and a tolerance $\epsilon > 0$, a polyhedral cone is generated by the set of the gradient vectors of the ϵ-active inequality constraints at the point x, which are the outward pointing normals to the corresponding faces of the feasible set, together with the gradient vectors of the equality constraints and their opposites, namely

$$\mathcal{G}(x,\epsilon) = \{-a_{i\bullet}\}_{i\in\mathcal{I}_\ell(x,\epsilon)} \cup \{a_{i\bullet}\}_{i\in\mathcal{I}_u(x,\epsilon)} \cup \{\pm a_{i\bullet}\}_{i\in\mathcal{E}}.$$

Such a cone is denoted by $\mathcal{K}(x,\epsilon)$ [25] and expressed as

$$\mathcal{K}(x,\epsilon) = \left\{ v \in \mathbb{R}^n \mid v = \sum_{i=1}^{|\mathcal{G}(x,\epsilon)|} t_i v_i, v_i \in \mathcal{G}(x,\epsilon), t_i \geq 0, i = 1,\ldots,|\mathcal{G}(x,\epsilon)| \right\}. \tag{2}$$

The polar of $\mathcal{K}(x,\epsilon)$, denoted by $\mathcal{K}^\circ(x,\epsilon)$, is stated as

$$\begin{aligned}\mathcal{K}^\circ(x,\epsilon) &= \left\{ w \in \mathbb{R}^n \mid v^\top w \leq 0, \forall v \in \mathcal{K}(x,\epsilon) \right\} \\ &= \{w \in \mathbb{R}^n \mid a_{i\bullet}w \geq 0, i \in \mathcal{I}_\ell(x,\epsilon);\ a_{i\bullet}w \leq 0, i \in \mathcal{I}_u(x,\epsilon); a_{i\bullet}w = 0, i \in \mathcal{E}\}.\end{aligned} \tag{3}$$

Therefore, the set $\mathcal{K}^\circ(x,\epsilon)$ is the cone of the feasible directions from x, associated with the ϵ-active inequality and equality constraints. Consequently, $\mathcal{K}^\circ(x,0)$ comprises all the feasible directions emanating from the feasible point x. Notice that, if $\mathcal{K}(x,\epsilon) = \{0\}$ then $\mathcal{K}^\circ(x,\epsilon) = \mathbb{R}^n$.

3 The PSIFA algorithm

Despite being based on ideas from the pattern search method combined with elements from the implicit filtering strategy, our approach does not require the iterates to lie on a rational lattice, as in [23, 25], neither depends on compact simple bounds for rescaling the problem variables, as assumed in [20]. Therefore, specific tools and results have been developed, accompanied by an original theoretical analysis. The main ingredients of PSIFA are described next, in preparation for the algorithm, which is presented at the end of this section.

3.1 Building the pattern

The next result provides the essential elements for building the pattern of directions at each iteration, which is the core of our method. It is an extension of Proposition 8.2 of Lewis and Torczon [25], devised to encompass nondegenerate as well as degenerate feasible sets. It comes as a consequence of our approach not being restricted to a rational lattice.

Proposition 3.1. *Given $x \in \Omega$ and $\delta > 0$, assume that the columns of the matrix $V \equiv V(x, \delta) \in \mathbb{R}^{n\times p}$ contain a generator for the cone $\mathcal{K}(x, \delta)$, and the columns of $N \in \mathbb{R}^{n\times q}$ generate the null space of $V^\top$. Then, for any $\epsilon \in [0, \delta]$, a set of positive generators for $\mathcal{K}^\circ(x, \epsilon)$ can be obtained from the columns of N, $-N$, $V(V^\top V)^+$ and $-V(V^\top V)^+$.*

Proof. Given $x \in \Omega$ and $\delta > 0$, let $\mathcal{K} \equiv \mathcal{K}(x, \delta)$ and $\mathcal{K}^\circ \equiv \mathcal{K}^\circ(x, \delta)$. If $w \in \mathcal{K}^\circ$, then for all $v \in \mathcal{K}$ we have that $w^\top v \leq 0$. From the first hypothesis, any $v \in \mathcal{K}$ is such that $v = V\lambda$, $\lambda \geq 0$, and thus $w^\top V\lambda \leq 0$.

Let $w = w_{\mathcal{N}} + w_{\mathcal{R}}$, with $w_{\mathcal{N}} \in \mathcal{N}(V^\top)$, the null space of $V^\top$, and $w_{\mathcal{R}} \in \mathcal{R}(V)$, the range of V. Therefore, as $w_{\mathcal{N}}^\top V\lambda = 0$, we obtain $w_{\mathcal{R}}^\top V\lambda \leq 0$. Moreover, $w_{\mathcal{R}} = Vy$, in which y is a solution of the problem

$$\min_{y\in\mathbb{R}^p} \|Vy - w\|_2^2. \tag{4}$$

The minimum norm solution of (4) is given by $y = V^+w$. In view of the previously established relationships and the properties of the pseudo inverse V^+ we have

$$0 \geq w_{\mathcal{R}}^\top V\lambda = \left[VV^+w\right]^\top V\lambda = w^\top \left[VV^+\right]^\top V\lambda = w^\top VV^+V\lambda = w^\top V\lambda = \tau^\top\lambda,$$

where $\tau = V^\top w$. Since $\tau^\top\lambda \leq 0$ holds for all $\lambda \geq 0$, it follows that $\tau \leq 0$.

Now, from the singular value decomposition of V, it turns out that $(V^\top V)^+V^\top = V^+$. Consequently, the minimum norm solution of (4) may be expressed as

$$y = V^+w = (V^\top V)^+V^\top w = (V^\top V)^+\tau.$$

As a result,

$$w_{\mathcal{R}} = Vy = V(V^\top V)^+\tau, \text{ with } \tau \leq 0.$$

From the second hypothesis, as the columns of N generate $\mathcal{N}(V^\top)$, any $w \in \mathcal{K}^\circ$ can be writen as

$$w = [N \ \ -N]\zeta + (-V(V^\top V)^+\bar{\tau}), \ \text{ with } \zeta \geq 0 \text{ and } \bar{\tau} \geq 0.$$

Hence, the columns of N, $-N$ and $-V(V^\top V)^+$ form a positive generator for $\mathcal{K}^\circ$. However, for $\epsilon \in [0, \delta)$, the cone $\widetilde{\mathcal{K}} \equiv \mathcal{K}(x, \epsilon)$ may be generated by just a subset of the columns of V. Without loss of generality, assume that the first r columns of V, $r \leq p$, do not belong to the generator subset. By considering $w \in \widetilde{\mathcal{K}}^\circ \equiv \mathcal{K}^\circ(x, \epsilon)$, with the previous reasoning we obtain $w^\top V\lambda \leq 0$ with $\lambda \geq 0$ and $\lambda_1 = \cdots = \lambda_r = 0$. Thus, defining $\tau = V^\top w$, we have $\tau^\top\lambda \leq 0$ with $\tau_{r+1}, \ldots, \tau_p \leq 0$ but $\tau_1, \ldots, \tau_r$ are sign free. Therefore, the first r columns of $V(V^\top V)^+$ and their opposites, the last $p - r$ columns of $-V(V^\top V)^+$ and the columns of N and $-N$ positively generate $\widetilde{\mathcal{K}}^\circ$. To consider all the cases, by taking the columns of $V(V^\top V)^+$, $-V(V^\top V)^+$, N and $-N$ we always obtain a set of positive generators for $\mathcal{K}^\circ(x, \epsilon)$, for any $\epsilon \in [0, \delta]$, and the proof is complete. □

Due to the feasible feature of our approach, if problem (1) contains equality constraints, they must be satisfied by the generated sequence of iterates. Hence, we introduce the set of generators of equality constraints

$$\mathcal{W}_{eq} = \{\pm\, a_{i\bullet} \mid i \in \mathcal{E}\}. \tag{5}$$

The genuine inequality constraints, on the other hand, are handled based on their relative position with respect to the point of interest x and the given tolerance ϵ to detect the boundary. Therefore, given $x \in \Omega$ and $\epsilon > 0$, the generators of the inequality constraints are

$$\mathcal{W}_{ineq}(x, \epsilon) = \{-a_{i\bullet} \mid i \in \mathcal{I}_\ell(x, \epsilon)\} \cup \{a_{i\bullet} \mid i \in \mathcal{I}_u(x, \epsilon)\}. \tag{6}$$

Moreover, the face defined by the numerically binding inequality constraints at x is further explored by PSIFA. Such a face is detected by means of a tighter and fixed tolerance $\bar{\epsilon} \in (0, \epsilon)$, being handled by the set of indices

$$\mathcal{I}_{bin}(x, \bar{\epsilon}) = \mathcal{I}_\ell(x, \bar{\epsilon}) \cup \mathcal{I}_u(x, \bar{\epsilon})$$

and the generators

$$\mathcal{W}_{bin}(x, \bar{\epsilon}) = \{a_{i\bullet} \mid i \in \mathcal{I}_{bin}(x, \bar{\epsilon})\}. \tag{7}$$

We stress that the additional generators in (7) improve the further analysis of the current face and are considered separately from (5) and (6) because, in case the current face is non-optimal, the algorithm should be able to abandon it.

At the current iterate $x^k \in \Omega$, given the tolerances ϵ_k and $\bar{\epsilon}$ such that $0 < \bar{\epsilon} < \epsilon_k$, the three sets $\mathcal{W}_{eq}$, $\mathcal{W}_{ineq}(x^k, \epsilon_k)$ and $\mathcal{W}_{bin}(x^k, \bar{\epsilon})$ constitute the *current working set*.

Based on Proposition 3.1, the pattern of directions $\mathcal{P}^k \subset \mathbb{R}^n$ is built within the following procedure.

Procedure for building the pattern

Input: $x^k \in \Omega$, $\epsilon_k > 0$, $\bar{\epsilon} > 0$.
Output: $P^k \in \mathbb{R}^{n \times r_k}$.

Step 1. Build the working sets, i.e., the sets $\mathcal{W}_{eq}$, $\mathcal{W}_{ineq}(x^k, \epsilon_k)$ and $\mathcal{W}_{bin}(x^k, \bar{\epsilon})$, stated in (5), (6) and (7), respectively.

