Optimization Letters
DOI 10.1007/s11590-008-0078-5

ORIGINAL PAPER

A derivative-free algorithm for systems of nonlinear inequalities

G. Liuzzi · S. Lucidi

Received: 13 February 2007 / Accepted: 4 February 2008
© Springer-Verlag 2008

Abstract Recently a new derivative-free algorithm has been proposed for the solution of linearly constrained finite minimax problems. This derivative-free algorithm is based on a smoothing technique that allows one to take into account the non-smoothness of the max function. In this paper, we investigate, both from a theoretical and computational point of view, the behavior of the minmax algorithm when used to solve systems of nonlinear inequalities when derivatives are unavailable. In particular, we show an interesting property of the algorithm, namely, under some mild conditions regarding the regularity of the functions defining the system, it is possible to prove that the algorithm locates a solution of the problem after a finite number of iterations. Furthermore, under a weaker regularity condition, it is possible to show that an accumulation point of the sequence generated by the algorithm exists which is a solution of the system. Moreover, we carried out numerical experimentation and comparison of the method against a standard pattern search minimization method. The obtained results confirm that the good theoretical properties of the method correspond to interesting numerical performance. Moreover, the algorithm compares favorably with a standard derivative-free method, and this seems to indicate that extending the smoothing technique to pattern search algorithms can be beneficial.

Keywords Derivative-free methods · Nonlinear inequality systems

G. Liuzzi (✉)
CNR - Consiglio Nazionale delle Ricerche,
IASI - Istituto di Analisi dei Sistemied Informatica “A. Ruberti”,
Viale Manzoni 30, 00185 Rome, Italy
e-mail: liuzzi@iasi.cnr.it

S. Lucidi
Dipartimento di Informatica e Sistemistica “Antonio Ruberti”,
Università di Roma “La Sapienza”, via Ariosto, 25, 00185 Rome, Italy
e-mail: lucidi@dis.uniroma1.it

1 Introduction

We consider the following system of linear and nonlinear inequalities

$$\begin{cases} g(x) \leq 0 \\ Ax \leq b, \end{cases} \tag{1}$$

where $x \in \mathbb{R}^n$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $A \in \mathbb{R}^{p\times n}$, $b \in \mathbb{R}^p$. We assume that the derivatives of the functions $g_i(x)$, $i = 1, \ldots, m$, can be neither calculated nor approximated explicitly. In recent years, many derivative-free algorithms have been proposed to solve various types of optimization problems. (See, for example, the survey paper [5].)

In particular, a derivative-free algorithm for the solution of linearly constrained finite minimax problems has been recently proposed in [7]. It has been shown that such an algorithm is able to produce a sequence of points that admits at least an accumulation point which is a stationary point of the non-smooth minimax problem.

In this paper, we reformulate the problem of finding a feasible solution for system (1) into the following minimax problem

$$\begin{array}{ll} \min & f(x) \\ s.t. & Ax \leq b. \end{array} \tag{2}$$

where $f(x)$ is defined by

$$f(x) = \max\{0, g_1(x), \ldots, g_m(x)\}. \tag{3}$$

Given Problem (2), we solve it by using the method proposed in [7].

The main theoretical result of the paper is that, under some mild regularity conditions on the functions defining system (1), it is possible to prove that the algorithm of [7] locates a solution of system (1) after a finite number of iterations. Furthermore, under a weaker regularity condition, it is possible to show that an accumulation point of the sequence generated by the algorithm exists which is a solution of system (1). Moreover, numerical experimentation has been carried out showing the possible efficiency of the method also in comparison with other approaches.

The paper is organized as follows. In Sect. 2, we briefly comment on some basic definitions and introduce the working assumptions used in the paper. In Sect. 3, we recall the approach proposed in [7], which consists of converting problem (2) into a smooth one by using a suitable smoothing function. In this section, we prove a new property of this smoothing function (Proposition 2), which plays a fundamental role in establishing the finite termination of the algorithm. In Sect. 4, we briefly describe the algorithm proposed in [7] and we prove that, when it is applied to the solution of systems (1), it shows new theoretical properties (Propositions 4 and 5). Finally, in Sect. 5 some numerical results are reported which confirm the good properties of the method.

Before concluding this section, we introduce some notation and background material. We denote by $\mathcal{D}$ the polyhedral set defined by the linear inequality constraints, namely

$$\mathcal{D} = \{x \in \mathbb{R}^n : \; Ax \le b\}, \quad \text{and} \quad \mathcal{F} = \{x \in \mathbb{R}^n : \; g(x) \le 0\} \cap \mathcal{D}$$

the feasible set defined by system (1). We denote by $a_j^\top$, $j = 1, \ldots, p$, the rows of matrix A. We introduce the function $g_0(x) = 0$ so that $f(x)$ defined in (3) can be rewritten as $\max_{i=0,\ldots,m}\{g_i(x)\}$.

The following index sets will be used in the remainder of the paper.

$$I_\pi(x) := \{i : g_i(x) \ge 0\}, \quad I_\nu(x) := \{i : g_i(x) < 0\}. \tag{4}$$

2 Definitions and assumptions

We begin this section by stating the following fundamental assumption regarding differentiability of the functions defining System (1) and imposing radial unboundedness of the max function $f(x)$.

