ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
"Antonio Ruberti"
CONSIGLIO NAZIONALE DELLE RICERCHE

**G. Liuzzi, S. Lucidi, M. Sciandrone**

# A DERIVATIVE-FREE ALGORITHM FOR LINEARLY CONSTRAINED FINITE MINIMAX PROBLEMS

**R. 607 2004**

**G. Liuzzi** – Dipartimento di Informatica e Sistemistica "A. Ruberti" - Universitá di Roma "La Sapienza", Via Buonarroti 12, 00185 Rome, Italy.

**S. Lucidi** – Dipartimento di Informatica e Sistemistica "A. Ruberti" - Universitá di Roma "La Sapienza", Via Buonarroti 12, 00185 Rome, Italy.

**M. Sciandrone** – Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti" (IASI) - CNR, Viale Manzoni 30, 00185 Rome, Italy.

ISSN: 1128–3378

Collana dei Rapporti
Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti"
Consiglio Nazionale delle Ricerche

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: iasi@iasi.cnr.it
URL: http://www.iasi.cnr.it

## Abstract

In this paper we propose a new derivative-free algorithm for linearly constrained finite minimax problems. As it is well-known, standard derivative-free algorithms manage to locate points which only satisfy weak necessary optimality conditions for such a class of nonsmooth problems. In this work we define a new derivative-free algorithm globally converging toward standard stationary points for the finite minimax problem. To this end, we convert the original problem into a smooth one by using a smoothing technique based on the exponential penalty function of Kort and Bertsekas. This technique depends on a smoothing parameter which controls the approximation to the finite minimax problem. The proposed method is based on a sampling of the smooth function along a suitable search direction and on a particular updating rule for the smoothing parameter depending on the sampling stepsize. We show also that the proposed approach can be used for defining derivative-free algorithms for general constrained minimization problems. Finally, we report numerical results on a set of standard test problems.

*Key words:* Derivative-free optimization, linearly constrained finite minimax problems, nonsmooth optimization.

*AMS subject classifications:* 65Y05, 65K05, 90C56

## 1. Introduction

Many problems of interest in real world applications can be modelled as finite minimax problems. This class of problems arises, for instance, in the solution of approximation problems, systems of nonlinear equation, nonlinear programming problems and multi-objective problems. Many algorithms have been developed for the solution of finite minimax problems which require the knowledge of first or second order derivatives of the functions involved in the definition of the problem. Unfortunately, in some engineering applications, like some of those arising in optimal design problems, the function values are obtained by direct measurements (which are often affected by numerical error or random noise) or are the result of complex simulation programs so that first order derivatives cannot be explicitly calculated or approximated. Moreover, the nonsmoothness of the minimax problem does not allow us to employ some off-the-shelf derivative free method to solve the problem. These latter methods are indeed based on a well-established convergence theory which, in order to guarantee convergence towards a stationary point, requires first order derivatives to be continuous even though they cannot be computed. In particular, if the continuity assumption on the derivatives is relaxed, it is no more possible to prove global convergence of the derivative free method to a stationary point but it is only possible to prove convergence towards a point where the (Clarke) generalized directional derivative is non negative with respect to every search direction. In appendix a more detailed discussion on this subject is reported.

In this paper we consider a particular class of nonsmooth problems, namely, the problem of minimizing the maximum among a finite number of smooth functions. Problems of this class have the valuable feature that they can be approximated by a smooth problem. This smooth approximation of the minimax problem can be achieved by using different techniques (see [10], [11], [12], [13], [15], [18], [19], [20], [21]). In particular, we consider an approximation approach based on a so-called smoothing function which depends on a precision parameter (see [2], [4] and [5]). The definition of a solution method based on a smoothing technique requires a twofold care. From a computational point of view, a trade-off should be found between the goodness of the approximation and the problem of limiting the ill-conditioning due to the nonsmoothness of the minimax problem at the solutions. From a theoretical point of view, it should be guaranteed the convergence of the minimization algorithm towards stationary points of the original minimax problem. In particular, in [4] a class of algorithms for the solution of the minimax problem has been proposed which takes into account the above two requirements. This is accomplished by using a feedback precision-adjustment rule which updates the precision parameter during the optimization process of the smoothing function. Roughly speaking, the idea behind the proposed updating rule is that of updating the parameter only when the minimization method has carried out a significant improvement. However, these updating rules are based upon the knowledge of the first derivatives of the problem.

In this paper we propose a derivative-free method which is based on a sampling of the smooth

function along suitable search directions and on a particular updating rule for the smoothing parameter depending on the sampling stepsize. In this way, we manage both to prove convergence of the method to a stationary point of the minimax problem and to reduce the negative effects of the ill-conditioning of the smoothing approach.
In Section 2, we describe the minimax problem, its properties and the smoothing function. In Section 3, we report some convergence result for a general derivative-free approach to solve the minimax problem. In Section 4, we report the proposed derivative-free algorithm and its convergence analysis. Finally, Section 5 is devoted to some results of our method.

## 2. Problem definition and smooth approximation

In this paper we consider the solution of finite minimax problems where the variables are subject to linear inequality constraints. In particular, we consider problems of the following form

$$\begin{aligned} &\min \quad f(x), \\ &s.t. \quad Ax \le b, \end{aligned} \tag{1}$$

where $x \in \Re^n$, $b \in R^m$, $A \in R^{m\times n}$ and

$$f(x) = \max_{1\le i\le q} f_i(x).$$

We require the following assumption to hold:

**Assumption 1.** *$f_i : R^n \to R$, $i = 1, \dots, q$ are twice continuously differentiable functions on $R^n$ and the function $f(x)$ is radially unbounded on the feasible set $\mathcal{F}$.*

Note that, even though every function $f_i(x)$, $i = 1, \dots, q$, is twice continuously differentiable, we assume that their gradients can be neither calculated nor approximated explicitly.
We denote by $B(x)$ the following set of indices:

$$B(x) = \{i = 1, \dots, q : f_i(x) = f(x)\}. \tag{2}$$

Furthermore, we indicate by $\mathcal{F}$ the *feasible set* of Problem (1), namely:

$$\mathcal{F} = \{x \in R^n : Ax \le b\}.$$

For every feasible point $x \in \mathcal{F}$, we define the *set of indices of active constraints*:

$$I(x) \quad = \quad \{j = 1, \dots, m : a_j^T x = b_i\}, \tag{3}$$

and the *cone of feasible directions*

$$T(x) = \{d \in R^n : a_j^T d \le 0, \; j \in I(x)\}. \tag{4}$$

The directional derivative of the max function $f(x)$ is given by (see, e.g. [2])

$$Df(x, d) = \max_{i\in B(x)} \{\nabla f_i(x)^T d\}.$$

We define $\bar{x} \in \mathcal{F}$ a stationary point of Problem (1) if

$$Df(\bar{x}, d) \geq 0, \qquad \text{for all } d \in T(\bar{x}). \tag{5}$$

In particular, the following proposition shows a different characterization of the stationary points of Problem (1).