Step 2. Build the matrices V and $\overline{V}$, whose columns are the generators that comprise the sets $\mathcal{W}_{eq} \cup \mathcal{W}_{ineq}(x^k, \epsilon_k)$ and $\{a_{i\bullet} \mid i \in \mathcal{E}\} \cup \mathcal{W}_{bin}(x^k, \bar{\epsilon})$, respectively

Step 3. Compute bases for the null spaces $\mathcal{N}(V^\top)$ and $\mathcal{N}(\overline{V}^\top)$, that define the columns of matrices N and $\overline{N}$, respectively.

Step 4. Compute the matrix $F = V(V^\top V)^+$.

Step 5. Normalize each of the columns of the matrices N, F, and $\overline{N}$.

Step 6. Build the pattern matrix as

$$P^k = \begin{bmatrix} N & -N & F & -F & \overline{N} & -\overline{N} \end{bmatrix} \in \mathbb{R}^{n \times r_k},$$

so that its columns d_i^k, $i = 1, \ldots, r_k$, define the set $\mathcal{P}^k$.

Although an opportunistic poll could be performed (cf. [21]) among the directions that ensemble the pattern set $\mathcal{P}^k$, we have adopted a complete poll. Our choice aims to collect enough information for computing the so-called stencil gradient, as defined below, and potentially accelerate the practical performance of the method. Moreover, as the generated sequence of points is not restricted to a rational lattice, following ideas from [10, 27], a nonmonotone sufficient decrease condition has been imposed to ensure global convergence. As detailed along the convergence analysis, relaxing the monotonicity turned to be crucial to accomodate the general nature of Φ, the noisy component of the objective function of problem (1).

3.2 The potential quasi-Newton step

After the pattern search phase, if an effective decrease of the objective function is achieved at the trial point, and with the aim of accelerating the method, a quasi-Newton BFGS step is computed, as suggested in [20]. The demanded gradients are replaced by the so-called *stencil gradient* [19], which may be seen as the gradient of a linear regression model for the objective function [5].

Let W be the matrix whose columns are all the search directions that have generated feasible points in the pattern search phase at the current iteration, i.e.

$$W = [w_1 \ \ w_2 \ \ \cdots \ \ w_q] \in \mathbb{R}^{n\times q}.$$

The stencil gradient is defined by

$$\nabla f(x, W, \alpha)^\top = \frac{1}{\alpha}\delta(f, x, W, \alpha)^\top W^+,$$

where α is the successful step length of the pattern search phase and $\delta(f, x, W, \alpha)$ denotes the vector of objective function differences, that is

$$\delta(f, x, W, \alpha)^\top = [f(x + \alpha w_1) - f(x), f(x + \alpha w_2) - f(x), \dots, f(x + \alpha w_q) - f(x)]\,.$$

The quasi-Newton step d_{QN} is obtained by solving the linear system

$$B_k d = -\nabla f(x^k, W_k, \alpha_k) \tag{8}$$

in which $B_k \in \mathbb{R}^{n\times n}$ is the current model Hessian and B_0 is given (e.g. the identity).

Due to the approximate nature of the stencil gradient, there is no guarantee that the solution of (8) is a descent direction. Moreover, it might be that $x^k + d_{QN} \notin \Omega$. Thus, in case the problem has equality constraints, by assembling the matrix V_{eq} whose columns are the vectors $\{a_{i\bullet} \mid i \in \mathcal{E}\}$, the direction d_{QN} should be projected onto the null space $\mathcal{N}(V_{eq}^\top)$ to ensure feasibility.

Next, the maximum feasible step length along d_{QN} (or its projection, also denoted by d_{QN}, for simplicity), should be computed by a *minimum ratio test* as follows:

- compute $t_\ell = \min\limits_{i:a_{i\bullet}d_{QN}<0} \left\{\dfrac{\ell_i - a_{i\bullet}x^k}{a_{i\bullet}d_{QN}}\right\}$ and $t_u = \min\limits_{i:a_{i\bullet}d_{QN}>0} \left\{\dfrac{u_i - a_{i\bullet}x^k}{a_{i\bullet}d_{QN}}\right\}$;
- set $t_{\max} = \min\{t_\ell, t_u, 1\}$.

Finally, starting with the step length $t_{\max}$, and pursuing $f(x^k + td_{QN}) < f(x^k)$, a backtracking line search is performed, reducing the step length at most $j_{\max}$ times. In case the aforementioned simple decrease is obtained, we define $x_{QN} = x^k + td_{QN}$. Recalling that we already have a trial point from the pattern search phase, namely $x_{PS} = x^k + \alpha_k d_j^k$, such that $f(x_{PS}) < f(x^k)$, we may choose between x_{PS} and x_{QN} to define x^{k+1}, possibly the one that decreases the objective function most. Otherwise, if the quasi-Newton step is unsuccessful, then we simply set $x^{k+1} = x_{PS}$.

An estimate for the full model Hessian is maintained, following [20], using the standard BFGS update

$$B_k = B_{\bar{k}} - \frac{B_{\bar{k}} s s^\top B_{\bar{k}}}{s^\top B_{\bar{k}} s} + \frac{y y^\top}{y^\top s},$$

based on the vectors $s = x^k - x^{\bar{k}}$ and $y = \nabla f(x^k, W_k, \alpha_k) - \nabla f(x^{\bar{k}}, W_{\bar{k}}, \alpha_{\bar{k}})$, where $\bar{k}$ is the previous iteration in which a quasi-Newton step has been computed.

3.3 The algorithm

Algorithm 1: PSIFA (Pattern Search Implicit Filtering Algorithm)

Input. $x^0 \in \Omega$, $\{\widetilde{\alpha}_0, \alpha_{\max}, \epsilon_0, \bar{\epsilon}, \gamma\} \in (0, +\infty)$, $p > 1$, $\rho_1 > 1$, $\rho_2 \in (0,1)$, $\{\eta_k\} \subset \mathbb{R}_{++}$ such that $\sum_{k=0}^{\infty} \eta_k < \infty$.

Initialization. Based on x^0, ϵ_0 and $\bar{\epsilon}$, obtain P^0 by means of the *Procedure for building the pattern.*

For $k = 0, 1, \ldots$ **do**

Step 1. Set $\alpha = \widetilde{\alpha}_k$.

Step 2. If there exists $d_j^k \in P^k$, $j = 1, \ldots, r_k$, such that

$$x^k + \alpha d_j^k \in \Omega \quad \text{and} \quad \min_{1 \leq j \leq r_k} \{f(x^k + \alpha d_j^k)\} \leq f(x^k) - \gamma \alpha^p + \eta_k,$$

set $\alpha_k = \alpha$, $d_j^k = d^k$, $x_{PS} = x^k + \alpha d_j^k$, and

if ($k = 0$ or $x^k \neq x^{k-1}$) then $\widetilde{\alpha}_{k+1} = \min\{\alpha_{\max}, \rho_1 \alpha\}$, else $\widetilde{\alpha}_{k+1} = \alpha$.

Otherwise, set $\alpha_k = 0$, $\widetilde{\alpha}_{k+1} = \rho_2 \widetilde{\alpha}_k$, and $x_{PS} = x^k$.

Step 3. If $f(x_{PS}) < f(x^k)$ then set $x^{k+1} = \arg\min\{f(x_{PS}), f(x_{QN})\}$, where x_{QN} is obtained from a stencil gradient-based BFGS step (cf. § 3.2).

Otherwise, set $x^{k+1} = x_{PS}$.

Step 4. Set $\epsilon_{k+1} = \min\{1, \widetilde{\alpha}_{k+1}\}$.

Step 5. Obtain P^{k+1}, based on x^{k+1}, ϵ_{k+1} and $\bar{\epsilon}$, by means of the *Procedure for building the pattern.*

Remarks concerning Algorithm 1:

1. *PSIFA is well defined.* From the dynamics of the algorithm, the next iterate x^{k+1} is always well defined, being either distinct from x^k (the *successful* iterations), or coinciding with x^k (the *unsuccessful* iterations). Such a nomenclature is classic, see e.g. [25].

2. *Further analysis of the current face.* The generators corresponding to the binding inequality constraints, stated in (7), together with the gradients of the equality constraints, provide additional elements to the pattern, namely the columns of the matrices $\overline{N}$ and $-\overline{N}$. This is a peculiarity of PSIFA, that yields a further analysis of the current face. Consequently, premature stops at the optimal face but at a point that is non-optimal are avoided. Besides, preliminary experiments [13] have revealed that this further analysis improves the practical performance of the algorithm.

3. *Computing x_{QN}.* After the pattern search iterate x_{PS} has been computed and, besides fulfilling the nonmonotone sufficient decrease condition of Step 2, the point x_{PS} provides a simple decrease of the objective function f, a further improvement is tried at Step 3. This procedure is based on an approximate quasi-Newton step, as detailed in Subsection 3.2. The stencil gradient based quasi-Newton step comes from Kelley's implicit filtering [20], but PSIFA demanded proper enhancements to ensure feasibility with respect to the general linear constraints within the set Ω, such as the projection onto the null space $\mathcal{N}(V_{eq}^{\top})$, as well as the minimum ratio test.

4. *Extrapolation strategy for updating the step length.* Another feature of PSIFA is a simple extrapolation strategy, adopted for updating the step length. Every time the initial step length of the current iteration satisfies the nonmonotone sufficient decrease condition of Step 2, the first tentative step length of the next iteration may increase. This is accomplished by means of the updating $\widetilde{\alpha}_{k+1} = \min\{\alpha_{\max}, \rho_1 \alpha\}$, whenever $k = 0$ or $x^k \neq x^{k-1}$.

5. *Updating the tolerance.* Step 4 was taken so that the relationship $\epsilon_k = \mathcal{O}(\widetilde{\alpha}_k)$ holds (see, e.g. [22]). Such a connection is important for the convergence analysis of the algorithm, developed at the next section, being an assumption of Theorem 4.11.