Assumption 1 The functions $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are twice continuously differentiable functions on $\mathbb{R}^n$, and the function $f(x)$ is radially unbounded on the set $\mathcal{D}$; that is, for every sequence $\{x_k\} \subset \mathcal{D}$ satisfying $\lim_{k\to\infty} \|x_k\| = +\infty$,

$$\lim_{k\to\infty} f(x_k) = +\infty.$$

We require Assumption 1 to hold true throughout the paper and we note that it is exactly the same assumption used in [7]. It is needed to guarantee boundedness of the iterates.

By exploiting the particular structure of Problem (2), that is, the expression of the max function $f(x)$, it is possible to state an equivalent characterization of its stationary points.

Definition 1 A point $\bar{x} \in \mathcal{D}$ is a stationary point of Problem (2) if λ_i, $i \in B(\bar{x})$, exist such that

$$\lambda_i \ge 0, \quad i \in B(\bar{x}), \quad \sum_{i \in B(\bar{x})} \lambda_i = 1, \tag{5}$$

$$\left(\sum_{i \in B(\bar{x})} \lambda_i \nabla g_i(\bar{x}) \right)^\top d \ge 0, \quad \text{for all } d \in T(\bar{x}), \tag{6}$$

where

$$B(x) = \left\{ i \in \{0, 1, \ldots, m\} : g_i(x) = \max_{j=0,1,\ldots,m} \{g_j(x)\} \right\},$$

$$T(x) = \{d \in \mathbb{R}^n : a_j^\top d \le 0, \text{ for all } j \text{ s.t. } a_j^\top x = b_j\}.$$

$T(x)$, in particular, is the *cone of feasible directions* with respect to the linear inequality constraints.

Now we introduce some assumptions which are standard assumptions in a constrained context.

Assumption 2 At any point $x \in \mathcal{D}$, a vector $z \in T(x)$ exists such that $\nabla g_i(x)^\top z < 0$, for all i with $g_i(x) > 0$.

Assumption 3 At any point $x \in \mathcal{D}$, a vector $z \in T(x)$ exists such that $\nabla g_i(x)^\top z < 0$, for all i with $g_i(x) \geq 0$.

Assumptions 2 and 3 will be explicitly recalled when needed.

Provided that $\mathcal{D} \neq \phi$, Assumptions 2 and 3 both imply that the feasible region defined by system (1) is not empty, but the second one is stronger in that it implies that $\mathcal{F}$ has a non-empty interior. In the case of convex functions, Assumptions 2 and 3 are minimal requirements for the feasibility and the strict feasibility of system (1). These connections between Assumptions 2 and 3 and the feasibility of system (1) are described by the following proposition.

Proposition 1 *Assume that $\mathcal{D} \neq \phi$. Then*

(i) *if Assumption 2 holds, then the feasible set $\mathcal{F}$ is not empty*;
(ii) *if Assumption 3 holds, then a point $\tilde{x} \in \mathbb{R}^n$ exists such that $g(\tilde{x}) < 0$ and $A\tilde{x} \leq b$.*
Moreover, if functions $g_1, \ldots, g_m$ are convex, then
(iii) *the feasible set $\mathcal{F}$ is not empty only if Assumption 2 holds*;
(iv) *a point $\tilde{x} \in \mathbb{R}^n$ exists such that $g(\tilde{x}) < 0$ and $A\tilde{x} \leq b$, only if Assumption 3 holds.*

Proof The proofs of points (i) and (ii) follow from standard arguments and, hence, we omit them. For completeness' sake we report the details of these proofs in [6].

Let $\tilde{x} \in \mathcal{F}$ and x be any given point within $\mathcal{D}$. First we show that $(\tilde{x} - x) \in T(x)$. To this aim, let j be an index such that $a_j^\top x = b$. We can write

$$a_j^\top (\tilde{x} - x) = a_j^\top \tilde{x} - b \leq 0,$$

which, by taking into account that index j was arbitrary, means $(\tilde{x} - x) \in T(x)$.

Point (iii). By the convexity assumption we have that, for all $i = 1, \ldots, m$,

$$0 \geq g_i(\tilde{x}) \geq g_i(x) + \nabla g_i(x)^\top (\tilde{x} - x).$$

This implies that, if $g_i(x) > 0$, we must have $\nabla g_i(x)^\top (\tilde{x} - x) < 0$. Therefore, letting $z = \tilde{x} - x$, we obtain

$$\nabla g_i(x)^\top z < 0, \quad \text{for all } i \text{ s.t. } g_i(x) > 0.$$

Point (iv). To this aim, let $\tilde{x} \in \mathcal{F}$ be such that $g(\tilde{x}) < 0$. By the convexity assumption we have that, for all $i = 1, \dots, m$,

$$0 > g_i(\tilde{x}) \geq g_i(x) + \nabla g_i(x)^\top (\tilde{x} - x).$$

This implies that, if $g_i(x) \geq 0$, we must have $\nabla g_i(x)^\top (\tilde{x} - x) < 0$. Therefore, letting $z = \tilde{x} - x$, we obtain

$$\nabla g_i(x)^\top z < 0, \quad \text{for all } i \text{ s.t. } g_i(x) \geq 0,$$

thus completing the proof. □

We end this section by recalling the following lemma, which is a slight modification of Theorem 2.2 of Ref. [4] and which will be used in subsequent sections.