**Proposition 2.1.** *A point $\hat{x} \in \mathcal{F}$ is a stationary point of Problem (1) if and only if $\lambda_i$, $i \in B(\hat{x})$ exist such that*

$$\lambda_i \geq 0, \qquad i \in B(\hat{x}), \qquad \sum_{i \in B(\hat{x})} \lambda_i = 1, \tag{6}$$

$$\left( \textstyle\sum_{i \in B(\hat{x})} \lambda_i \nabla f_i(\hat{x}) \right)^T d \geq 0, \qquad \textit{for all} \quad d \in T(\hat{x}). \tag{7}$$

**Proof.** If $\hat{x} \in \mathcal{F}$ is a stationary point of Problem (1) then there exists at least an index $j \in B(\hat{x})$ such that $\nabla f_j(x)^T d \geq 0$. Then conditions (6) and (7) hold with $\lambda_j = 1$ and $\lambda_i = 0$ for all $i \neq j$. If $\hat{x} \in \mathcal{F}$ satisfies conditions (6) and (7) then we can write:

$$0 \leq \left( \sum_{i \in B(\hat{x})} \lambda_i \nabla f_i(\hat{x}) \right)^T d \leq q \max_{i \in B(\hat{x})} \nabla f_i(\hat{x})^T d,$$

which shows that $\bar{x}$ is a stationary point of Problem (1). □

In order to find a stationary point of Problem (1) we adopt a smoothing technique ([2], [5], [4], [10]) which consists in solving a sequence of smooth problems approximating the minimax problem in the limit. Let $\mu > 0$ be a smoothing parameter and define

$$f(x, \mu) = \mu \ln \sum_{i=1}^{m} \exp\left( \frac{f_i(x)}{\mu} \right),$$

which is sometimes referred to as an exponential penalty function [2]. An alternative expression for $f(x, \mu)$ is given by

$$f(x, \mu) = f(x) + \mu \ln \sum_{i=1}^{m} \exp\left( \frac{f_i(x) - f(x)}{\mu} \right).$$

We report some properties of $f(x; \mu)$ [10].

**Proposition 2.2.**

(i) $f(x, \mu)$ *is increasing with respect to* $\mu$*, and*

$$f(x) \leq f(x, \mu) \leq f(x) + \mu \ln m; \tag{8}$$

(ii) $f(x,\mu)$ *is twice continuously differentiable for all* $\mu > 0$, *and*

$$\nabla_x f(x,\mu) = \sum_{i=1}^{m} \lambda_i(x,\mu)\nabla f_i(x), \tag{9}$$

$$\nabla_x^2 f(x,\mu) = \sum_{i=1}^{m}\left(\lambda_i(x,\mu)\nabla^2 f_i(x) + \frac{1}{\mu}\lambda_i(x,\mu)\nabla f_i(x)\nabla f_i(x)^T\right) - \frac{1}{\mu}\left(\sum_{i=1}^{m}\lambda_i(x,\mu)\nabla f_i(x)\right)\left(\sum_{i=1}^{m}\lambda_i(x,\mu)\nabla f_i(x)\right)^T, \tag{10}$$

*where*

$$\lambda_i(x,\mu) = \frac{\exp(f_i(x)/\mu)}{\sum_{j=1}^{m}\exp(f_j(x)/\mu)} \in (0,1), \qquad \sum_{i=1}^{m}\lambda_i(x,\mu) = 1. \tag{11}$$

## 3. Derivative-free convergence conditions

A derivative free algorithm for Problem (1) has to take into account two different difficulties. The first one is due to the nonsmoothness of Problem (1). The second difficulty is due to the fact that stationary points of Problem (1), as stated by (5) and Proposition 2.1, are characterized by first order derivatives of the component functions $f_i(x)$, $i = 1,\dots,q$ which are not available. In order to take into account the nonsmoothness of Problem (1), we employ the following smooth approximating problem

$$\min_{x\in\mathcal{F}} f(x,\mu), \tag{12}$$

where the approximating parameter $\mu$ will be adaptively reduced during the optimization process.
In order tackle the second difficulty, that is the lack of first derivatives, we try to get first order information by sampling the objective function along a suitable set of search directions. Hence, to solve Problem (12), we follow the approach proposed in [8] which is based on a suitable sampling of the objective function along a set of search directions which positively span a "consistent" approximation of the cone of feasible directions (see (4)).

The first step toward the definition of a derivative-free method is to associate a suitable set of search directions with each point $x_k$ produced by the algorithm. This set of directions should have the property that the local behavior of the objective function along them provides sufficient information to overcome the lack of the gradient. Before stating this property, we formally define the set of indices of $\epsilon$-active constraints and the cone of first order $\epsilon$-feasible variations.

**Definition 3.1.**

$$I(x;\epsilon) = \{j : a_j^T x \ge b_j - \epsilon\}.$$

$$T(x;\epsilon) = \{d \in R^n : a_j^T d \le 0, \forall j \in I(x;\epsilon)\}.$$

As concerns the above sets $I(x;\epsilon)$ and $T(x;\epsilon)$ we can state the following Proposition ([8]).

**Proposition 3.2.** *Let $\{x_k\}$ be a sequence of iterates converging towards a point $\bar{x} \in \mathcal{F}$. Then, there exists a value $\epsilon^* > 0$ (depending on $\bar{x}$ only) such that for every $\epsilon \in (0, \epsilon^*]$ there exists $\bar{k}_\epsilon$ such that*

$$I(x_k; \epsilon) = I(\bar{x}), \tag{13}$$
$$T(x_k; \epsilon) = T(\bar{x}), \tag{14}$$

*for all $k \geq \bar{k}_\epsilon$.*

**Proof.** See the proof of Proposition 1 in [8]. □

Now we can formally introduce the assumption which the sets of search directions are required to satisfy.

**Assumption 2.** *Given $x_k$. The set of search directions*

$$D_k = \{d_k^i, \quad i = 1, \ldots, r_k\}, \qquad \text{with} \qquad \|d_k^i\| = 1,$$

*satisfies (for some constant $\bar{\epsilon} > 0$): $\bigcup D_k$ is a finite set, $r_k$ is uniformly bounded and*

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

The proposition which follows states a general convergence result. In particular, it points out the minimal requirements on the sampling of the smoothing function along the directions $d_k^i$, $i = 1, \ldots, r_k$ and on the updating of the smoothing parameter which are able to guarantee global convergence of the method towards a stationary point of the original minimax problem (1).

**Proposition 3.3.** *Let $\{x_k\}$ be a sequence of feasible points and $\bar{x}$ be a limit point of a subsequence $\{x_k\}_K$ for some infinite set $K \subseteq \{0, 1, \dots\}$. Let $\{D_k\}$, with $D_k = \{d_k^1, \dots, d_k^{r_k}\}$, be a sequence of sets of directions which satisfy Assumption 2 and $J_k = \{i \in \{1, \dots, r_k\} : d_k^i \in T(x_k, \epsilon)\}$ with $\epsilon \in (0, \min\{\bar{\epsilon}, \epsilon^\star\}]$ (where $\bar{\epsilon}$ and $\epsilon^\star$ are defined in Assumption 2 and Proposition 3.2, respectively).*
*Suppose that the following conditions hold:*

*(a) for each $k \in K$ and $i \in J_k$, there exist $y_k^i$ and scalars $\xi_k^i > 0$ such that:*

$$y_k^i + \xi_k^i d_k^i \in \mathcal{F} \tag{15}$$

$$f(y_k^i + \xi_k^i d_k^i, \mu_k) \geq f(y_k^i, \mu_k) - o(\xi_k^i); \tag{16}$$

*(b) and, furthermore,*

$$\lim_{k \to \infty, k \in K} \mu_k = 0; \tag{17}$$

$$\lim_{k \to \infty, k \in K} \frac{\max_{i \in J_k} \{\xi_k^i, \|x_k - y_k^i\|\}}{\mu_k} = 0. \tag{18}$$

*Then, $\bar{x}$ is a stationary point of the minimax problem (1).*

**Proof.** Recalling condition (16), for all $k \in K$, we have

$$f(y_k^i + \xi_k^i d_k^i, \mu_k) - f(y_k^i, \mu_k) \geq -o(\xi_k^i), \quad i \in J_k.$$