4 Convergence analysis

To develop the convergence analysis of Algorithm 1, we provide next the hypotheses that the generated sequence must fulfill, namely feasibility and nonmonotone sufficient decrease [10, 27].

Conditions upon the generated sequence. Given a step length $\alpha_k \geq 0$, the iterate is set as

$$x^{k+1} = \begin{cases} x^k + \alpha_k d^k, & \alpha_k > 0 \\ x^k, & \alpha_k = 0. \end{cases}$$

In both cases,

C1. the feasibility is preserved, i.e.

$$x^k + \alpha_k d^k \in \Omega; \tag{9}$$

C2. a nonmonotone sufficient decrease is obtained, i.e.

$$f(x^{k+1}) \leq f(x^k) - \gamma \alpha_k^p + \eta_k, \tag{10}$$

where $p > 1$ and $\gamma \in (0,1)$ are algorithmic constants and $\{\eta_k\} \subset \mathbb{R}_{++}$ is a pre-established forcing sequence that satisfies $\sum_{k=0}^{\infty} \eta_k < +\infty$.

We stress that C1 and C2 are both ensured by Step 2. Moreover, the potential improvement provided by Step 3 has a practical purpose, preserving these hypotheses, without interfering upon the theoretical results developed along this section.

Next we state two important definitions to our analysis. The first one is a tool for handling the noise [20], whereas the second is classic (e.g. [4]), included for completeness and for establishing the notation.

Definition 4.1. *(Local norm of the noise) Let $\mathcal{S}(x,t,\mathcal{V})$ be a set of points determined by directions in $\mathcal{V} \subset \mathbb{R}^n$, from $x \in \Omega$ and step length $t > 0$, i.e. points of the form $z = x + td$, for all $d \in \mathcal{V}$. The* local norm of the noise *is defined by*

$$\|\Phi\|_{\mathcal{S}(x,t,\mathcal{V})} = \max_{z \in \{x\} \cup \mathcal{S}(x,t,\mathcal{V})} |\Phi(z)|. \tag{11}$$

Definition 4.2. *(Tangent cone) Given a subset $\Omega \subset \mathbb{R}^n$ and a vector $x \in \Omega$, the vector $d \in \mathbb{R}^n$ is said to be a tangent of Ω at x if either $d = 0$ or there exists a sequence $\{x^k\} \subset \Omega$ such that $x^k \neq x$ for all k and $x^k \to x$, $\frac{x^k - x}{\|x^k - x\|} \to \frac{d}{\|d\|}$. The set of all tangents to Ω at x is called the* tangent cone *of Ω at x and is denoted by $\mathcal{T}_\Omega(x)$.*

It is immediate from Definition 4.2 that the feasible directions with respect to the set Ω at a given point $x \in \Omega$ are also tangent directions. On the other hand, as Ω is a convex polytope, though not necessarily bounded, all the tangent directions are also feasible. Therefore,

$$\mathcal{T}_\Omega(x) = \mathcal{K}^\circ(x, 0). \tag{12}$$

From Proposition 3.1, given $\delta > 0$, the pattern $\mathcal{P}^k$ contains a generator set for the cone $\mathcal{K}^\circ(x^k, \epsilon_k)$, with $\epsilon_k \in [0, \delta]$. Hence, from (12), the pattern $\mathcal{P}^k$ contains a generator set for the cone $\mathcal{T}_\Omega(x^k)$.

A direct consequence of the hypothesis that the norm of the noise goes to zero faster than the step length is presented below.

Proposition 4.3. *(Property of the noise) Given a point $\bar{x} \in \Omega$, a step length $t > 0$ and a set of directions $\mathcal{V} \subset \mathbb{R}^n$, if $\Phi(\bar{x})$ exists and $\lim_{t \downarrow 0} \frac{\|\Phi\|_{\mathcal{S}(\bar{x},t,\mathcal{V})}}{t} = 0$ then $\Phi(\bar{x}) = 0$.*

Proof. Assume, for the sake of a contradiction, that $\Phi(\bar{x}) = a \neq 0$. By Definition 4.1, for $t > 0$,

$$\begin{aligned}\|\Phi\|_{\mathcal{S}(\bar{x},t,\mathcal{V})} &= \max_{z \in \{\bar{x}\} \cup \mathcal{S}(\bar{x},t,\mathcal{V})} |\Phi(z)| \\ &= \max\left\{ |\Phi(\bar{x})|, \max_{d \in \mathcal{V}} |\Phi(\bar{x} + td)| \right\} \\ &\geq |\Phi(\bar{x})| = |a|.\end{aligned}$$

Dividing the previous expression by $t > 0$ and taking the limit as $t \downarrow 0$ we obtain

$$\lim_{t \downarrow 0} \frac{\|\Phi\|_{\mathcal{S}(\bar{x},t,\mathcal{V})}}{t} \geq \lim_{t \downarrow 0} \frac{|a|}{t} = +\infty,$$

that contradicts the hypothesis and concludes the proof. □

Under the assumption that the norm of the noise decreases faster than the step length, a necessary optimality condition for problem (1) is given next.

Theorem 4.4. *(Necessary optimality condition) Let the components of the objective function of problem (1) be such that f_s is differentiable at $x^* \in \Omega$ and Φ satisfies*

$$\lim_{t \downarrow 0} \frac{\|\Phi\|_{\mathcal{S}(x^*,t,\mathcal{T}_\Omega(x^*))}}{t} = 0. \tag{13}$$

If x^* is a local minimizer of (1) then

$$\nabla f_s(x^*)^\top d \geq 0, \ \forall d \in \mathcal{T}_\Omega(x^*). \tag{14}$$

Proof. If $\mathcal{T}_\Omega(x^*) = \{0\}$, the result is trivially satisfied. Let $d \in \mathcal{T}_\Omega(x^*)$ be an arbitrary nonzero tangent at $x^* \in \Omega$. From Definition 4.2, there exist sequences $\{\xi^k\} \subset \mathbb{R}^n$ and $\{x^k\} \subset \Omega$ such that $x^k \neq x^*$ for all k, $\xi^k \to 0$, $x^k \to x^*$ and

$$\frac{x^k - x^*}{\|x^k - x^*\|} = \frac{d}{\|d\|} + \xi^k. \tag{15}$$

Based on (15), define $t_k = \|x^k - x^*\|$ and $d^k = \frac{d}{\|d\|} + \xi^k$. Then $\{t_k\} \subset \mathbb{R}_{++}$ is such that $t_k \downarrow 0$, $d^k \to \frac{d}{\|d\|}$ and $x^* + t_k d^k = x^k \in \Omega$ for all k.

As x^* is a local minimizer of problem (1) we have

$$0 \leq f(x^* + t_k d^k) - f(x^*) \tag{16a}$$
$$= f_s(x^* + t_k d^k) + \Phi(x^* + t_k d^k) - f_s(x^*) - \Phi(x^*) \tag{16b}$$
$$\leq f_s(x^* + t_k d^k) - f_s(x^*) + \|\Phi\|_{\mathcal{S}(x^*, t_k, \mathcal{T}_\Omega(x^*))}, \tag{16c}$$

where (16b) comes from the definition of function f and (16c) follows from Definition 4.1 and Proposition 4.3, observing that, due to the polyhedral and convex nature of Ω, as both x^k and x^* belong to Ω, then $x^k - x^* = t_k d^k$ is a feasible direction, and thus $d^k \in \mathcal{T}_\Omega(x^*)$, as the cone of feasible directions and the cone of tangent directions coincide.

In view of Taylor's expansion, as $\lim_{k\to+\infty} x^* + t_k d^k = x^*$ and the sequence $\{d^k\}$ is bounded, for k large enough, (16c) becomes

$$0 \leq t_k \nabla f_s(x^*)^\top d^k + o(t_k) + \|\Phi\|_{\mathcal{S}(x^*, t_k, \mathcal{T}_\Omega(x^*))}.$$

Dividing the previous inequality by $t_k > 0$ and taking the limit as $k \to \infty$, it follows that $t_k \downarrow 0$, and we have

$$0 \leq \lim_{k\to+\infty} \nabla f_s(x^*)^\top d^k \;+\; \lim_{t_k \downarrow 0} \frac{o(t_k)}{t_k} \;+\; \lim_{t_k \downarrow 0} \frac{\|\Phi\|_{\mathcal{S}(x^*, t_k, \mathcal{T}_\Omega(x^*))}}{t_k}.$$

Since $d^k \to d/\|d\|$ and the second and the third limits are equal to zero, the proof is complete. □

Theorem 4.4 establishes a necessary optimality condition for the noisy problem (1), based on which we define next a stationary point for (1).

Definition 4.5. *(Stationary point) A point $x^* \in \mathbb{R}^n$ is* stationary *for problem* (1) *if the components of the objective function are such that f_s is differentiable at $x^* \in \Omega$, Φ satisfies* (13) *and* (14) *holds.*

The following result is very useful in our convergence analysis, stating that at non-stationary points, the pattern set always contains a descent direction for the smooth component of the objective function.

Proposition 4.6. *Let $\mathcal{P}^k = \{d_1^k, \ldots, d_{r_k}^k\}$ be the pattern of directions used by Algorithm 1 at its k-th iteration, and let $x^k \in \Omega$ be the k-th iterate generated by this algorithm. If there exists $d \in \mathcal{T}_\Omega(x^k)$ such that $\nabla f_s(x^k)^\top d < 0$ then there exists $d_j^k \in \mathcal{P}^k$ such that $\nabla f_s(x^k)^\top d_j^k < 0$.*

Proof. Let $d \in \mathcal{T}_\Omega(x^k) = \mathcal{K}^\circ(x^k, 0)$. By Proposition 3.1, the pattern $\mathcal{P}^k$ contains a set of positive generators for $\mathcal{K}^\circ(x^k, 0)$, i.e. there exists $w \in \mathbb{R}_+^{r_k}$ such that $d = P^k w$. Hence, $0 > \nabla f_s(x^k)^\top d = \nabla f_s(x^k)^\top P^k w$. As w is a nonnegative vector, the previous relationship assures that, for all $j = 1, \ldots, r_k$, $\nabla f_s(x^k)^\top d_j^k \geq 0$ cannot hold. Therefore, there must exist $d_j^k \in \mathcal{P}^k$ such that $\nabla f_s(x^k)^\top d_j^k < 0$. □

Next we provide a theoretical consequence of the nonmonotone sufficient decrease condition adopted by PSIFA, being not only crucial to the main convergence result, but also an endorsement to the practical stopping criteria of the algorithm. It is worth mentioning that the thesis of our theorem is assumed in [25, Hypothesis 6] in order to obtain a stronger convergence result.