Lemma 1 *Let $\hat{x}$ be any given point in $\mathcal{D}$. If Assumption 3 holds, then an open neighborhood $\mathcal{B}(\hat{x}; \rho)$ of $\hat{x}$ and a direction $d \in T(\hat{x})$ exist such that, for all $i \in I_\pi(\hat{x})$, we have*

$$\nabla g_i(x)^\top d \leq -1, \quad \forall\, x \in \mathcal{B}(\hat{x}; \rho). \tag{7}$$

3 Smooth approximation and preliminary results

To handle the non differentiability of function $f(x)$, the following twice continuously differentiable smoothing function (described in [1, 12]) is used in [7]:

$$f(x; \mu) = \mu \ln\left(1 + \sum_{i=1}^{m} \exp\left(\frac{g_i(x)}{\mu}\right)\right) = \mu \ln \sum_{i=0}^{m} \exp\left(\frac{g_i(x)}{\mu}\right)$$

whose gradient is given by

$$\nabla f(x; \mu) = \sum_{i=0}^{m} \lambda_i(x; \mu) \nabla g_i(x), \tag{8}$$

where

$$\lambda_i(x; \mu) = \frac{\exp\left(\frac{g_i(x)}{\mu}\right)}{1 + \sum_{j=1}^{m} \exp\left(\frac{g_j(x)}{\mu}\right)} \in (0, 1), \quad i = 0, 1, \dots, m, \tag{9}$$

and $\sum_{i=0}^{m} \lambda_i(x; \mu) = 1$.

By using Lemma 1, it is possible to show that the minimization of the smoothing function $f(x; \mu)$ can have a strong connection with the problem of finding a point which satisfies the system of inequalities (1). In particular, the following proposition shows that, in every point $x \in \mathcal{D}$ which does not satisfy the system $g(x) \leq 0$, the smoothing function, with sufficiently small values of the parameter μ, has the directional derivative along a suitable feasible direction which is smaller than a negative fixed constant. This property will be exploited in the next section for showing the finite termination of the algorithm proposed in [7] when it is used for solving system (1).

Proposition 2 *Let $\hat{x}$ be any given point in $\mathcal{D}$. If Assumption 3 holds, then a direction $d \in T(\hat{x})$, a scalar $\sigma(\hat{x}) > 0$ and a scalar $\mu(\hat{x}) > 0$ exist such that for all $x \in \mathcal{B}(\hat{x}; \sigma(\hat{x})) \cap \mathcal{D}$ and $g(x) \not\leq 0$, $\mu \in (0, \mu(\hat{x})]$,*

$$\nabla f(x;\mu)^\top d \leq -\frac{1}{2(m+1)}.$$

Proof By Assumption 3, we have that the hypotheses of Lemma 1 are satisfied at $\hat{x}$ for $I = I_\pi(\hat{x})$. Let $\mathcal{B}(\hat{x}; \rho)$ and d be the neighborhood and the direction considered in Lemma 1. By continuity, we can find a neighborhood $\mathcal{B}(\hat{x}; \sigma(\hat{x})) \subseteq \mathcal{B}(\hat{x}; \rho)$ such that for $i \notin I_\pi(\hat{x})$ and $x \in \mathcal{B}(\hat{x}; \sigma(\hat{x}))$, we have $g_i(x) < 0$; it follows that $I_\pi(x) \subseteq I_\pi(\hat{x})$ and $I_\nu(\hat{x}) \subseteq I_\nu(x)$, for $x \in \mathcal{B}(\hat{x}; \sigma(\hat{x}))$.

Let now $x \in \mathcal{B}(\hat{x}; \sigma(\hat{x})) \cap \mathcal{D}$ be an infeasible point with respect to the nonlinear inequality constraints; then, there must exists at least an index $i \in I_\pi(x)$ such that $g_i(x) > 0$ which implies $I_\pi(x) \neq \phi$. By recalling expression (8), we can write

$$\nabla f(x;\mu)^\top d = \sum_{i\in I_\nu(\hat{x})} \lambda_i(x;\mu)\nabla g_i(x)^\top d + \sum_{i\in I_\pi(\hat{x})} \lambda_i(x;\mu)\nabla g_i(x)^\top d.$$

By Lemma 1 we have that $\nabla g_i(x)^\top d \leq -1$, $i \in I_\pi(\hat{x})$, so that we can write

$$\nabla f(x;\mu)^\top d \leq \sum_{i\in I_\nu(\hat{x})} \lambda_i(x;\mu)\nabla g_i(x)^\top d - \sum_{i\in I_\pi(\hat{x})} \lambda_i(x;\mu). \tag{10}$$