By the Mean-Value Theorem, we can write

$$f(y_k^i + \xi_k^i d_k^i, \mu_k) - f(y_k^i, \mu_k) = \xi_k^i \nabla_x f(u_k^i, \mu_k)^T d_k^i, \qquad i \in J_k, \tag{19}$$

where $u_k^i = y_k^i + t_k^i \xi_k^i d_k^i$, with $t_k^i \in (0, 1)$. By using again the Mean-Value Theorem, we can write

$$\xi_k^i \nabla_x f(u_k^i, \mu_k)^T d_k^i = \xi_k^i \nabla_x f(x_k, \mu_k)^T d_k^i + \xi_k^i (u_k^i - x_k)^T \nabla_x^2 f(\tilde{u}_k^i, \mu_k) d_k^i,$$

where $\tilde{u}_k^i = x_k + \tilde{t}_k^i (u_k^i - x_k)$, with $\tilde{t}_k^i \in (0, 1)$.
Recalling (9), (10), (11), Assumption 2 and the fact that $\{x_k\}_K$ is converging, we have that $\{x_k\}_K$, $\{\lambda_j(\tilde{u}_k^i, \mu_k)\}$, $\{d_k^i\}$ for all $i, j$ are bounded sequences. Therefore, we can find constants $c_1$ and $c_2$ such that

$$\xi_k^i \nabla_x f(u_k^i, \mu_k)^T d_k^i \leq \xi_k^i \nabla_x f(x_k, \mu_k)^T d_k^i + \xi_k^i \left( c_1 + \frac{1}{\mu_k} c_2 \right) \|u_k^i - x_k\|. \tag{20}$$

By (16), (19) and (20), we obtain

$$\nabla_x f(x_k, \mu_k)^T d_k^i + \left( c_1 + \frac{1}{\mu_k} c_2 \right) \|u_k^i - x_k\| \geq -\frac{o(\xi_k^i)}{\xi_k^i}$$

from which, taking into account (9) we can write

$$\left( \sum_{j=1}^m \lambda_j(x_k, \mu_k) \nabla f_j(x_k) \right)^T d_k^i + \left( c_1 + \frac{1}{\mu_k} c_2 \right) \|u_k^i - x_k\| \geq -\frac{o(\xi_k^i)}{\xi_k^i} \tag{21}$$

Since $\cup_{k\in K} D_k$ is a finite set by Assumption 2 and recalling the boundedness of the sequence $\{\lambda_j(x_k,\mu_k)\}$, $j=1,\dots,m$, there exist an infinite set $\bar K \subseteq K$ and, given the fact that $r_k$ is uniformly bounded, a finite set $J \subseteq \{1,2,\dots\}$ and $\bar{d}^j \in \Re^n$, $j \in J$, such that

$$\lim_{\substack{k\to\infty\\ k\in\bar K}} x_k = \bar x \tag{22}$$

$$\lim_{\substack{k\to\infty\\ k\in\bar K}} \lambda_j(x_k,\mu_k) = \bar\lambda_j \qquad j=1,\dots,m \tag{23}$$

$$J_k = J, \qquad \forall k \in \bar K, \tag{24}$$

$$d_k^j = \bar d^j, \qquad \forall j \in J \text{ and } \; k \in \bar K. \tag{25}$$

Moreover, recalling that $u_k^i = y_k^i + t_k^i \xi_k^i d_k^i$, with $t_k^i \in (0,1)$, we have that

$$\left(c_1 + \frac{1}{\mu_k} c_2\right) \|u_k^j - x_k\| \le \left(c_1 + \frac{1}{\mu_k} c_2\right) (\|y_k^j - x_k\| + \xi_k^j), \qquad \forall j \in J,$$

which, by using (18), implies

$$\lim_{\substack{k\to\infty\\ k\in\bar K}} \left(c_1 + \frac{1}{\mu_k} c_2\right) \|u_k^j - x_k\| = 0, \qquad \forall j \in J. \tag{26}$$

We note that expression (11) can be rewritten as

$$\lambda_j(x,\mu) = \frac{\exp((f_j(x) - f(x))/\mu)}{\sum_{l=1}^m \exp((f_l(x) - f(x))/\mu)}, \qquad j=1,\dots,m,$$

so that it is easily seen that

$$\begin{aligned} \bar\lambda_j &\ge 0 \quad \forall j, \\ \bar\lambda_j &= 0 \quad \forall j \notin B(\bar x). \end{aligned} \tag{27}$$

Furthermore, since $\sum_{j=1}^m \lambda_j(x_k,\mu_k) = 1$, for all $k$, then

$$\sum_{j=1}^m \bar\lambda_j = 1. \tag{28}$$

Now, taking limits for $k \to \infty$, $k \in \bar K$ in (21), and by recalling (26), we obtain

$$\left(\sum_{j=1}^m \bar\lambda_j \nabla f_j(\bar x)\right)^T \bar d^i \ge 0, \qquad \forall i \in J. \tag{29}$$

Now, Proposition 3.2 and Assumption 2 imply that for $k \in K$,

$$T(\bar x) = T(x_k;\epsilon) = cone\{D_k \cap T(x_k;\epsilon)\} = cone\{d_k^i\}_{i\in J_k}. \tag{30}$$

Hence, by (30), (24) and (25) we have that

$$T(\bar x) = cone\{\bar d^i\}_{i\in J}, \tag{31}$$

so that, for every $d \in T(\bar{x})$, by using (31) we can write

$$d = \sum_{i \in J} \beta_i \bar{d}^i, \tag{32}$$

with $\beta_i \geq 0$, for all $i \in J$. From (32) and recalling (29), we obtain

$$\left( \sum_{j=1}^{m} \bar{\lambda}_j \nabla f_j(\bar{x}) \right)^T d = \sum_{i \in J} \beta_i \left( \sum_{j=1}^{m} \bar{\lambda}_j \nabla f_j(\bar{x}) \right)^T \bar{d}^i \geq 0,$$

which, along with (27) and (28), proves the proposition. □

The above proposition is a non trivial extension of similar results established in the context of derivative-free methods for smooth optimization (see, for instance, [8]). The major novelty of Proposition 3.3 is relation (18) which relates the convergence rate of the smoothing parameter with that of the sampling step-sizes. Roughly speaking, this condition requires first order derivatives to be approximated faster than the original minimax problem. Indeed, Proposition 3.3 has two crucial aspects.

- The first one consists in showing that when $x_k \to \bar{x}$ and $\mu_k \to 0$, eventually,

$$\nabla_x f(x_k, \mu_k)^T d_k^i = \left( \sum_{j=1}^{m} \lambda_j(x_k, \mu_k) \nabla f_j(x_k) \right)^T d_k^i \geq 0 \quad \forall\, i \in J_k.$$

- The second aspect consists in exploiting the fact that the bounded sequence $\{(\lambda_1(x_k, \mu_k), \ldots, \lambda_m(x_k, \mu_k))\}$ has an accumulation point which allows us to overcome the difficulty tied to the indefiniteness of $\nabla_x f(x_k, \mu_k)$ in the limit.

The sampling of the smooth objective function along the directions $d_k^i$, $i \in J_k$, introduces a further difficulty, namely that $\nabla_x f(x_k, \mu_k)^T d_k^i$ is approximated by the following quantity

$$\nabla_x f(u_k^i, \mu_k)^T d_k^i = \left( \sum_{j=1}^{m} \lambda_j(u_k^i, \mu_k) \nabla f_j(u_k^i) \right)^T d_k^i,$$

where, for every index $i \in J_k$, we have different bounded sequences $\{(\lambda_1(u_k^i, \mu_k), \ldots, \lambda_m(u_k^i, \mu_k))\}$. This raises the problem that each of these sequences converges to its own limit while the optimality condition (29) requires them to have the same limit point. In order to guarantee the existence of a unique limit point of the sequences $\{(\lambda_1(u_k^i, \mu_k), \ldots, \lambda_m(u_k^i, \mu_k))\}$, for all $i \in J_k$, it is necessary that $\|u_k^i - x_k\|$, $i \in J_k$, tends to zero faster than $\mu_k$, where $\|u_k^i - x_k\|$ can be seen as a measure of the degree of approximation of first order derivatives and $\mu_k$ gives a measure of the degree of approximation of the original minimax problem.
To conclude, we note that since Proposition 3.3 poses only an upper bound on the convergence rate of $\mu_k$ towards zero, it allows us to choose an updating rule for the smoothing parameter which conciliate global convergence with the problem of avoiding the ill-conditioning of the smooth approximating problem.