Theorem 4.7. *Assume that the objective function of problem* (1) *is bounded below in Ω. Let $\{\eta_k\} \subset \mathbb{R}_{++}$ be such that $\sum_{k=0}^{\infty} \eta_k < +\infty$, $\{x^k\} \subset \Omega$ be generated by Algorithm 1 and $\{\alpha_k\} \subset \mathbb{R}_+$ be the associated sequence of step lengths. Then*

$$\lim_{k\to+\infty} \alpha_k = 0.$$

Proof. In view of the boundedness of f in Ω, there exists $M \in \mathbb{R}$ such that

$$f(x^0) - M \geq f(x^0) - f(x^k).$$

Using a telescopic sum reasoning, we have

$$f(x^0) - M \geq \sum_{j=0}^{k-1} \left(f(x^j) - f(x^{j+1}) \right). \tag{17}$$

From condition C2 (10), the design of Step 2 and the assumption $\eta_k > 0$, for all k it holds

$$f(x^k) - f(x^{k+1}) \geq \gamma \alpha_k^p - \eta_k,$$

which applied to (17) gives

$$f(x^0) - M \geq \sum_{j=0}^{k-1} \left(\gamma \alpha_j^p - \eta_j, \right) \quad \Rightarrow \quad \sum_{j=0}^{k-1} \gamma \alpha_j^p \leq f(x^0) - M + \sum_{j=0}^{k-1} \eta_j.$$

Taking the inequality on the right to the limit as $k \to \infty$, because $\sum_{j=0}^{\infty} \eta_j < \infty$, it follows that $\sum_{j=0}^{\infty} \gamma \alpha_j^p < \infty$, so that the series is convergent. As $\gamma > 0$ and $p > 1$ are constant values, we conclude that $\lim_{k \to \infty} \alpha_k = 0$, and the proof is complete. □

Before proceeding with the asymptotic convergence analysis, we establish next a result concerning the finite convergence of PSIFA.

Proposition 4.8. *Let $\{x^k\} \subset \Omega$ be a sequence generated by the Algorithm 1 applied to problem (1) and let $\{\alpha_k\} \subset \mathbb{R}_+$ be the associated sequence of step lengths. If there exists an index $\bar{k}$ such that $\alpha_k = 0$ for all $k \geq \bar{k}$, the iterate $x^{\bar{k}}$ fulfills*

$$\lim_{t \downarrow 0} \frac{\|\Phi\|_{\mathcal{S}(x^{\bar{k}}, t, \mathcal{T}_\Omega(x^{\bar{k}}))}}{t} = 0, \tag{18}$$

the value $\Phi(x^{\bar{k}})$ exists, and f_s is differentiable at $x^{\bar{k}}$, then the point $x^{\bar{k}}$ is stationary for problem (1).

Proof. First, let $\bar{x}$ denote $x^{\bar{k}}$ along this proof. Reasoning by contradiction, assume that $\bar{x}$ is not stationary for problem (1). Thus, by Definition 4.5, as $\bar{x} \in \Omega$, there exists $d \in \mathcal{T}_\Omega(\bar{x})$ such that $\nabla f_s(\bar{x})^\top d < 0$. Therefore, by Proposition 4.6, there exists $d_j^{\bar{k}} \in \mathcal{P}^{\bar{k}}$ such that $\nabla f_s(\bar{x})^\top d_j^{\bar{k}} < 0$. We will introduce the notations $\bar{d} := d_j^{\bar{k}}$ and $a := -\nabla f_s(\bar{x})^\top \bar{d}$, which is strictly positive.

The pattern set $\mathcal{P}^{\bar{k}}$ is built as in Subsection 3.1, so it contains a positive generator for $\mathcal{K}^\circ(\bar{x}, 0) \equiv \mathcal{T}_\Omega(\bar{x})$, which coincides with the cone of feasible directions for Ω from $\bar{x}$. Hence, there exists $\bar{t} > 0$ such that $\bar{x} + t\bar{d} \in \Omega$ for all $t \in [0, \bar{t}]$. Consequently, for $t > 0$, we have $\|\Phi\|_{\mathcal{S}(\bar{x}, t, \mathcal{T}_\Omega(\bar{x}))} \geq \max_{z \in \{\bar{x}\} \cup \{\bar{x} + t\bar{d}\}} |\Phi(z)|$, and, in view of (18),

$$\lim_{t \downarrow 0} \frac{\|\Phi\|_{\mathcal{S}(\bar{x}, t, \{\bar{d}\})}}{t} = 0.$$

Therefore,

$$|\Phi(\bar{x} + t\bar{d})| = o(t). \tag{19}$$

From the differentiability of f_s at $\bar{x}$ we have $f_s(\bar{x} + t\bar{d}) = f_s(\bar{x}) - ta + o(t)$, so that, because $f = f_s + \Phi$, we obtain

$$f(\bar{x} + t\bar{d}) = f(\bar{x}) + \Phi(\bar{x} + t\bar{d}) - \Phi(\bar{x}) - ta + o(t) = f(\bar{x}) - ta + o(t),$$

where the last equaility comes from (19) and Proposition 4.3. As a result, for $\hat{t}$ positive and sufficiently small, it follows that $f(\bar{x} + t\bar{d}) \le f(\bar{x}) - ta/2$, for all $t \in [0, \hat{t}]$.

Without loss of generality, assume that $t \in [0, \min\{1, \bar{t}, \hat{t}\}]$. As $p > 1$, one has $t^p \le t$ and thus

$$f(\bar{x} + t\bar{d}) \le f(\bar{x}) - ta/2 \le f(\bar{x}) - t^p a/2.$$

If $\gamma \le a/2$, as $\eta_{\bar{k}} > 0$, then the relationship above implies that

$$f(\bar{x} + t\bar{d}) \le f(\bar{x}) - \gamma t^p + \eta_{\bar{k}} \tag{20}$$

is satisfied for all $t \in [0, \min\{1, \bar{t}, \hat{t}\}]$. On the other hand, if $\gamma > a/2$, then (20) holds for any $t \in \left[0, \min\left\{1, \bar{t}, \hat{t}, \left(\frac{\eta_{\bar{k}}}{\gamma - a/2}\right)^{1/p}\right\}\right]$. All in all, for t small enough and positive, it is possible to leave $\bar{x}$ along a direction from the pattern, keeping feasibility and fullfilling the nonmonotone sufficient decrease condition. Moreover, from the updating scheme of Step 2, as $0 < \widetilde{\alpha}_k \le t$ eventually becomes valid, the step length will be updated as $\alpha_k = \widetilde{\alpha}_k$, contradicting the assumption $\alpha_k = 0$, for all $k \ge \bar{k}$. □

Assuming from now onwards that the finite convergence set up at Proposition 4.8 does not occur, we proceed with our analysis. The next result establishes that, in a neighborhood of a limit point, the set of indices of the active constraints may be estimated from the set of indices of the ϵ-active constraints. It is an adaptation of Proposition 1 of [27], and the proof follows with simple adjustments. We highlight the usage of a convenient description for the index sets, encompassing the indices of the equality constraints that might be present at the problem formulation.

Lemma 4.9. *Let $\Omega \subset \mathbb{R}^n$ be the feasible set of problem* (1) *and let $\{x^k\} \subset \Omega$ be a sequence converging to $\bar{x} \in \Omega$. Then*

1. *there exists a scalar $\bar{\epsilon} > 0$, depending only on $\bar{x}$, such that for all $\epsilon \in (0, \bar{\epsilon}]$, there exists k_ϵ such that*
$$\mathcal{I}(x^k, \epsilon) = \mathcal{I}(\bar{x}),\ \forall k \ge k_\epsilon, \tag{21}$$
where
$$\mathcal{I}(x^k, \epsilon) = \left\{ i \in \{1, \dots, m\} \mid a_{i\bullet} x^k - \ell_i \le \epsilon \ \text{ or } \ u_i - a_{i\bullet} x^k \le \epsilon \right\},$$
$$\mathcal{I}(\bar{x}) = \{ i \in \{1, \dots, m\} \mid a_{i\bullet} \bar{x} - \ell_i = 0 \ \text{ or } \ u_i - a_{i\bullet} \bar{x} = 0 \}.$$

2. *For all k large enough we have*
$$\mathcal{I}(x^k, \epsilon_k) \subseteq \mathcal{I}(\bar{x}), \tag{22}$$
where $\epsilon_k > 0$ for all k and $\lim_{k\to\infty} \epsilon_k = 0$.

The next auxiliary result establishes that, under the convergence of the sequence of iterates, the associated sequence of polar cones to the ϵ_k-active constraints eventually stabilizes.

Lemma 4.10. *Let $\{x^k\} \subset \Omega$ be a sequence that converges to $\bar{x} \in \Omega$. Then, for k large enough it holds*

$$\mathcal{K}^\circ(x^k, \epsilon_k) = \mathcal{K}^\circ(\bar{x}, 0), \tag{23}$$

in which the sequence $\{\epsilon_k\}$ is such that $\epsilon_k > 0$ for all k and $\epsilon_k \to 0$.