Let $\bar{\imath} \in I_\pi(\hat{x})$ be an index such that $g_{\bar{\imath}}(x) = \max_{i\in I_\pi(\hat{x})}\{g_i(x)\}$. Since

$$\frac{\exp(g_i(x)/\mu)}{\exp(g_{\bar{\imath}}(x)/\mu)} \leq 1, \quad \text{for all } i \in \{1,\ldots,m\},$$

then

$$\sum_{i\in I_\pi(\hat{x})} \lambda_i(x;\mu) \geq \lambda_{\bar{\imath}}(x;\mu) \geq \frac{1}{1+1+\sum_{i=1,i\neq\bar{\imath}}^m \frac{\exp(g_i(x)/\mu)}{\exp(g_{\bar{\imath}}(x)/\mu)}} \geq \frac{1}{1+m}. \tag{11}$$

By considering (10) and (11), we get

$$\nabla f(x;\mu)^\top d \leq \sum_{i\in I_\nu(\hat{x})} \lambda_i(x;\mu)\nabla g_i(x)^\top d - \frac{1}{1+m}. \tag{12}$$

Now, since $I_\nu(\hat{x}) \subseteq I_\nu(x)$, for $x \in \mathcal{B}(\hat{x}; \sigma(\hat{x}))$, by expression (9), it follows that, for any given $x \in \mathcal{B}(\hat{x}; \sigma(\hat{x})) \cap \mathcal{D}$ not feasible,

$$\lim_{\mu \to 0^+} \lambda_i(x; \mu) = 0, \quad i \in I_\nu(\hat{x}).$$

Hence, by the boundedness of $\nabla g_i(x)^\top d$ over $\mathcal{B}(\hat{x}; \sigma(\hat{x}))$, a $\mu(\hat{x}) > 0$ exists such that for all $\mu \in (0, \mu(\hat{x})]$ we have

$$\sum_{i \in I_\nu(\hat{x})} \lambda_i(x; \mu) \nabla g_i(x)^\top d < \frac{1}{2(m+1)}. \tag{13}$$

The result follows from (13) and (12). □

4 DF Algorithm and convergence properties

In this section we briefly recall the method proposed in [7] and prove its convergence properties in relation to the system of nonlinear inequalities (1). Following [7], we employ the smooth approximating problem

$$\min_{x \in \mathcal{D}} f(x; \mu), \tag{14}$$

which, as $\mu \to 0$, approximates the non-smooth Problem (2). The approximating parameter μ is adaptively reduced by the algorithm during the optimization process.

At this point, we briefly recall the main idea which is behind the derivative-free method. Specifically, the lack of first-order information can be overcome by sampling the smoothing function along a suitable set of search directions. By this we mean that, at each iteration, the set of search directions must satisfy the following assumption [8,9,11].

Assumption 4 Let $\{x_k\}$ be a sequence of points belonging to $\mathcal{D}$ and $\{D_k\}$ be a sequence of sets of search directions. Then, for all k, $D_k = \{d_k^i : \|d_k^i\| = 1, i = 1, \ldots, r_k\}$, and, for some constant $\bar{\nu} > 0$,

$$cone\{D_k \cap T(x_k; \nu)\} = T(x_k; \nu) \quad \forall \nu \in [0, \bar{\nu}], \tag{15}$$

where $T(x; \nu) = \{d \in \mathbb{R}^n : a_j^\top d \leq 0, \forall j \text{ s.t. } a_j^\top x \geq b_j - \nu\}$. Moreover, $\bigcup_{k=0}^{\infty} D_k$ is a finite set and r_k is bounded.

Assumption 4 guarantees that every set D_k contains a sufficiently rich set of directions to overcome the lack of first order derivatives. In particular, condition (15) implies that, at every iteration k, the set of directions D_k must contain the positive generators of the cones $T(x_k; \nu)$ for all $\nu \in [0, \bar{\nu}]$. Roughly speaking, if the point x_k is near the boundary of $\mathcal{D}$, then a subset of directions of D_k must conform to the local geometry of the boundary of $\mathcal{D}$ near x_k. On the contrary, if the point x_k is suitably in the interior of $\mathcal{D}$ then the set D_k must contain a positive spanning set for $\mathbb{R}^n$.

An example on how to compute sets of directions satisfying the above assumption along with a discussion on its meaning can be found in [9,11].

Here, for the sake of clarity, we report the algorithm proposed in [7] for the solution of linearly constrained finite minimax problems. For a detailed description and analysis of the method we refer the interested reader to [7].