## 4. A derivative-free method and global convergence result

In this section we define an algorithm for the solution of Problem (1). The proposed method stems from the union of a derivative-free approach for smooth and linearly constrained optimization with a suitable handling of the smoothing parameter $\mu$. In particular, the derivative-free method samples the smoothing function value along a finite set of search directions and decreases the sampling stepsize and the smoothing parameter if a sufficiently improved objective function value is not attained. The sampling strategy and the updating rule of the smoothing parameter are guided by the convergence conditions of Proposition 3.3. The derivative-free technique for sampling the smoothing function is based on the *Feasible descent method 2* proposed in [8] for a class of smooth optimization problems including linearly constrained problems. The formal description of the algorithm is reported below.

**DF Algorithm**

**Data.** $x_0 \in \mathcal{F}$, $\tilde{\alpha}_0 > 0$, $\mu_0$, $\gamma > 0$, $\theta \in (0,1)$, $\bar{\epsilon} > 0$.

**Step 0.** Set $k = 0$.

**Step 1.** *(Computation of search directions)*

Choose a set of directions $D_k = \{d_k^1, \dots, d_k^{r_k}\}$ satisfying Assumption 2.

**Step 2.** *(Minimization on the cone$\{D_k\}$)*

**Step 2.1.** *(Initialization)*

Set $i = 1$, $y_k^i = x_k$, $\tilde{\alpha}_k^i = \tilde{\alpha}_k$.

**Step 2.2.** *(Computation of the initial stepsize)*

Compute the maximum steplength $\bar{\alpha}_k^i$ such that $y_k^i + \bar{\alpha}_k^i d_k^i \in \mathcal{F}$ and set $\hat{\alpha}_k^i = \min\{\bar{\alpha}_k^i, \tilde{\alpha}_k^i\}$.

**Step 2.3.** *(Test on the search direction)*

If $\hat{\alpha}_k^i > 0$ and $f(y_k^i + \hat{\alpha}_k^i d_k^i, \mu_k) \le f(y_k^i, \mu_k) - \gamma(\hat{\alpha}_k^i)^2$,

compute $\alpha_k^i$ by the *Expansion Step*($\bar{\alpha}_k^i, \hat{\alpha}_k^i, d_k^i; \alpha_k^i$) and set $\tilde{\alpha}_k^{i+1} = \alpha_k^i$;

otherwise set $\alpha_k^i = 0$ and $\tilde{\alpha}_k^{i+1} = \theta\tilde{\alpha}_k^i$.

**Step 2.4.** *(New point)*

Set $y_k^{i+1} = y_k^i + \alpha_k^i d_k^i$.

**Step 2.5** *(Test on the minimization on the cone*$\{D_k\}$)

If $i = r_k$ go to Step 3;

otherwise set $i = i + 1$ and go to Step 2.2.

**Step 3.** *(Main iteration)*

Find $x_{k+1} \in \mathcal{F}$ such that $f(x_{k+1}, \mu_k) \le f(y_k^{i+1}, \mu_k)$;

otherwise set $x_{k+1} = y_k^{i+1}$.

Set $\tilde{\alpha}_{k+1} = \tilde{\alpha}_k^{i+1}$,

choose $\mu_{k+1} = \min\left\{\mu_k, \max_{i=1,\dots,r_k}\{(\tilde{\alpha}_k^i)^{1/2}, (\alpha_k^i)^{1/2}\}\right\}$,

set $k = k + 1$, and go to Step 1.

**Expansion Step** $(\bar{\alpha}_k^i, \hat{\alpha}_k^i, y_k^i, d_k^i; \alpha_k^i)$**.**

**Data.** $\gamma > 0$, $\delta \in (0,1)$.

**Step 1.** Set $\alpha = \hat{\alpha}_k^i$.

**Step 2.** Let $\tilde{\alpha} = \min\{\bar{\alpha}_k^i, \{\alpha/\delta\}\}$.
If $\alpha = \bar{\alpha}_k^i$ or $f(y_k^i + \tilde{\alpha} d_k^i, \mu_k) > f(y_k^i, \mu_k) - \gamma\tilde{\alpha}^2$ set $\alpha_k^i = \alpha$ and stop.

**Step 3.** Set $\alpha = \tilde{\alpha}$ and go to Step 2.

At Step 1 a suitable set of search directions $d_k^1, \dots, d_k^{r_k}$ is determined. At Step 2 the behavior of the objective function is evaluated along each search direction. In particular, if the search direction is feasible and is of sufficient decrease, the behavior of the objective function along this direction is further investigated by executing an *Expansion Step* until a suitable increase of the objective function is detected or the trial point reaches the boundary of the feasible region. In Step 3, the new point $x_{k+1}$ can be the point $y_k^{i+1}$ produced by Steps 1-2 or any point where the objective function is improved with respect to $f(y_k^{r_k}, \mu_k)$. This fact allows us to adopt any approximation scheme for the objective function to produce a new better point. Then the smoothing parameter $\mu_k$ is reduced whenever $\max_{i=1,\dots,r_k}\{(\tilde{\alpha}_k^i)^{1/2}, (\alpha_k^i)^{1/2}\}$ gets smaller than the current smoothing value $\mu_k$. We recall that $\max_{i=1,\dots,r_k}\{(\tilde{\alpha}_k^i)^{1/2}, (\alpha_k^i)^{1/2}\}$ can be viewed as a stationarity measure of the current iterate, see for example [6]. Thus, according to the updating rule, the smoothing parameter is reduced whenever an approximation of a stationary point of the smoothing function is obtained with a finer and finer precision.

The following proposition describes some properties concerning the sequences of points and of objective function values generated by Algorithm DF.

**Proposition 4.1.** *Let* $\{x_k\}$, $\{\mu_k\}$ *be the sequences generated by* DF *Algorithm. Then*

(a) $\{x_k\}$ *is well defined;*

(b) *the sequence* $\{f(x_k, \mu_k)\}$ *is monotonically decreasing;*

(c) *the sequences* $\{f(x_{k+1}, \mu_k)\}$ *and* $\{f(x_k, \mu_k)\}$ *are both convergent and have the same limit;*

(d) *the sequence* $\{x_k\}$ *is bounded;*

(e) *every cluster point of* $\{x_k\}$ *belongs to* $\mathcal{F}$.

**Proof.** To prove assertion (a), it suffices to show that the Expansion Step, when performed along a direction $d_k^i$ from $y_k^i$, for $i \in \{1, \dots, r_k\}$, terminates in a finite number $j$ of steps. If this were not true, we would have for some $k$ and $i \in \{1, \dots, r_k\}$ that $\hat{\alpha}_k^i > 0$ and

$$y_k^i + \delta^{-j}\hat{\alpha}_k^i d_k^i \in \mathcal{F} \qquad \text{for all } j = 0, 1, \dots,$$

and, recalling point (i) of Proposition 2.2,

$$f(y_k^i + \delta^{-j}\hat{\alpha}_k^i d_k^i) \le f(y_k^i + \delta^{-j}\hat{\alpha}_k^i d_k^i, \mu_k) < f(y_k^i, \mu_k) - \gamma\Big(\delta^{-j}\hat{\alpha}_k^i\Big)^2,$$

for all $j = 0, 1, \ldots$, which violates the assumption that $f$ is bounded below on $\mathcal{F}$.