Proof. The cones in (23) may be expressed as

$$\mathcal{K}^\circ(x^k, \epsilon_k) = \left\{ d \in \mathbb{R}^n \mid a_{i\bullet} d \geq 0, i \in \mathcal{I}_\ell(x^k, \epsilon_k); a_{i\bullet} d \leq 0, i \in \mathcal{I}_u(x^k, \epsilon_k); a_{i\bullet} d = 0, i \in \mathcal{E} \right\} \tag{24}$$

and

$$\mathcal{K}^\circ(\bar{x}, 0) = \{ d \in \mathbb{R}^n \mid a_{i\bullet} d \geq 0, i \in \mathcal{I}_\ell(\bar{x}, 0);\ a_{i\bullet} d \leq 0, i \in \mathcal{I}_u(\bar{x}, 0);\ a_{i\bullet} d = 0, i \in \mathcal{E} \}. \tag{25}$$

From the hypothesis that $x^k \to \bar{x}$ and applying the first item of Lemma 4.9, there exists $\bar{\epsilon} > 0$ such that, for any $\epsilon \in (0, \bar{\epsilon}]$, there exists an integer k_ϵ for which for all $k \geq k_\epsilon$, it holds $\mathcal{I}(x^k, \epsilon) = \mathcal{I}(\bar{x})$. Moreover, as $0 < \epsilon_k \to 0$, there exists an integer $\bar{k}$ such that $\epsilon_k \in (0, \bar{\epsilon}]$, for all $k \geq \bar{k}$, and thus $\mathcal{I}(x^k, \epsilon_k) = \mathcal{I}(\bar{x})$, for $k \geq \max\{k_\epsilon, \bar{k}\}$. Using this relationship into (24) and (25) and since $\mathcal{I}(x^k, \epsilon_k) = \mathcal{I}_\ell(x^k, \epsilon_k) \cup \mathcal{I}_u(x^k, \epsilon_k) \cup \mathcal{E}$ and $\mathcal{I}(\bar{x}) = \mathcal{I}_\ell(\bar{x}, 0) \cup \mathcal{I}_u(\bar{x}, 0) \cup \mathcal{E}$, we obtain the desired result. □

Next, we present the main convergence result of Algorithm 1.

Theorem 4.11. *Let $\{x^k\} \subset \Omega$ be a sequence generated by the Algorithm 1 applied to problem (1) with $\epsilon_k = \mathcal{O}(\widetilde{\alpha}_k)$, and let $\{\mathcal{P}^k\}$, with $\mathcal{P}^k = \{d_1^k, d_2^k, \cdots, d_{r_k}^k\}$, be a sequence of sets of pattern search directions generated by this algorithm, with uniformly bounded directions and $\{\eta_k\}$ be the associated forcing sequence. Assume that f admits a lower bound in Ω and verifies $f = f_s + \Phi$, with $f_s : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable and $\Phi : \mathbb{R}^n \to \mathbb{R}$ such that*

$$\lim_{\substack{k\to\infty \\ \alpha_k \neq 0}} \frac{\|\Phi\|_{\mathcal{S}(x^k, \alpha_k, \mathcal{P}^k)}}{\alpha_k} = 0. \tag{26}$$

If $x^* \in \Omega$ is a limit point of the sequence $\{x^k\}$, then x^* is stationary for problem (1), i.e.,

$$\nabla f_s(x^*)^\top d \geq 0, \quad \forall d \in \mathcal{T}_\Omega(x^*). \tag{27}$$

Proof. Let $\mathbb{K} \subset \{1, 2, \ldots\}$ be an infinite set of indices such that $\lim_{k\in\mathbb{K}} x^k = x^*$, $x^* \in \Omega$.

If there exists $\bar{k} \in \mathbb{K}$ such that $\alpha_k = 0$ for all $k \in \mathbb{K}, k \geq \bar{k}$, then, by the mechanism of the algorithm, $x^* = x^{\bar{k}}$. Therefore, reasoning as in Proposition 4.8 and under the assumptions thereof, it follows that x^* is stationary for problem (1).

Hence, for all $k \in \mathbb{N}$, we may assume that $x^k \neq x^*$. As a consequence, there exists an infinite set of indices $\mathbb{K}_1 \subset \mathbb{K}$ such that

$$0 < \alpha_k = \widetilde{\alpha}_k,\ \forall k \in \mathbb{K}_1. \tag{28}$$

Consider now an infinite set $\mathbb{K}_2 \subset \mathbb{K}_1$ for which $\{\alpha_k\}_{k\in\mathbb{K}_2}$ is monotonically decreasing. The existence of $\mathbb{K}_2$ is ensured by Theorem 4.7 and by the assumption that $\alpha_k > 0$ for all $k \in \mathbb{K}_1$. Using a more convenient notation for the latter sequence, we have

$$\{\alpha_{k_\ell}\}_{\ell=1}^{\infty},\ k_\ell \in \mathbb{K}_2, \text{ and for all } \ell \in \mathbb{N} \text{ it holds } \alpha_{k_\ell} > \alpha_{k_{\ell+1}}. \tag{29}$$

Notice that between the indices k_ℓ and $k_{\ell+1}$ there exists at least one integer. In fact, assume that these two indices are consecutive, i.e. $k_{\ell+1} = k_\ell + 1$. As both k_ℓ and $k_{\ell+1}$ belong to $\mathbb{K}_2 \subset \mathbb{K}_1$ then from (28) and (29), $\widetilde{\alpha}_{k_\ell} > \widetilde{\alpha}_{k_{\ell+1}}$. However, from the updating scheme of Step 2, $\alpha_{k_{\ell+1}} < \alpha_{k_\ell} = 0$, a contradiction with (28). Thus, take the index $k_\ell - 1$. The previous reasoning leads to $k_\ell - 1 \notin \mathbb{K}_1$. Therefore, $\alpha_{k_\ell - 1} = 0$, which implies that

$$\widetilde{\alpha}_{k_\ell - 1} > \widetilde{\alpha}_{k_\ell} = \alpha_{k_\ell}.$$

Observe that, from Step 2, the subsequence of tentative step lenghts satisfies

$$\{\widetilde{\alpha}_{k-1}\}_{k\in\mathbb{K}_2} = \frac{1}{\rho_2}\{\alpha_k\}_{k\in\mathbb{K}_2},$$

where we have resumed the simplified notation for not overloading the presentation. In view of Theorem 4.7, it follows that

$$\lim_{k\in\mathbb{K}_2} \widetilde{\alpha}_{k-1} = 0. \tag{30}$$

Moreover, for any $k \in \mathbb{K}_2$ there exists $i \in \{1, \ldots, r_k\}$ such that $x^k + \alpha_k d_i^k$ satisfies the nonmonotone sufficient decrease condition (10), but for all $j \in \{1, \ldots, r_k\}$, the trial points $x^k + \widetilde{\alpha}_{k-1} d_j^k$ did not achieve success in such a condition. Thus, for any $k \in \mathbb{K}_2$ and for all $d_j^k \in \mathcal{P}^k$, since $\eta_k > 0$ for all k, we have

$$f(x^k + \widetilde{\alpha}_{k-1} d_j^k) \; > \; f(x^k) - \gamma(\widetilde{\alpha}_{k-1})^p + \eta_k \; > \; f(x^k) - \gamma(\widetilde{\alpha}_{k-1})^p$$

so that

$$\begin{aligned} f_s(x^k + \widetilde{\alpha}_{k-1} d_j^k) + \Phi(x^k + \widetilde{\alpha}_{k-1} d_j^k) \;&>\; f_s(x^k) + \Phi(x^k) - \gamma(\widetilde{\alpha}_{k-1})^p \\ \Longrightarrow f_s(x^k + \widetilde{\alpha}_{k-1} d_j^k) \;&>\; f_s(x^k) - \gamma(\widetilde{\alpha}_{k-1})^p - \Phi(x^k + \widetilde{\alpha}_{k-1} d_j^k) + \Phi(x^k) \\ \Longrightarrow f_s(x^k + \widetilde{\alpha}_{k-1} d_j^k) \;&>\; f_s(x^k) - \gamma(\widetilde{\alpha}_{k-1})^p - 2\|\Phi\|_{S(x^k, \widetilde{\alpha}_{k-1}, \mathcal{P}^k)}. \end{aligned}$$

Replacing $f_s(x^k + \widetilde{\alpha}_{k-1} d_j^k)$ by its Taylor expansion in the previous inequality, for k large enough, and from the uniform limitation of the search directions, we have

$$f_s(x^k) + \widetilde{\alpha}_{k-1} \nabla f_s(x^k)^\top d_j^k + o(\widetilde{\alpha}_{k-1}) > f_s(x^k) - \gamma(\widetilde{\alpha}_{k-1})^p - 2\|\Phi\|_{S(x^k, \widetilde{\alpha}_{k-1}, \mathcal{P}^k)}.$$

Simplifying and dividing both sides of this relationship by $\widetilde{\alpha}_{k-1} > 0$ we obtain

$$\nabla f_s(x^k)^\top d_j^k + \frac{o(\widetilde{\alpha}_{k-1})}{\widetilde{\alpha}_{k-1}} > -\gamma(\widetilde{\alpha}_{k-1})^{p-1} - \frac{2\|\Phi\|_{S(x^k, \widetilde{\alpha}_{k-1}, \mathcal{P}^k)}}{\widetilde{\alpha}_{k-1}}. \tag{31}$$

As the subsequence $\{x^k\}_{k\in\mathbb{K}_2}$ can be regarded as a sequence that converges to x^*, applying Lemma 4.10 to $\{x^k\}_{k\in\mathbb{K}_2}$ we have that, for all sufficiently large $k \in \mathbb{K}_2$, it holds

$$\mathcal{K}^\circ(x^k, \epsilon_k) = \mathcal{K}^\circ(x^*, 0), \tag{32}$$

as $0 < \epsilon_k \to 0$, from the assumption that $\epsilon_k = \mathcal{O}(\widetilde{\alpha}_k)$, from (28) and Theorem 4.7. Consequently, for all sufficiently large $k \in \mathbb{K}_2$ we also have

$$\mathcal{K}(x^k, \epsilon_k) = \mathcal{K}(x^*, 0). \tag{33}$$

Concerning the pattern $\mathcal{P}^k$, from Proposition 3.1, it is built from the matrix $V(x^k, \epsilon_k)$, whose columns are the rows of A associated with the constraints that are ϵ_k-active at x^k, the generators of the cone $\mathcal{K}(x^k, \epsilon_k)$. From (33), for all sufficiently large $k \in \mathbb{K}_2$,

$$V(x^k, \epsilon_k) = V(x^*, 0).$$

Therefore, for all sufficiently large $k \in \mathbb{K}_2$, we have $\mathcal{P}^k = \mathcal{P}$, where $\mathcal{P}$ is the pattern search set built from the columns of $V(x^*, 0)$. In other words,

$$\{\mathcal{P}^k\}_{k\in\mathbb{K}_2} \to \mathcal{P}, \tag{34}$$

that is, $d_j^k \to d_j$, with $d_j^k \in \mathcal{P}^k$, $k \in \mathbb{K}_2$ and $d_j \in \mathcal{P}$. Moreover, from Proposition 3.1 and from (32), the set $\mathcal{P}$ contains a set of positive generators for $\mathcal{K}^\circ(x^*, 0) = \mathcal{T}_\Omega(x^*)$.