DF Algorithm

Data. $x_0 \in \mathcal{D}$, $\mu_0 > 0$, $init\_step_0 > 0$, $\gamma > 0$, $\theta \in (0,1)$, $\nu > 0$, $k := 0$.
While ($g(x_k) \not\leq 0$) **then**
  Choose a set of directions $D_k = \{d_k^1, \ldots, d_k^{r_k}\}$ satisfying Assumption 4;
  $i := 1;\ y_k^i := x_k;\ \tilde{\alpha}_k^i := init\_step_k;$
  Repeat
    Compute the maximum steplength $\bar{\alpha}_k^i$ such that $A(y_k^i + \bar{\alpha}_k^i d_k^i) \leq b$;
    $\hat{\alpha}_k^i := \min\{\bar{\alpha}_k^i, \tilde{\alpha}_k^i\}$;

    If $\left(\hat{\alpha}_k^i > 0 \text{ and } f(y_k^i + \hat{\alpha}_k^i d_k^i; \mu_k) \leq f(y_k^i; \mu_k) - \gamma(\hat{\alpha}_k^i)^2\right)$
      compute α_k^i by the *Expansion Step*($\bar{\alpha}_k^i, \hat{\alpha}_k^i, y_k^i, d_k^i; \alpha_k^i$); and set
      $\tilde{\alpha}_k^{i+1} := \alpha_k^i$;
    Else $\alpha_k^i := 0$; $\tilde{\alpha}_k^{i+1} := \theta\tilde{\alpha}_k^i$; **End If**
    $y_k^{i+1} := y_k^i + \alpha_k^i d_k^i;\ i := i+1;$
  Until ($i > r_k$)
  Choose $x_{k+1} \in \mathcal{D}$ such that $f(x_{k+1}; \mu_k) \leq f(y_k^i; \mu_k)$;
  $init\_step_{k+1} := \tilde{\alpha}_k^i;\ \mu_{k+1} := \min\left\{\mu_k, \max_{i=1,\ldots,r_k} \{(\tilde{\alpha}_k^i)^{1/2}, (\alpha_k^i)^{1/2}\}\right\};$
  $k := k+1;$
End While

Expansion Step ($\bar{\alpha}_k^i, \hat{\alpha}_k^i, y_k^i, d_k^i; \alpha_k^i$).
Data. $\gamma > 0$, $\delta \in (0,1)$.
$\alpha := \hat{\alpha}_k^i$;
Repeat $\alpha_k^i := \alpha;\ \alpha := \min\{\bar{\alpha}_k^i, (\alpha/\delta)\};$
Until ($\alpha_k^i < \bar{\alpha}_k^i$ **and** $f(y_k^i + \alpha d_k^i; \mu_k) \leq f(y_k^i; \mu_k) - \gamma\alpha^2$)

As shown in [7], a significant role in carrying out the convergence analysis of Algorithm DF is played by the index set K defined as follows:

$$K = \{k : \mu_{k+1} < \mu_k\}. \tag{16}$$

Recalling the results described in [7] we can report the following proposition describing some asymptotic properties of the Algorithm DF.

Proposition 3 *Suppose that Algorithm DF does not stop after finitely many iterations on a point $\bar{x}$ satisfying system (1). Let $\{x_k\}$ be the sequence produced by the algorithm and let $\{x_k\}_K$ be the subsequence corresponding to the subset of indices K. Then,*

(i) *the sequence $\{x_k\}$ is bounded*;
(ii) $\lim_{k\to\infty} \mu_k = 0$;
(iii) *every accumulation point $\bar{x}$ of $\{x_k\}_K$ is a stationary point of Problem (2)*;
(iv) *for every set $\bar{K} \subseteq K$ such that $\lim_{k\to\infty, k\in\bar{K}} x_k = \bar{x}$ and for every $d \in T(\bar{x})$*

$$\lim_{k\to\infty, k\in\bar{K}} \nabla f(x_k; \mu_k)^\top d \geq 0.$$

Proof Point (i) follows from point (iii) of Proposition 5 of [7]. Point (i) of Proposition 6 of [7] and the updating rule of the parameter μ_k in Algorithm DF imply point (ii) of the proposition. Point (iii) directly follows from Corollary 1 of [7].

The proof of (iv) follows with minor modifications from the proofs of Theorem 1 and Proposition 4 of Ref. [7]. In fact, in Theorem 1, it is shown that the sequences produced by the algorithm satisfy the hypothesis of Proposition 4. Then, the proof of Proposition 4 is done by showing that

$$\lim_{k\to\infty, k\in\bar{K}} \nabla f(x_k; \mu_k)^\top d \geq 0$$

holds for every $d \in T(\bar{x})$. See, for instance, relations (21) and (29) of Ref. [7]. □

As said before, the previous proposition follows from the analysis performed in [7] concerning the behaviour of Algorithm DF in solving a general minimax problem. In what follows, we study the case where Algorithm DF is used for finding a point which satisfies system (1), in other words the case where Algorithm DF is used for solving Problem (2), which is a particular minimax problem deriving from system (1).

A first result can be obtained by combining point (iii) of Proposition 3 and Assumption 2. In fact, this assumption guarantees that the accumulation points of the sequence $\{x_k\}_K$ produced by Algorithm DF are solutions of system (1).

Proposition 4 *Under Assumption 2, Algorithm DF either*

(i) *stops after finitely many iterations on a point $\bar{x}$ satisfying system (1) or*
(ii) *it produces an infinite sequence of points $\{x_k\}$ such that every limit point $\bar{x}$ of the subsequence $\{x_k\}_K$, where K is defined by (16), satisfies system (1).*

Proof If an index $\bar{k}$ exists such that $g(x_{\bar{k}}) \leq 0$ then Algorithm DF returns the feasible point $\bar{x} = x_{\bar{k}}$. If this is not the case, then an infinite sequence $\{x_k\}$ is produced and we can apply point (iii) of Proposition 3 to conclude that every limit point of the subsequence $\{x_k\}_K$ is stationary for Problem (2). Therefore, from Proposition 1 we know that λ_i, $i \in B(\bar{x})$, exist such that (5) and (6) hold.