To prove assertion (b) we note that the instructions of the algorithm imply

$$f(x_{k+1}, \mu_k) \le f(x_k, \mu_k).$$

Since $\mu_{k+1} \le \mu_k$ and $f(x, \mu)$ is increasing with respect to $\mu$ (see (i) of Proposition 2.2), we have

$$f(x_{k+1}, \mu_{k+1}) \le f(x_{k+1}, \mu_k) \le f(x_k, \mu_k), \tag{33}$$

so that assertion (b) is proved.
Now, by Assumption 1 and recalling (8) and (33) we have $f(x_{k+1}) \le f(x_{k+1}, \mu_{k+1})$, so that also assertion (c) is proved.
By assertion (b) we have $f(x_k, \mu_k) \le f(x_0, \mu_0)$ for all $k$, and hence

$$x_k \in \{x \mid f(x, \mu_k) \le f(x_0, \mu_0)\}.$$

From (i) of Proposition 2.2 we get that for any $x$ such that

$$f(x, \mu_k) \le f(x_0, \mu_0)$$

we have

$$f(x) \le f(x_0, \mu_0),$$

so that we can write

$$x_k \in \{x \mid f(x, \mu_k) \le f(x_0, \mu_0)\} \subseteq \{x \mid f(x) \le f(x_0, \mu_0)\}.$$

By Assumption 1 it follows that the set $\{x \mid f(x) \le f(x_0, \mu_0)\}$ is bounded, and hence assertion (d) is proved.
To prove the assertion (e), we note that the instructions of Algorithm DF imply that $x_k \in \mathcal{F}$ for all $k$. Since $\mathcal{F}$ is a closed set, the assertion follows. □

The proposition which follows states some results concerning the sampling technique adopted. More in particular, point (i) guarantees that the sampling points tend to cluster more and more. Point (ii) ensures the existence of sufficiently large step lengths providing feasible points along the search directions.

**Proposition 4.2.** *Let $\{x_k\}$ be the sequence produced by Algorithm DF. Then:*

*(i) we have*

$$\lim_{k\to\infty} \max_{1\le i\le \tilde{r}_k} \left\{\alpha_k^i\right\} = 0, \tag{34}$$

$$\lim_{k\to\infty} \max_{1\le i\le \tilde{r}_k} \left\{\tilde{\alpha}_k^i\right\} = 0, \tag{35}$$

$$\lim_{k\to\infty} \max_{1\le i\le \tilde{r}_k} \left\|x_k - y_k^i\right\| = 0. \tag{36}$$

*(ii) $\bar{\alpha}_k^i \ge \epsilon/c - \|x_k - y_k^i\|$ whenever $d_k^i \in T(x_k, \epsilon)$ and $\epsilon > 0$, where $c = \max_{j=1,\ldots,m} \|a_j\|$.*

**Proof.** To prove assertion (i), we note from the construction of $\alpha_k^i$ and $y_k^{i+1}$ in Step 2.3 that

$$f(y_k^{i+1},\mu_k) \le f(y_k^i,\mu_k) - \gamma(\alpha_k^i)^2, \tag{37}$$

and from the construction of $\tilde{\alpha}_k^{i+1}$ that

$$\begin{aligned} \text{either} \qquad & \tilde{\alpha}_k^{i+1} = \alpha_k^i & (38)\\ \text{or} \qquad & \tilde{\alpha}_k^{i+1} = \theta\tilde{\alpha}_k^i, & (39)\end{aligned}$$

for each $k$ and each $i \in \{1,\dots,r_k\}$. Summing (37) for $i = 1,...,r_k$ and using the construction of $x_{k+1}$ in Step 3 yields

$$f(x_{k+1},\mu_k) \le f(x_k,\mu_k) - \gamma\sum_{i=1}^{r_k}(\alpha_k^i)^2.$$

Recalling point (c) of Proposition 4.1, $\{f(x_k,\mu_k)\}$ and $\{f(x_{k+1},\mu_k)\}$ are both convergent and have the same limit, and $\{\sum_{i=1}^{r_k}(\alpha_k^i)^2\} \to 0$, thus proving (34).
For all $k$ we have

$$\tilde{\alpha}_k^i = (\theta)^{p_k^i}\,\alpha_{m_k^i}^{l_k^i}, \tag{40}$$

where $m_k^i$ is the biggest iteration index such that $m_k^i \le k$ and $l_k^i \le r_{m_k^i}$ is the biggest direction index such that (38) holds and the exponent $p_k^i$ is given by

$$\begin{aligned} p_k^i &= i - l_k^i, && \text{if } m_k^i = k & (41)\\ p_k^i &= i + r_{k-1} + \cdots + r_{m_k^i} - l_k^i, && \text{otherwise.} & (42)\end{aligned}$$

Then, let $i$ be an arbitrary integer such that the set $K^i = \{k \in \{0,1,...\} : i \le r_k\}$ has infinitely many elements. If $m_k^i \to \infty$, as $k \to \infty$ with $k \in K^i$, then, by (40) and (34), we get (35).
Otherwise, assume that $m_k^i$ is uniformly bounded above, then $m_k^i < k$ for all $k \in K^i$ sufficiently large so that $p_k^i$ is given by (42). Since $r_{m_k^i} \ge l_k^i$ and $r_l \ge 1$ for $l = m_k^i+1,\dots,k-1$, this then implies $p_k^i \ge i + (k-1-m_k^i)$ so that $p_k^i \to \infty$ as $k \to \infty$, $k \in K^i$. Hence, by (40) and $\theta < 1$ we get (35).
Then, we note from the updating formula for $y_k^i$ in Step 2.4 that

$$x_k - y_k^i = -\sum_{l=1}^{i-1}\alpha_k^l d_k^l.$$

Then, using (34), $\|d_k^l\| = 1$ for $1 \le l \le r_k$, $i \le r_k$, and the assumption that $\{r_k\}$ is uniformly bounded, we obtain (36).

To prove assertion (ii), notice that $d_k^i \in T(x_k;\epsilon)$ and the definition of $\bar{\alpha}_k^i$ in Step 2.2 imply either $\bar{\alpha}_k^i = +\infty$ or the existence of an index $\bar{\jmath} \notin I(x_k;\epsilon)$ such that

$$a_{\bar{\jmath}}^T(y_k^i + \bar{\alpha}_k^i d_k^i) = b_{\bar{\jmath}}.$$

Then, solving for $\bar{\alpha}_k^i$ and using $0 < a_{\bar{\jmath}}^T d_k^i \le c$, yields

$$\begin{aligned} \bar{\alpha}_k^i &= \left(b_{\bar{\jmath}} - a_{\bar{\jmath}}^T y_k^i\right)/(a_{\bar{\jmath}}^T d_k^i)\\ &\ge \left(b_{\bar{\jmath}} - a_{\bar{\jmath}}^T y_k^i\right)/c \end{aligned}$$

$$\begin{aligned} &= \left(b_{\bar{\jmath}} - a_{\bar{\jmath}}^T x_k + a_{\bar{\jmath}}^T (x_k - y_k^i)\right)/c \\ &\geq \left(\epsilon + a_{\bar{\jmath}}^T (x_k - y_k^i)\right)/c \\ &\geq \left(\epsilon - \|x_k - y_k^i\| c\right)/c, \end{aligned}$$

where the second inequality follows from $\bar{\jmath} \notin I(x_k;\epsilon)$ and the definition of $I(x_k;\epsilon)$. □

The next proposition establishes the convergence properties of Algorithm DF.