Now, from Proposition 4.6, if there exists $d \in \mathcal{T}_\Omega(x^*)$ such that $\nabla f_s(x^*)^\top d < 0$ then it must exist $d_j \in \mathcal{P}$ such that $\nabla f_s(x^*)^\top d_j < 0$.

Taking the limit for $k \in \mathbb{K}_2$ in (31) we have

$$\lim_{k\in\mathbb{K}_2} \nabla f_s(x^k)^\top d_j^k + \lim_{k\in\mathbb{K}_2} \frac{o(\widetilde{\alpha}_{k-1})}{\widetilde{\alpha}_{k-1}} \geq -\gamma \lim_{k\in\mathbb{K}_2} (\widetilde{\alpha}_{k-1})^{p-1} - 2 \lim_{k\in\mathbb{K}_2} \frac{\|\Phi\|_{S(x^k,\widetilde{\alpha}_{k-1},\mathcal{P}^k)}}{\widetilde{\alpha}_{k-1}},$$

so, from (26), (30) and (34) we obtain

$$\nabla f_s(x^*)^\top d_j \geq 0, \quad \forall d_j \in \mathcal{P}.$$

As a result, it does not exist any $d \in \mathcal{T}_\Omega(x^*)$ such that $\nabla f_s(x^*)^\top d < 0$, concluding the proof. □

5 Numerical experiments

The implementation of Algorithm 1 was coded in `matlab` R2014a 64-bits, and it will be denoted by `psifa`. The experiments were run in a notebook Dell Inspiron with a 2.6 GHz Intel Core i74510U processor (16GB RAM).

The parameters used for `psifa` were as follows: initial step length $\alpha_0 = 1$, initial tolerance $\epsilon_0 = 1$, tolerance for the binding constraints $\bar{\epsilon} = 10^{-10}$, extrapolation constant $\rho_1 = 2$, shrinking constant $\rho_2 = 0.5$, maximum step length $\alpha_{\max} = 10$, constants of the nonmonotone sufficient decrease condition $\gamma = 10^{-5}$ and $p = 1.5$, forcing sequence $\eta_0 = \gamma$, $\eta_k = \gamma/k^3$, $k \geq 1$, and maximum number of reductions along the quasi-Newton step $j_{\max} = 3$. With the practical purpose of avoiding reevaluations of the objective function, PSIFA carries over the history of function evaluations, what is particularly useful when dealing with expensive functions (see also [23, § 8.5]).

We have compared the performance of `psifa` with the routine `patternsearch`, from the optimization toolbox of `matlab`, as well as with two versions of `imfil`[1], the default and a version enriched with the directions generated by `psifa`. We stress that `imfil` demands compact simple bounds for the variables, performing a normalization to work within the box $[0, 1]^n$ and basically it rests upon coordinate directions. As already mentioned, additional constraints, either linear or nonlinear, are handled by `imfil` within the objective function by means of an extreme barrier approach (cf. [3]). However, `imfil` allows the user to provide supplementary directions. By providing to `imfil` the directions generated by `psifa`, we obtain the version '`imfil` (enriched)'. As this latter version maintains the pre-established and sequentially decreasing updating of the step lengths, inherent to `imfil`, the effect of the extrapolating strategy is employed solely by `psifa`. Moreover, the extra directions are used by `imfil` (enriched) after performing the polling along the coordinate directions, usually turning it less efficient than `psifa` and `patternsearch`. Last but not the least, `psifa` checks feasibility before evaluating the objective function, whereas `imfil` (enriched), similarly to `imfil`, addresses directly only simple bounds and handles the general linear constraints within the objective function.

[1] The `matlab` code is available at `http://www4.ncsu.edu/~ctk/darts/imfil.m`.

Additionally, a generalized pattern search algorithm guided by simplex derivatives (SID-PSM) [8, 7], also written in `matlab`[2], has been considered for addressing the nondegenerate problems. Denoted as `sid-psm`, such a solver does not handle equality constraints directly and faces numerical difficulties by turning each equality into two inequalities. Therefore, only for running `sid-psm`, the problems with equality constraints in the original formulation had to be manually pre-processed, in order to eliminate the equalities, with a consequent decreasing of the problem dimension.

There are two stopping conditions: attaining a sufficient small tentative step length ($\widetilde{\alpha}_k < 10^{-8}$) or reaching the budget on the number of function evaluations. Equivalent conditions were set for `patternsearch`, `imfil` and `sid-psm`. Recalling that the compared approaches generate feasible points, for declaring that a problem has been solved we have used the test [28]

$$f(x^0) - f(x) \geq (1-\tau)[f(x^0) - f_{\text{low}}], \tag{35}$$

where x^0 is the initial point, f_{low} is the smallest function value obtained among the compared solvers within a pre-established budget for function evaluations and $\tau > 0$ is the level of accuracy.

Despite having performed the analysis with varying levels of accuracy, namely $\tau \in \{10^{-1}, 10^{-3}, 10^{-5}\}$, we will report in detail only the results associated with $\tau = 10^{-3}$, as they offer a typical and significant perspective of the methods upon consideration. We have used the data profile tool of Moré and Wild [28] and the perfomance profiles of Dolan and Moré [11], with the measure of the number of demanded function evaluations, for presenting graphically the comparative results, always adopting 2000 function evaluations as the budget.

5.1 Linearly constrained problems from the literature

We have solved all the 32 linearly (and bound) constrained problems from the collection [17], adding to the smooth objective function of [17] the artificial nonsmooth noise

$$\Phi(x) = \mu\|x - x^*\|^2|\cos(80\|x - x^*\|)|, \tag{36}$$

being x^* the solution reported in [17], and $\mu > 0$ a parameter that controls the level of the noise. The dimensions varied as $2 \leq n \leq 16$ and $1 \leq m \leq 17$. For a sequence $\{x^k\}$ that converges to x^*, notice that $\Phi(x^k)$ goes to zero faster than the step length, and thus it fullfils the theoretical assumption (26) of the convergence result. We have adopted two levels of noise, $\mu = 0.01$ and $\mu = 0.05$. The corresponding data [28] and performance [11] profiles are depicted in Figures 1 and 2, respectively, with logarithmic scale in the horizontal axis, for better visualization. Performance profiles [11] were also prepared, from which we have collected, for each solver, the percentage of problems solved with the least amount of functional evaluations (efficiency) and the percentage of problems for which the criterion (35) was verified with $\tau = 10^{-3}$ (robustness). These results are summarized in Table 1, where the efficiency of `sid-psm` and the robustness of `psifa` emerge, for both levels of noise. Indeed, the percentage values corroborate the curves depicted in Figures 1 and 2.

The distribution of problems according with the type of linear constraints is as follows:

- 8 problems have only simple bounds on the variables;
- 10 problems have both simple bounds on the variables and inequality constraints;
- 7 problems have both simple bounds on the variables and equality constraints;

[2]See `http://www.mat.uc.pt/sid-psm/`.

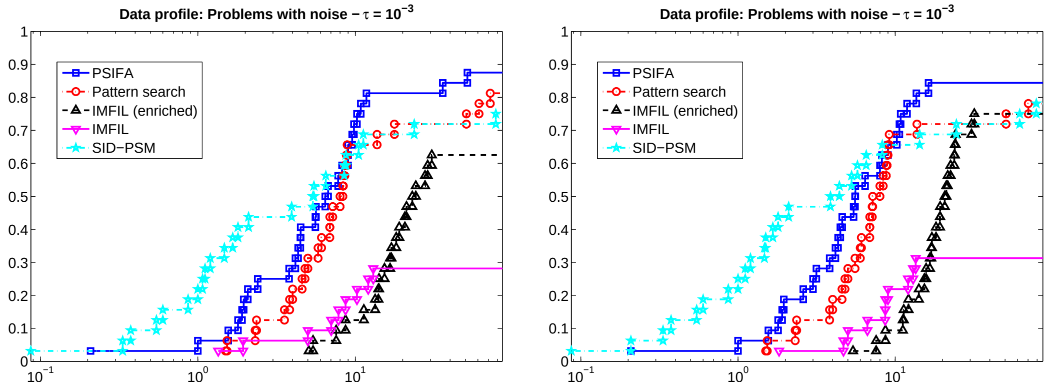

Figure 1: Data profiles of functional evaluations for the 32 linearly constrained problems from [17] and the synthetic noise (36) with $\mu = 0.01$ (*left*) and $\mu = 0.05$ (*right*).

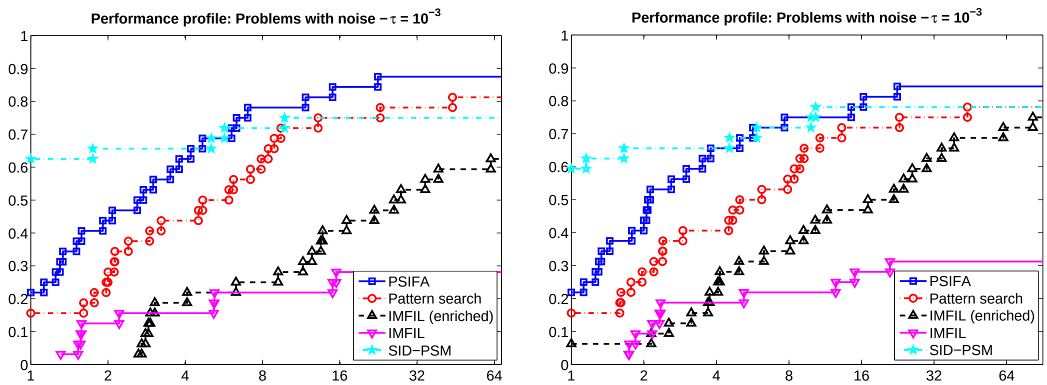

Figure 2: Performance profiles of functional evaluations for the 32 linearly constrained problems from [17] and the synthetic noise (36) with $\mu = 0.01$ (*left*) and $\mu = 0.05$ (*right*).