Let us suppose by contradiction, that $\bar{x}$ is not feasible, that is, at least one index $i \in \{1, \ldots, m\}$ exists such that $g_i(\bar{x}) > 0$. By Assumption 2, we know that $z \in T(\bar{x})$ exists such that

$$\nabla g_i(\bar{x})^\top z < 0, \quad \text{for all } i \text{ s.t. } g_i(\bar{x}) > 0.$$

Now, by considering that $B(\bar{x}) \subseteq \{i : g_i(\bar{x}) > 0\}$, the above relation along with the non-negativity of the multipliers λ_i yields

$$\left(\sum_{i\in B(\bar{x})} \lambda_i \nabla g_i(\bar{x})\right)^\top z < 0, \tag{17}$$

which contradicts (6). □

A quite stronger result can be stated under Assumption 3. In fact, this assumption allows us to recall Proposition 2 and to state the following proposition.

Proposition 5 *Under Assumption 3, Algorithm DF stops after finitely many iterations at a point $\bar{x}$ which satisfies system (1).*

Proof Reasoning by contradiction, let us suppose that an infinite sequence $\{x_k\}$ is produced such that

$$g(x_k) \not\leq 0, \tag{18}$$

for all integers k.

First, Point (i) of Proposition 3 ensures that the sequence $\{x_k\}$ is bounded. Then, let $\bar{x}$ be an accumulation point of the subsequence $\{x_k\}_K$. Therefore an index set $\bar{K} \subseteq K$ exists such that

$$\lim_{k\to\infty, k\in\bar{K}} x_k = \hat{x}. \tag{19}$$

If $\sigma(\hat{x})$ and $\mu(\hat{x})$ are the scalars introduced in Proposition 2, then (18), (19) and point (ii) of Proposition 3 imply that there exists an index $\hat{k} \in \bar{K}$ such that, for all $k \geq \hat{k}$, $k \in \bar{K}$, we have

$$g(x_k) \not\leq 0, \quad x_k \in \mathcal{B}(\hat{x}; \sigma(\hat{x})) \cap \mathcal{D}, \;\; \mu_k \in (0, \mu(\hat{x})].$$

Hence, by Proposition 2, we have that a direction $\hat{d} \in T(\hat{x})$ exist such that

$$\nabla f(x_k; \mu_k)^\top \hat{d} \leq -\frac{1}{2(m+1)}, \tag{20}$$

for all $k \geq \hat{k}$, $k \in \bar{K}$.

Now, relation (20) contradicts point (iv) of Proposition 3 thus completing the proof. □

Assumption 2 of Proposition 4 and Assumption 3 of Proposition 5 can be replaced by the convexity of the functions $g_i(x)$, $i = 1, \ldots, m$. In particular, point (iii) of Propositions 1 and 4 imply the following corollary.

Corollary 1 *Assume that functions $g_1, \ldots, g_m$ are convex. Then, Algorithm DF either stops after finitely many iterations at a solution for system (1), or it produces an infinite sequence of points $\{x_k\}$ such that every limit point of the subsequence $\{x_k\}_K$, where K*

is defined by (16)*, is a solution for system* (1) *if and only if the feasible set* $\mathcal{F}$ *is not empty.*

Then, point (iv) of Propositions 1 and 5 yield the following corollary.

Corollary 2 *Assume that functions* $g_1, \dots, g_m$ *are convex and that a point* $\tilde{x} \in \mathbb{R}^n$ *exists such that* $g(\tilde{x}) < 0$ *and* $A\tilde{x} \leq b$*. Then, Algorithm DF stops after finitely many iterations on a point* $\bar{x}$ *which solves system* (1).

5 Numerical experience

In this section we report some preliminary numerical results obtained by Algorithm DF. Far from being an extensive testing, the computational experience is principally intended to show that, besides the interesting theoretical properties, Algorithm DF is reliable. Algorithm DF is implemented in double precision Fortran 90 and run on a computer with an Intel 3.2 GHz Pentium IV processor with 1GB RAM.

The test problems were all selected from the well-known CUTEr [3] (Constrained and unconstrained testing environment, revisited) test set. Among all the CUTEr test problems, we selected 83 problems according to the following criteria.

- The provided initial point is (obviously) infeasible but it must satisfy the linear constraints (if any). This because Algorithm DF handles directly the linear constraints by explicitly forcing feasibility of all iterates with respect to them. To this end, the initial point x_0 is required to belong to the set $\mathcal{D}$.
- The problem should have at least one nonlinear inequality constraint. Indeed, many problems of the CUTEr collection have only nonlinear equality constraints, so that finding a feasible point amounts to solving a system of nonlinear equalities. This latter problems have long been studied in the literature, and some very efficient and reliable methods exist for their solution.
- The number of variables should not exceed 200 for fixed dimension problems and, more or less, 50 for variably dimensioned problems. The rationale for this particular choice has been that of selecting a consistent number of problems but, at the same time, of guaranteeing that many of them have a dimension suitable for a derivative-free code.