**Proposition 4.3.** *Let $\{x_k\}$ be the sequence generated by Algorithm DF. Then, a limit point of the sequence $\{x_k\}$ exists which is a stationary point of the minimax problem (1).*

**Proof.** By using the results of Proposition 4.2, we get that

$$\lim_{k\to\infty} \mu_k = 0. \tag{43}$$

Let $\{x_k\}_K$ be the subsequence corresponding to the subset indices $K$ such that

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

Relation (43) guarantees that the index set $K$ has infinitely many elements.
Let now $\bar{x}$ be an accumulation point of the subsequence $\{x_k\}_K$ and $\epsilon \in (0, \min\{\bar{\epsilon}, \epsilon^\star\}]$ where $\bar{\epsilon}$ and $\epsilon^\star$ are defined in Algorithm DF and Proposition 3.2 respectively. Let

$$J_k = \{i \in \{1,\dots,r_k\} : d_k^i \in D_k \cap T(x_k,\epsilon)\}.$$

Now, Proposition 3.2 and Assumption 2 imply that for $k \in K$,

$$T(\bar{x}) = T(x_k;\epsilon) = cone\{D_k \cap T(x_k;\epsilon)\} = cone\{d_k^i\}_{i\in J_k}. \tag{44}$$

For all the search directions $d_k^i$, $i \in J_k$, that is $d_k^i \in T(x_k;\epsilon)$, point (ii) of Proposition 4.2 shows that

$$\bar{\alpha}_k^i \geq \epsilon/c - \left\|x_k - y_k^i\right\|,$$

which, using also point (i) of Proposition 4.2 implies there exists an index $\bar{k}$ such that, for all $k \geq \bar{k}$ and $k \in K$,

$$\alpha_k^i/\delta < \bar{\alpha}_k^i \quad \text{and} \quad \hat{\alpha}_k^i < \bar{\alpha}_k^i, \tag{45}$$

where $\hat{\alpha}_k^i = \min\{\bar{\alpha}_k^i, \tilde{\alpha}_k^i\}$.
Then, the construction of $\alpha_k^i$ in Step 2.3 implies, for each $i \in J_k$, either

$$y_k^i + \frac{\alpha_k^i}{\delta} d_k^i \in \mathcal{F}, \qquad f(y_k^i + \frac{\alpha_k^i}{\delta} d_k^i, \mu_k) > f(y_k^i, \mu_k) - \gamma(\frac{\alpha_k^i}{\delta})^2$$

if an Expansion Step is performed or otherwise

$$y_k^i + \hat{\alpha}_k^i d_k^i \in \mathcal{F}, \qquad f(y_k^i + \hat{\alpha}_k^i d_k^i, \mu_k) > f(y_k^i, \mu_k) - \gamma(\hat{\alpha}_k^i)^2.$$

Letting $\xi_k^i = \alpha_k^i/\delta$ in the first case and $\xi_k^i = \hat{\alpha}_k^i$ in the second case, then, by the instructions of Algorithm DF we have

$$f(y_k^i + \xi_k^i d_k^i, \mu_k) \geq f(y_k^i, \mu_k) - \gamma(\xi_k^i)^2. \tag{46}$$

From the updating formula for $y_k^i$ in Step 2.4 of Algorithm DF, we note that

$$\|y_k^i - x_k\| \le \sum_{l=1}^{i-1} \alpha_k^l \le \delta \sum_{l=1}^{i-1} \xi_k^l \le \delta r_k \max_{j \in J_k} \{\xi_k^j\}. \tag{47}$$

Moreover, by considering relations (45), we have that

$$\max_{i \in J_k} \{\xi_k^i, \|x_k - y_k^i\|\} \le \max\{1, \delta r_k\} \max_{i \in J_k} \{\xi_k^i\} \le r_k \max_{i \in J_k} \{\tilde{\alpha}_k^i, \alpha_k^i\}. \tag{48}$$

For all $k \in K$, by the instruction at Step 3 of Algorithm DF, we know that

$$\mu_k^2 > \max_{j=1,\dots,r_k} \left\{ (\tilde{\alpha}_k^j), (\alpha_k^j) \right\}, \tag{49}$$

so that, by (48) and (49), we obtain

$$\lim_{k \to \infty, k \in K} \frac{\max_{i \in J_k} \{\xi_k^i, \|x_k - y_k^i\|\}}{\mu_k} = 0. \tag{50}$$

Finally, (43), (46), (50) and the result of Proposition 3.3 conclude the proof. □

**Corollary 4.4.** *Let $\{x_k\}$ be the sequence produced by Algorithm DF and let $\{x_k\}_K$ be the subsequence corresponding to the subset of indices $K$ such that*

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

*Then, every accumulation point of $\{x_k\}_K$ is a stationary point of the minimax problem (1).*

## 5. Numerical results

The aim of the computational experiments is to investigate the ability of the proposed algorithm to locate a good approximation of a solution point of the finite minimax problem (1). We report the numerical results obtained by Algorithm DF both on a set of 33 unconstrained minimax problems with $n \in [1, 200]$, $q \in [2, 501]$ (see [14] and [4] for a description of these problems) and on a set of 5 linearly constrained minimax problems with $n \in [2, 20]$, $q \in [3, 14]$ and $m \in [1, 4]$ (see [17] for a description of these test problems).
As regards the choice of the parameters which define Algorithm DF, we use the following values

$$\begin{array}{lll} \tilde{\alpha}_0 = 1.0, & \mu_0 = 1.0, & \gamma = 10^{-6}, \\ \theta = 0.5, & \delta = 0.5, & \bar{\epsilon} = 1.0\,. \end{array}$$

As for the search directions, in the linearly constrained setting we use the computation strategy proposed in [16] whereas, in the unconstrained case, we use $D_k = D = \{\pm e_1, \dots, \pm e_n\}$. In the latter case, we further exploit the fact that set $D_k$ is constant. First of all, we modify Step 2 by adopting the step size updating strategy proposed in [7] where each search direction $e_i$, $i = 1, \dots, n$, has associated its own step size. Furthermore, in Step 3 a point $\hat{x}$ is computed by performing a linesearch along an additional direction described at Step 4 of Algorithm 3 in [7]. Then, $x_{k+1} = \hat{x}$ provided that $f(\hat{x}, \mu_k) \le f(y_k^{i+1}, \mu_k)$, otherwise we set $x_{k+1} = y_k^{i+1}$. We note that in the linearly constrained case we always set $x_{k+1} = y_k^{i+1}$.

As concerns the stopping condition, we choose to stop the algorithm when $\tilde{\alpha}_k \leq 10^{-4}$ in the constrained case and $\max_{i=1,\dots,n} \tilde{\alpha}_k^i \leq 10^{-4}$ in the unconstrained case. Furthermore, we also stop the computation whenever the code reaches a total of 50000 function evaluations.
Table 1 reports the numerical results obtained by Algorithm DF. Beside the problem name, the table reports, for every problem, the number $n$ of variables, the number $q$ of component functions and the number $m$ of linear constraints. We denote by $f(\bar{x})$ the minimum value obtained by Algorithm DF, by $\bar{\mu}$ the value of the smoothing parameter when the stopping condition is met and by $f^\star$ the value of the known solution. Furthermore, we denote by