Table 1: Performance of the solvers for the 32 linearly constrained problems from [17].

	$\mu = 0.01$		$\mu = 0.05$	
Solver	Efficiency	Robustness	Efficiency	Robustness
`psifa`	21.87%	87.50%	21.87%	84.37%
`patternsearch`	15.62%	81.25%	15.62%	78.12%
`imfil` (enriched)	0.00%	62.50%	6.25%	75.00%
`imfil`	0.00%	28.12%	0.00%	31.25%
`sid-psm`	62.25%	75.00%	56.25%	78.12%

- 7 problems only have equality constraints.

Two aspects deserve attention. Firstly, for running `sid-psm`, the 7 equality constrained problems have been turned into lower-dimensional unconstrained problems. For the same purpose, the 7 equality constrained problems with simple bounds have been turned into lower-dimensional simple bounded problems. The aforementioned 14 problems, together with the 8 original simple bounded problems, amount to more than two-thirds of the 32 test problems from [17] with a very beneficial structure for being solved by `sid-psm`, somehow favoring the efficiency of this algorithm. Secondly, just 14 original problems from [17] actually possess compact simple bounds in their formulations. For each of the remaining 18 problems to be addressed by `imfil`, an artificial bounding box $\ell_x \leq x \leq u_x$ was built componentwise, based on the reported solution x^* and the initial point x^0, as follows:

$$(\ell_x)_i = \min\{x_i^*, x_i^0\} - 10 \quad \text{and} \quad (u_x)_i = \max\{x_i^*, x_i^0\} + 10, \quad i = 1, \ldots, n. \tag{37}$$

5.2 Problems with controlled degree of degeneracy

We have solved degenerate geometric problems with the feasible set defined by polyhedral 3D-cones with control of generated faces and vertex at the origin, as in Andreani et. al. [1]. The feasible set is $\Omega = \{x \in \mathbb{R}^3 \mid Ax \geq 0\}$, where the i-th row of matrix $A \in \mathbb{R}^{m\times 3}$ is the vector

$$a_{i\bullet}^T \equiv \begin{pmatrix} \sin\left(\frac{2\pi}{m}i\right)\left(\cos\left(\frac{2\pi}{m}\right) - 1\right) - \cos\left(\frac{2\pi}{m}i\right)\sin\left(\frac{2\pi}{m}\right) \\ \cos\left(\frac{2\pi}{m}i\right)\left(1 - \cos\left(\frac{2\pi}{m}\right)\right) - \sin\left(\frac{2\pi}{m}i\right)\sin\left(\frac{2\pi}{m}\right) \\ r\sin\left(\frac{2\pi}{m}\right) \end{pmatrix}, \quad i = 1, \ldots, m.$$

The number of faces varied with $m \in \{4, 5, 6, 9, 12, 15, 18\}$ and the opening of the edges was controlled by the parameter $r \in \{0.1, 1, 10\}$, which amounted to 21 instances for the feasible set.

We have chosen two possibilities for the f_s component of the objective function, either the quadratic

$$f_s^Q(x) = x_1^2 + x_2^2 + (x_3 + 1)^2 \tag{38}$$

or the function

$$f_s^N(x) = \sum_{i=1}^m \sqrt{a_{i\bullet}x}, \tag{39}$$

for which the feasible set plays the role of the domain as well. It is worth noticing that f_s^N is nonsmooth at the origin, being not covered by our theoretical analysis.

We have considered two types of noise, one that fullfills the theoretical assumption, namely (36), with $x^* = 0$ and $\mu = 0.05$, and also a *white noise*, defined as a piecewise constant function in a 3D-mesh of size 0.05, with amplitude randomly and uniformly generated in the interval $[-0.05, 0.05]$. Despite not being covered by the theoretical convergence analysis, nor having an analytic expression, the white noise might have a practical appeal, such as modelling lack of precision in the data, for instance.

The initial approximation was set as the point $x^0 = \left(\frac{r}{2}\cos\left(\frac{2\pi}{m}\right), \frac{r}{2}\sin\left(\frac{2\pi}{m}\right), 1\right)^T$, always strictly inside the polyhedral cone Ω.

With the synthetic expression for the noise, the solution is the origin for both choices of f_s, whereas with the white noise, it may not be the case. Indeed, when the synthetic noise (36) is used, the optimal value is 1 for the problem with f_s^Q, and zero for f_s^N. With the white noise, these optimal values slightly varied since the white noise is not necessarily zero at the solution.

The comparative results are shown in Figures 3–6, with logarithmic scale in the horizontal axis. For running `imfil` and '`imfil` (enriched)', an artificial bounding box has been introduced, as in (37). The algorithm `sid-psm` could not solve any degenerate problem, therefore it is absent from the comparative analysis of this subsection. The corresponding summary of the performance is shown in Tables 2 and 3, following the same pattern of the previous subsection, using $\tau = 10^{-3}$ in the test (35).

As can be seen in the profiles and in the percentage values of the tables, the performance of **psifa** is very good for the solved degenerate instances, not only for problems within the theoretical assumptions (Figures 3 and 4, and Table 2), but also for problems with white noise that does not vanish in a neighborhood of the solution, that is, in tests that are not covered by the theory (Figures 5 and 6, and Table 3). Clearly, the latter problems are much more difficult, giving a hard time to **patternsearch** and to **imfil**. In fact, in these degenerate instances, for the quadratic function, **patternsearch** had the best performance, whereas for the nonsmooth function, **patternsearch** is the second more efficient, and the enriched version of **imfil** is the second more robust. Concerning the nonsmooth function, by keeping track of the sequences of functional values $\{f(x^k)\}$ and step lenghts $\{\alpha_k\}$, we have noticed that the outstanding performance of **psifa** is mainly a consequence of the generated directions that ensembled the pattern. Indeed, such a set of directions also benefit the robustness of '**imfil** (enriched)'. Nevertheless, as already explained in the beginning of this section, the dynamic of '**imfil** (enriched)' using the extra directions of **psifa**, as well as its way of addressing the feasible set, overload its performance with excessively extra functional evaluations.

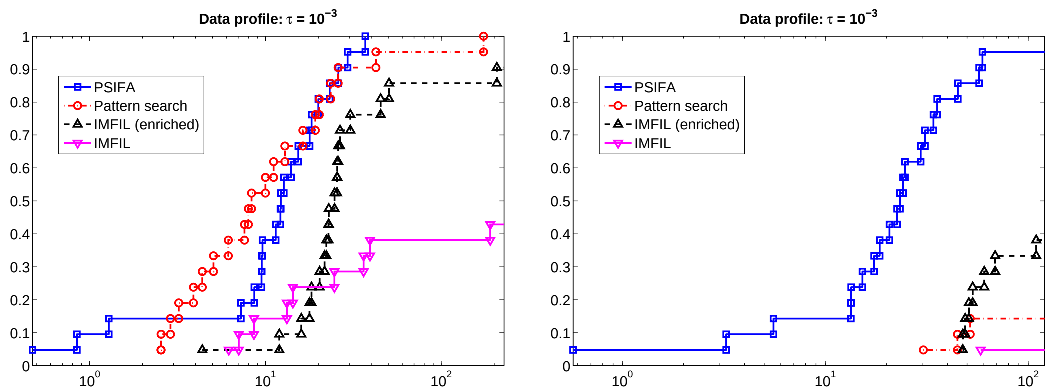

Figure 3: Data profiles of functional evaluations for problems with the synthetic noise (36) and either the quadratic function (*left*) f_s^Q, given by (38), or the nonsmooth function (*right*) f_s^N, given by (39).

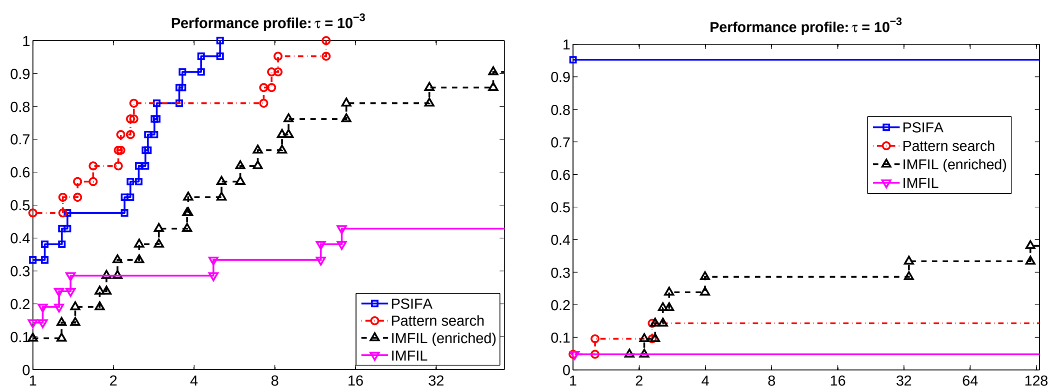

Figure 4: Performance profiles of functional evaluations for problems with the synthetic noise (36) and either the quadratic function (*left*) f_s^Q, given by (38), or the nonsmooth function (*right*) f_s^N, given by (39).

Table 2: Performance of the solvers for the 21 degenerate instances with the synthetic noise (36).