In the actual implementation of the code, we slightly modified the version used in the paper [7] by introducing the stopping condition on the feasibility of the current iterate $g(x_k) \leq 0$ and tightening the tolerance on the step length. In particular, we use as an alternate stopping condition $init\_step_k \leq 10^{-5}$ in the linearly constrained case, and $\max_{i=1,\dots,n} \tilde{\alpha}_k^i \leq 10^{-5}$ in the unconstrained case. We also impose a limit on the required number of function evaluations which must not exceed 2×10^5.

Let P be the set of selected test problems. For every $p \in P$, let n_p, m_p denote, respectively, the number of variables and the number of nonlinear constraints. We define

$$D_t(\tau) = |\{p \in P : t_p \leq \tau\}|,$$

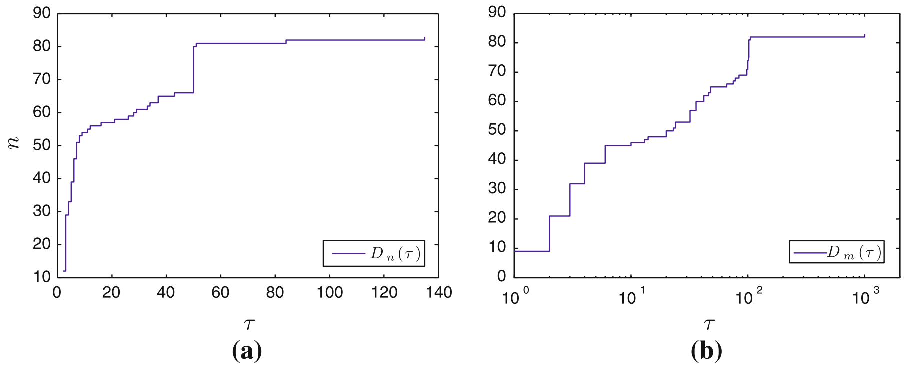

Fig. 1 Test set composition

with $t_p = n_p, m_P, \epsilon_p$. Functions $D_n(\tau)$ and $D_m(\tau)$, plotted in Fig. 1, tells us, for every choice of $\tau > 0$, how many problems have $n_p \leq \tau$ variables and $m_p \leq \tau$ nonlinear constraints, respectively, thus giving us cumulative information about the composition of the test set.

As regards the initial feasibility violation ϵ^0, we point out that all the problems have $\epsilon^0 > 8 \times 10^{-2}$. Moreover, 10 problems have $8 \times 10^{-2}\epsilon^0 \leq 1$ and almost half of the test problems have $\epsilon^0 \geq 10$.

Since many of the selected problems have also equality constraints, we briefly describe how the proposed approach can be adapted to handle these constraints. Consider the following system of equalities and inequalities.

$$\begin{cases} g(x) \leq 0, & h(x) = 0 \\ \hat{A}x \leq \hat{b}, & \tilde{A}x = \tilde{b}, \end{cases} \tag{21}$$

where, again, $x \in \mathbb{R}^n$, $\hat{b} \in \mathbb{R}^{\hat{p}}$, $\tilde{b} \in \mathbb{R}^{\tilde{p}}$, $\hat{A} \in \mathbb{R}^{\hat{p}\times n}$, $\tilde{A} \in \mathbb{R}^{\tilde{p}\times n}$ and $g : \mathbb{R}^n \to \mathbb{R}^{\hat{m}}$, $h : \mathbb{R}^n \to \mathbb{R}^{\tilde{m}}$. The function $f(x)$ becomes

$$f(x) = \max\{0, g_1(x), \ldots, g_{\hat{m}}(x), h_1(x), -h_1(x), \ldots, h_{\tilde{m}}(x), -h_{\tilde{m}}(x)\}.$$

The results obtained by applying Algorithm DF to the solution of the selected problems show a good overall behavior of the method and seem to point out that the algorithm can have also a practical interest beside the theoretical one. Indeed, Algorithm DF produces a point $x^\star$ with a feasibility violation less than 10^{-5} on 79 problems. Further, 69 problems are solved in less than 1,000 function evaluations and 51 in less than 100 function evaluations.

Furthermore, we compared Algorithm DF with another freely available derivative-free algorithm. In particular, we downloaded a general purpose minimization pattern search method from [10]. To solve Problem (1) using this pattern search method,we employed two reformulations of the problem by means of the following two objective functions,

$$f_{\max}\ (x) = \max\{0, g_1(x), \ldots, g_m(x), a_1^\top x - b_1, \ldots, a_p^\top x - b_p\},$$
$$f_{\max^2}(x) = \sum_{i=1}^{m} \max\{0, g_i(x)\}^2 + \sum_{j=1}^{p} \max\{0, a_j^\top x - b_j\}^2.$$

We will refer to the pattern search method using $f_{\max}$ as Algorithm $P_{\max}$ and to the pattern search method using $f_{\max^2}$ as Algorithm $P_{\max^2}$.