$$\Delta = \frac{f(\bar{x}) - f^\star}{1 + |f^\star|}$$

the error at the solution obtained by Algorithm DF.
From Table 1 it clearly emerges the ability of Algorithm DF to locate a good estimate of the minimum point of the minimax problem (1) (as reported in [17] and [4]) with a limited number of function evaluations especially for problems with a reasonable number of variables (say, less than 10) in a derivative free context. It is worth noting that for almost every problem, the value $\bar{\mu}$, that is the final value of the smoothing parameter, is of order $10^{-2}$ or slightly less than that. In order to better point out the efficiency of the proposed derivative free method and the usefulness of the smooth approximating function $f(x,\mu)$ along with the updating rule for $\mu$, we compare Algorithm DF with some reasonable modifications of it. First of all, it seems reasonable to try and modify Algorithm DF into $\text{DF}_{\texttt{mod1}}$ so as to always use the max function $f(x)$ instead of the smooth approximation $f(x,\mu)$. This will help us to evaluate the benefits of using an algorithm which is proved to be convergent toward stationary points of problem (1) instead of a method which is only convergent toward weak stationary points (see appendix A). Secondly, in order to judge the effectiveness of the updating rule for the smoothing parameter, we choose to compare Algorithm DF with algorithms $\text{DF}_{\texttt{mod2}}$ and $\text{DF}_{\texttt{mod3}}$ which can be obtained from Algorithm DF by dropping the updating rule for $\mu$ at Step 3 and choosing $\mu_0 = 1$ and $\mu_0 = 10^{-2}$ respectively.
The complete results obtained by the three modified versions of Algorithm DF ($\text{DF}_{\texttt{mod1}}$, $\text{DF}_{\texttt{mod2}}$ and $\text{DF}_{\texttt{mod3}}$) are reported in Appendix B. Here, for the sake of clarity, we only report a summary of the obtained results. In particular, for every algorithm, Table 2 indicates how many problems were solved with an error $\Delta < 10^{-3}$, $10^{-3} \leq \Delta < 10^{-1}$ and $\Delta \geq 10^{-1}$. From these results it clearly emerges the superiority of Algorithm DF with respect to algorithms $\text{DF}_{\texttt{mod1}}$ and $\text{DF}_{\texttt{mod2}}$. As for method $\text{DF}_{\texttt{mod3}}$ it has two failures ($\Delta \geq 10^{-1}$) more than DF but it still performs well and seems to have a behavior quite similar to that of DF itself. However, by a closer examination of the complete results of algorithms DF and $\text{DF}_{\texttt{mod3}}$ (Tables 1 and 5 respectively) it is easily seen that the two algorithms performs quite differently in terms of function evaluations. Indeed, if we sum up the number of function evaluations needed for those problems which are solved with an error $\Delta < 10^{-3}$ by each code, we get a 7135 figure for Algorithm DF as opposed to 21713 for Algorithm $\text{DF}_{\texttt{mod3}}$. This significant difference in terms of function evaluations between the two codes properly points out the fundamental importance of the updating rule for the smoothing parameter $\mu$ whose ultimate task is that of limiting the ill-conditioning of the approximating problem. Indeed, when we fix the smoothing parameter to $10^{-2}$ we come up with too an ill-conditioned problem from the early stages of the solution process. On the contrary, Algorithm DF tries to limit the ill-conditioning especially in the early stages of the minimization by appropriately decreasing the smoothing parameter at a suitable rate.

| PROBLEM | n | q | m | nF | $f(\bar{x})$ | $\bar{\mu}$ | $f^\star$ | $\Delta$ |
|---|---|---|---|---|---|---|---|---|
| crescent | 2 | 2 | 0 | 160 | 3.061E-03 | 1.105E-02 | 0.000E+00 | 3.061E-03 |
| polak 1 | 2 | 2 | 0 | 106 | 2.718E+00 | 7.812E-03 | 2.718E+00 | 7.654E-09 |
| lq | 2 | 2 | 0 | 343 | -1.411E+00 | 7.812E-03 | -1.414E+00 | 1.158E-03 |
| mifflin 1 | 2 | 2 | 0 | 65 | -1.000E+00 | 1.210E-02 | -1.000E+00 | 0.000E+00 |
| mifflin 2 | 2 | 2 | 0 | 188 | -9.980E-01 | 7.813E-03 | -1.000E+00 | 1.009E-03 |
| charalambous-conn 1 | 2 | 3 | 0 | 118 | 1.954E+00 | 9.882E-03 | 1.952E+00 | 4.631E-04 |
| charalambous-conn 2 | 2 | 3 | 0 | 208 | 2.003E+00 | 1.105E-02 | 2.000E+00 | 1.153E-03 |
| demyanov-malozemov | 2 | 3 | 0 | 84 | -3.000E+00 | 1.105E-02 | -3.000E+00 | 0.000E+00 |
| ql | 2 | 3 | 0 | 132 | 7.203E+00 | 1.105E-02 | 7.200E+00 | 3.575E-04 |
| hald-madsen 1 | 2 | 4 | 0 | 170 | 1.582E-02 | 1.105E-02 | 0.000E+00 | 1.582E-02 |
| rosen | 4 | 4 | 0 | 368 | -4.394E+01 | 7.906E-03 | -4.400E+01 | 1.347E-03 |
| hald-madsen 2 | 5 | 42 | 0 | 471 | 6.177E-03 | 7.906E-03 | 1.220E-04 | 6.055E-03 |
| polak 2 | 10 | 2 | 0 | 285 | 5.460E+01 | 7.813E-03 | 5.459E+01 | 1.134E-04 |
| maxq | 20 | 20 | 0 | 1858 | 0.000E+00 | 1.105E-02 | 0.000E+00 | 0.000E+00 |
| maxl | 20 | 40 | 0 | 891 | 0.000E+00 | 1.105E-02 | 0.000E+00 | 0.000E+00 |
| goffin | 50 | 50 | 0 | 2045 | 0.000E+00 | 7.813E-03 | 0.000E+00 | 0.000E+00 |
| polak 6.1 | 2 | 3 | 0 | 131 | 1.954E+00 | 1.118E-02 | 1.952E+00 | 4.760E-04 |
| polak 6.2 | 20 | 20 | 0 | 692 | 2.384E-09 | 1.105E-02 | 0.000E+00 | 2.384E-09 |
| polak 6.3 | 4 | 50 | 0 | 2055 | 6.253E-03 | 7.813E-03 | 2.637E-03 | 3.607E-03 |
| polak 6.4 | 4 | 102 | 0 | 1105 | 9.166E-03 | 7.813E-03 | 2.650E-03 | 6.499E-03 |
| polak 6.5 | 4 | 202 | 0 | 1890 | 9.181E-03 | 7.813E-03 | 2.650E-03 | 6.515E-03 |
| polak 6.6 | 3 | 50 | 0 | 374 | 6.531E-03 | 7.813E-03 | 4.500E-03 | 2.022E-03 |
| polak 6.7 | 3 | 102 | 0 | 335 | 7.141E-03 | 7.813E-03 | 4.505E-03 | 2.624E-03 |
| polak 6.8 | 3 | 202 | 0 | 369 | 7.263E-03 | 7.813E-03 | 4.505E-03 | 2.746E-03 |
| polak 6.9 | 2 | 2 | 0 | 91 | 1.162E-01 | 7.812E-03 | 0.000E+00 | 1.162E-01 |
| polak 6.10 | 1 | 25 | 0 | 129 | 1.784E-01 | 1.105E-02 | 1.782E-01 | 1.625E-04 |
| polak 6.11 | 1 | 51 | 0 | 136 | 1.784E-01 | 1.105E-02 | 1.783E-01 | 6.206E-05 |
| polak 6.12 | 1 | 101 | 0 | 153 | 1.784E-01 | 1.105E-02 | 1.784E-01 | 2.368E-05 |
| polak 6.13 | 1 | 501 | 0 | 153 | 1.784E-01 | 1.105E-02 | 1.784E-01 | 1.464E-05 |
| polak 6.14 | 100 | 100 | 0 | 3452 | 3.433E-09 | 1.105E-02 | 0.000E+00 | 3.433E-09 |
| polak 6.15 | 200 | 200 | 0 | 6891 | 3.433E-09 | 1.105E-02 | 0.000E+00 | 3.433E-09 |
| polak 6.16 | 100 | 50 | 0 | 3452 | 5.364E-09 | 1.105E-02 | 0.000E+00 | 5.364E-09 |
| polak 6.17 | 200 | 50 | 0 | 7233 | 1.023E-08 | 7.812E-03 | 0.000E+00 | 1.023E-08 |
| mad 1 | 2 | 3 | 1 | 43 | -3.896E-01 | 1.747E-02 | -3.897E-01 | 5.878E-05 |
| mad 2 | 2 | 3 | 1 | 42 | -3.304E-01 | 1.353E-02 | -3.304E-01 | -9.735E-10 |
| mad 4 | 2 | 3 | 2 | 72 | -4.489E-01 | 1.562E-02 | -4.489E-01 | 4.601E-07 |
| wong 2 | 10 | 6 | 3 | 236 | 2.522E+01 | 1.948E-02 | 2.431E+01 | 3.609E-02 |
| wong 3 | 20 | 14 | 4 | 451 | 1.076E+02 | 2.545E-02 | 1.337E+02 | -1.938E-01 |