	Quadratic function		Nonsmooth function	
Solver	Efficiency	Robustness	Efficiency	Robustness
`psifa`	33.33%	100.00%	95.24%	95.24%
`patternsearch`	47.62%	100.00%	4.76%	14.28%
`imfil` (enriched)	4.76%	90.48%	0.00%	38.10%
`imfil`	14.28%	42.86%	0.00%	4.76%

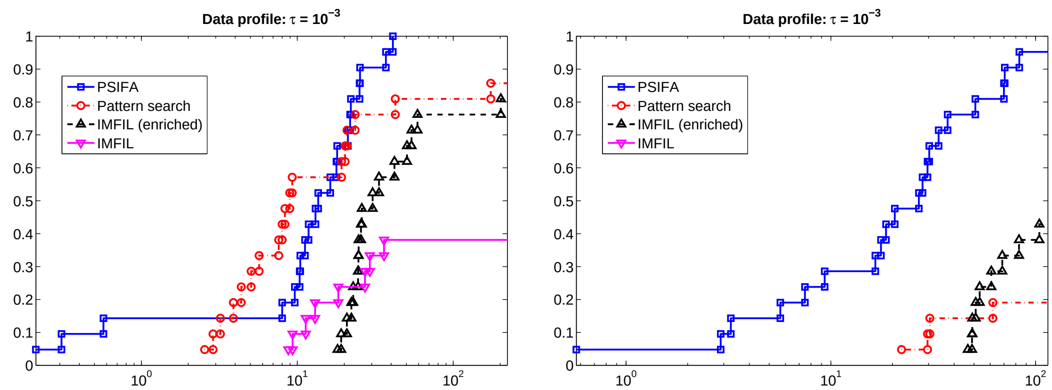

Figure 5: Data profiles of functional evaluations for problems with white noise and either the quadratic function (*left*) f_s^Q, given by (38), or the nonsmooth function (*right*) f_s^N, given by (39).

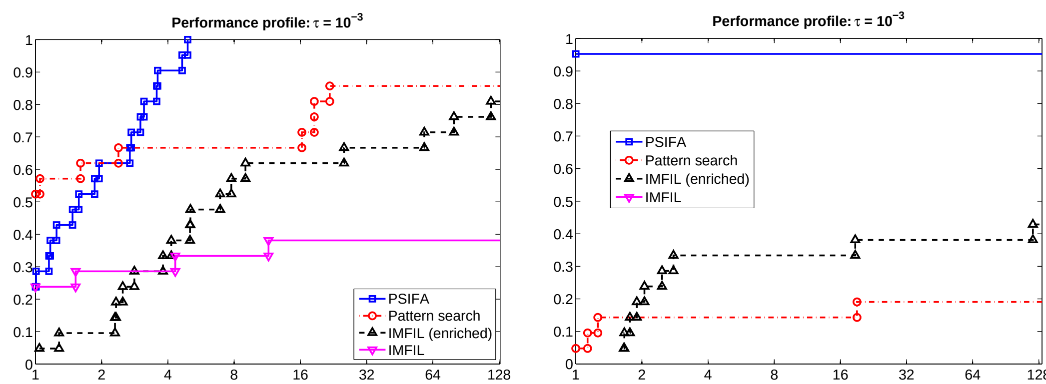

Figure 6: Performance profiles of functional evaluations for problems with white noise and either the quadratic function (*left*) f_s^Q, given by (38), or the nonsmooth function (*right*) f_s^N, given by (39).

6 Final remarks

We have proposed the algorithm PSIFA, that combines pattern search and implicit filtering ingredients for solving linearly constrained problems with noisy objective functions. Global convergence has been proved under assumptions upon the problem, including the hypothesis that the noise decays faster

Table 3: Performance of solvers for the 21 degenerate instances with white noise.

	Quadratic function		Nonsmooth function	
Solver	Efficiency	Robustness	Efficiency	Robustness
`psifa`	23.81%	100.00%	95.24%	95.24%
`patternsearch`	52.38%	85.71%	4.76%	19.05%
`imfil` (enriched)	0.00%	80.95%	0.00%	42.85%
`imfil`	23.81%	38.09%	0.00%	0.00%

than the step length as the generated sequence approaches stationarity. Our original convergence analysis rested upon a problem-based necessary optimality condition (Theorem 4.4), an auxiliary consequence of the hypotheses upon the problem (Proposition 4.3), specific features of our algorithm PSIFA (Propositions 4.6 and 4.8, Theorem 4.7 and Lemma 4.10), coming up with the global convergence result (Theorem 4.11). The outcomes obtained from numerical experiments with problems from the literature with a syntethic and artificial noise, and also with degenerate problems and white noise clearly show satisfactory performance of PSIFA when compared to Pattern Search, IMFIL and SID-PSM. These results were accomplished not only for problems that fulfill our assumptions, but also in situations that are not covered by the theory. We have also included experiments with an objective function that is not differentiable at the vertex of the polyhedral cone that constitutes both the feasible set of the problem, and the domain of the first component of the objective function. Such a behaviour indicates potential room for further improvements of the convergence results, in which the analysis of [6] and references therein might be helpful. We are currently working on a parameter identification problem with intrinsic noise, on the grounds of a case study from [20], for which PSIFA has shown a promising performance.

Acknowledgements

The authors are thankful to the anonymous reviewers, whose comments and concerns helped us to improve upon the preliminary versions of the manuscript.

Funding

This work was partially supported by Brazilian Funding Agencies *Fundação de Amparo à Pesquisa do Estado de São Paulo* - FAPESP (grants 2013/12964-4, 2013/05475-7, 2013/07375-0) and *Conselho Nacional de Desenvolvimento Científico e Tecnológico* - CNPq (grant 302915/2016-8).

References

[1] R. Andreani, A. Friedlander, and S. A. Santos. On the resolution of the generalized nonlinear complementarity problem. *SIAM Journal on Optimization*, 12(2):303–321, 2002.

[2] C. Audet and J. E. Dennis Jr. Analysis of generalized pattern searches. *SIAM Journal on Optimization*, 13(3):889–903, 2003.

[3] C. Audet and J. J. E. Dennis. A progressive barrier for derivative-free nonlinear programming. *SIAM Journal on Optimization*, 20(1):445–472, 2009.

[4] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Belmont, MA, USA, 1999.

[5] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2009.

[6] A. L. Custódio, J. E. Dennis Jr., and L. N. Vicente. Using simplex gradients of nonsmooth functions in direct search methods. *IMA Journal of Numerical Analysis*, 28(4):770–784, 2008.

[7] A. L. Custódio, H. Rocha, and L. N. Vicente. Incorporating minimum Frobenius norm models in direct search. *Computational Optimization and Applications*, 46(2):265–278, 2010.

[8] A. L. Custódio and L. N. Vicente. Using sampling and simplex derivatives in pattern search methods. *SIAM Journal on Optimization*, 18(2):537–555, 2007.

[9] R. De Leone, M. Gaudioso, and L. Grippo. Stopping criteria for linesearch methods without derivatives. *Mathematical Programming*, 30(3):285–300, 1984.

[10] M. A. Diniz-Ehrhardt, J. M. Martínez, and M. Raydan. A derivative-free nonmonotone line-search technique for unconstrained optimization. *Journal of Computational and Applied Mathematics*, 219(2):383–397, 2008. Special Issue dedicated to Prof. Claude Brezinski, on the occasion of his retirement from the University of Sciences and Technologies of Lille in June 2006.

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

[12] G. Fasano, G. Liuzzi, S. Lucidi, and F. Rinaldi. A linesearch-based derivative-free approach for nonsmooth constrained optimization. *SIAM Journal on Optimization*, 24(3):959–992, 2014.

[13] D. G. Ferreira. Sobre métodos de busca padrão para minimização de funções com restrições lineares (in Portuguese). Master's thesis, University of Campinas, Campinas (SP) Brazil, 2013.

[14] D. E. Finkel and C. T. Kelley. Convergence analysis of sampling methods for perturbed Lipschitz functions. *Pacific Journal of Optimization*, 5(2):339–350, 2009.

[15] R. Fletcher. *Practical Methods of Optimization.* John Wiley & Sons, Chichester, England, 1987.

[16] P. Gilmore and C. T. Kelley. An implicit filtering algorithm for optimization of functions with many local minima. *SIAM Journal on Optimization*, 5(2):269–285, 1995.

[17] W. Hock and K. Schittkowski. Test examples for nonlinear programming codes. In *Lecture Notes in Economics and Mathematical Systems*, volume 187. Springer, 1981.

[18] R. Hooke and T. A. Jeeves. Direct search solution of numerical and statistical problems. *SIAM Journal on Optimization*, 8(2):212–229, 1961.

[19] C. T. Kelley. *Iterative Methods for Optimization.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.

[20] C. T. Kelley. *Implicit Filtering.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2011.

[21] T. G. Kolda, R. M. Lewis, and V. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45(3):385–482, 2003.

[22] T. G. Kolda, R. M. Lewis, and V. Torczon. Stationarity results for generating set search for linearly constrained optimization. *SIAM Journal on Optimization*, 17(4):943–968, 2006.

[23] R. M. Lewis, A. Shepherd, and V. Torczon. Implementing generating set search methods for linearly constrained minimization. *SIAM Journal on Scientific Computing*, 29(6):2507–2530, 2007.

[24] R. M. Lewis and V. Torczon. Pattern search algorithms for bound constrained minimization. *SIAM Journal on Optimization*, 9(4):1082–1099, 1999.

[25] R. M. Lewis and V. Torczon. Pattern search methods for linearly constrained minimization. *SIAM Journal on Optimization*, 10(3):917–941, 2000.

[26] S. Lucidi and M. Sciandrone. A derivative-free algorithm for bound constrained optimization. *Computational Optimization and Applications*, 21(2):119–142, 2002.

[27] S. Lucidi, M. Sciandrone, and P. Tseng. Objective-derivative-free methods for constrained optimization. *Mathematical Programming*, 92(1):37–59, 2002.

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

[29] C. J. Price and I. D. Coope. Frames and grids in unconstrained and linearly constrained optimization: A nonsmooth approach. *SIAM Journal on Optimization*, 14(2):415–438, Feb. 2003.

[30] D. E. Stoneking, G. L. Bilbro, P. A. Gilmore, R. J. Trew, and C. T. Kelley. Yield optimization using a gaas process simulator coupled to a physical device model. *IEEE Transactions on Microwave Theory and Techniques*, 40(7):201–213, 1992.

[31] V. Torczon. On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, 7(1):1–25, 1997.