Moreover, since both Algorithms $P_{\max}$ and $P_{\max^2}$ do not handle explicitly the linear constraints, we also consider a modified version of Algorithm DF, which simply makes no distinction between linear and nonlinear constraints, thus considering the linear constraints implicitly. We will refer to this latter algorithm as Algorithm $\mathrm{DF}_{\mathrm{imp}}$.

Now, let S be the set of competing algorithms and denote by $t_{p,s}$ a performance measure of solver s when applied to the solution of problem p. Then, define [2]

$$D_t^s(\tau) = |\{p \in P : t_{p,s} \le \tau\}|.$$

As performance measures, we use the feasibility violation $\epsilon^\star$ on the solution point $x^\star$ (see Fig. 2a) and the number of function evaluations $\phi^\star$ required to get convergence (see Fig. 2b).

Figure 2 shows that Algorithm $\mathrm{DF}_{\mathrm{imp}}$ performs better than algorithms $P_{\max}$ and $P_{\max^2}$ both in terms of attained feasibility and required number of function evaluations. This can be partly due to the fact that the two methods $P_{\max}$ and $P_{\max^2}$ are based on an off-the-shelf pattern search algorithm which has been used without any tuning of the parameters. However, this can also indicate that it could be advantageous to extend the considered smoothing strategy in the context of pattern search methods. Furthermore, Fig. 2 shows that algorithm $\mathrm{DF}_{\mathrm{imp}}$ is outperformed by DF, and this clearly points out the beneficial effects of an explicit handling of the linear constraints.

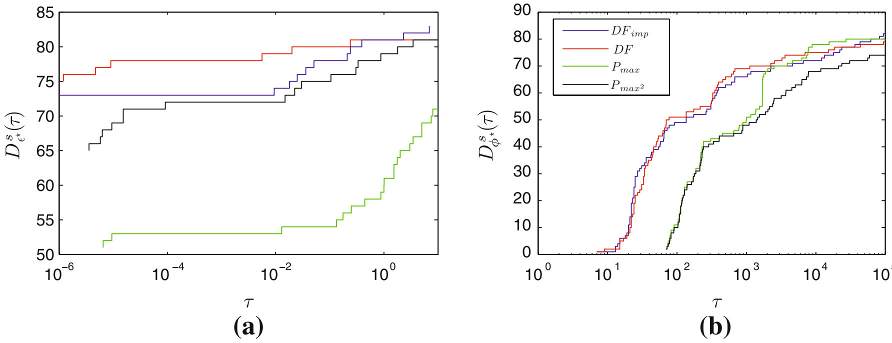

Fig. 2 Comparison between DF, $\mathrm{DF}_{\mathrm{imp}}$, $P_{\max}$ and $P_{\max^2}$

6 Conclusions

We have proved that the algorithm proposed in [7] for finite minimax problems has some additional properties when applied to the solution of systems of nonlinear inequalities when derivatives are not available. In particular, under suitable assumptions we have shown that Algorithm DF is able to locate a solution of system (1) after a finite number of iterations. The reported numerical results seem to show on the one hand that the use of a smoothing technique is an effective tool for the solution of system on nonlinear inequalities. On the other hand, the results also indicate the positive contribution deriving from an explicit handling of the linear constraints, when present. In conclusion, Algorithm DF can be considered a promising tool for the solution of systems of nonlinear inequalities.

References

1. Bertsekas, D.P.: Constrained Optimization and Lagrange Multipliers Methods. Academic Press, New York (1982)
2. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program **91**, 201–213 (2002)
3. Gould, N.I.M., Orban, D., Toint, Ph.L.: CUTEr and SifDec: a constrained and unconstrained testing environment, revisited. ACM T. Math. Softw. **29**, 373–394 (2003)
4. Han, S.P., Mangasarian, O.L.: Exact penalty functions in nonlinear programming. Math. Program **17**, 251–269 (1979)
5. Kolda, T.G., Lewis, R.M., Torczon, V.: Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods. SIAM Rev. **45**, 385–482 (2003)
6. Liuzzi, G., Lucidi, S.: A derivative-free algorithm for systems of nonlinear inequalities. TR 17/05, Dept. of Computer and Systems Science, "Sapienza" University of Rome (2006)
7. Liuzzi, G., Lucidi, S., Sciandrone, M.: A derivative-free algorithm for linearly constrained finite minimax problems. SIAM J. Optim. **16**, 1054–1075 (2006)
8. Lucidi, S., Sciandrone, M., Tseng, P.: Objective-derivative-free methods for constrained optimization. Math. Program Ser. A **92**, 37–59 (2002)
9. May, J.H.: Linearly constrained nonlinear programming: a solution method that does not require analytic derivatives. Ph.D. Thesis. Yale University (1974)
10. Torczon, V.: http://www.cs.wm.edu/~va/software/
11. Lewis, R.M., Torczon, V.: Pattern search methods for linearly constrained minimization. SIAM J. Optim. **10**, 917–941 (2000)
12. Xu, S.: Smoothing method for minimax problems. Comput. Optim. Appl. **20**, 267–279 (2001)