Table 1: Numerical performance of Algorithm DF

| | $\Delta < 10^{-3}$ | $\Delta < 10^{-1}$ | $\Delta \geq 10^{-1}$ |
|---|---|---|---|
| DF | 23 | 14 | 1 |
| $\mathrm{DF}_{\mathtt{mod1}}$ | 14 | 12 | 12 |
| $\mathrm{DF}_{\mathtt{mod2}}$ | 16 | 16 | 6 |
| $\mathrm{DF}_{\mathtt{mod3}}$ | 21 | 14 | 3 |

Table 2: Comparison of methods

## A. Appendix

Let $f : R^n \to R$ be a locally *Lipschitz* nonsmooth function namely such that for every $x \in R^n$

$$|f(y^{'}) - f(y^{''})| \leq L\|y^{'} - y^{''}\|$$

for all $y^{'}, y^{''}$ belonging to a ball of radius $\delta$ centered at $x$. The *generalized directional derivative* [3] of $f$ at $x$ in the direction $d$ is denoted by $f^0(x, d)$ and is defined as follows

$$f^0(x, d) = \limsup_{y \to x, t \to 0^+} \frac{f(y + td) - f(y)}{t} \tag{51}$$

Under the assumption that $f$ is Lipschitz near $x$, $f^0(x, d)$ is well defined.
The following proposition extends to Lipschitz continuous functions analogous results reported in [7] and [8] concerning general convergence conditions for derivative free methods.

**Proposition A.1.** *Let $\{x_k\}$ be a sequence of feasible points and $\bar{x}$ be a limit point of a subsequence $\{x_k\}_K$ for some infinite set $K \subseteq \{0, 1, \ldots\}$. Let $\{D_k\}$, with $D_k = \{d_k^1, \ldots, d_k^{r_k}\}$, be a sequence of sets of directions which satisfy Assumption 2 and $J_k = \{i \in \{1, \ldots, r_k\} : d_k^i \in T(x_k, \epsilon)\}$ with $\epsilon \in (0, \min\{\bar{\epsilon}, \epsilon^\star\}]$ (where $\bar{\epsilon}$ and $\epsilon^\star$ are defined in Assumption 2 and Proposition 3.2, respectively).*
*Suppose that the following conditions hold:*

*(a) for each $k \in K$ and $i \in J_k$, there exist $y_k^i$ and scalars $\xi_k^i > 0$ such that:*

$$y_k^i + \xi_k^i d_k^i \in \mathcal{F} \tag{52}$$

$$f(y_k^i + \xi_k^i d_k^i) \geq f(y_k^i) - o(\xi_k^i); \tag{53}$$

$$\lim_{k \to \infty, k \in K} \max_{i \in J_k} \{\xi_k^i\} = 0; \tag{54}$$

$$\lim_{k \to \infty} \max_{i \in J_k} \|x_k - y_k^i\| = 0. \tag{55}$$

*Then,*

$$\lim_{k \to \infty, k \in K} \min_{i \in J_k} \left\{ \min\{0, f^0(x_k, d_k^i)\} \right\} = 0. \tag{56}$$

**Proof.** Since $\bigcup_{k \in K} D_k$ is a finite set, there exist an infinite subset $K_1 \subseteq K$ and $J \subset \{1, 2, \ldots\}$ and $r \geq 1$ such that

$$J_k = J \qquad \text{for all } k \in K_1,$$

$$\{d_k^i\}_{i \in I_k} = \{\bar{d}^1, \ldots, \bar{d}^r\}, \qquad \|\bar{d}^i\| = 1 \qquad \text{for all } k \in K_1.$$

By using condition (55) it follows

$$\lim_{k \to \infty, k \in K_1} y_k^i = \bar{x}, \qquad i \in J. \tag{57}$$

Now, recalling condition (53), for all $k \in K_1$, we have

$$f(y_k^i + \xi_k^i \bar{d}^i) - f(y_k^i) \geq -o(\xi_k^i), \quad i \in J, \tag{58}$$

from which we obtain

$$\limsup_{k\to\infty, k\in K_1} \frac{f(y_k^i + \xi_k^i \bar{d}^i) - f(y_k^i)}{\xi_k^i} \geq 0. \tag{59}$$

Since $f(x)$ is locally Lipschitz near $\bar{x}$, by using (51), (54), and (57) we can write

$$f^o(\bar{x}, \bar{d}^i) \geq \limsup_{k\to\infty, k\in K_1} \frac{f(y_k^i + \xi_k^i \bar{d}^i) - f(y_k^i)}{\xi_k^i} \qquad i = 1, \dots, r,$$

so that, from (59), we obtain

$$f^o(\bar{x}, \bar{d}^i) \geq 0 \qquad i = 1, \dots, r, \tag{60}$$

which proves (56). □

## B. Appendix

Here we report the complete results for the modified versions of Algorithm DF, namely $\mathrm{DF}_{\mathtt{mod1}}$, $\mathrm{DF}_{\mathtt{mod2}}$ and $\mathrm{DF}_{\mathtt{mod3}}$

| PROBLEM | n | q | m | nF | $f(\bar{x})$ | $\bar{\mu}$ | $f^\star$ | $\Delta$ |
|---|---|---|---|---|---|---|---|---|
| crescent | 2 | 2 | 0 | 78 | 0.000E+00 | 1.105E-02 | 0.000E+00 | 0.000E+00 |
| polak 1 | 2 | 2 | 0 | 106 | 2.718E+00 | 7.812E-03 | 2.718E+00 | 7.654E-09 |
| lq | 2 | 2 | 0 | 86 | -1.395E+00 | 7.812E-03 | -1.414E+00 | 7.771E-03 |
| mifflin 1 | 2 | 2 | 0 | 185 | -1.000E+00 | 1.210E-02 | -1.000E+00 | 6.358E-08 |
| mifflin 2 | 2 | 2 | 0 | 74 | -1.000E+00 | 7.812E-03 | -1.000E+00 | 0.000E+00 |
| charalambous-conn 1 | 2 | 3 | 0 | 80 | 2.000E+00 | 7.812E-03 | 1.952E+00 | 1.618E-02 |
| charalambous-conn 2 | 2 | 3 | 0 | 81 | 2.000E+00 | 1.105E-02 | 2.000E+00 | 0.000E+00 |
| demyanov-malozemov | 2 | 3 | 0 | 84 | -3.000E+00 | 1.105E-02 | -3.000E+00 | 0.000E+00 |
| ql | 2 | 3 | 0 | 92 | 7.812E+00 | 7.812E-03 | 7.200E+00 | 7.470E-02 |
| hald-madsen 1 | 2 | 4 | 0 | 122 | 1.767E-01 | 1.105E-02 | 0.000E+00 | 1.767E-01 |
| rosen | 4 | 4 | 0 | 259 | -4.378E+01 | 7.906E-03 | -4.400E+01 | 4.821E-03 |
| hald-madsen 2 | 5 | 42 | 0 | 194 | 3.126E-01 | 7.906E-03 | 1.220E-04 | 3.124E-01 |
| polak 2 | 10 | 2 | 0 | 285 | 5.460E+01 | 7.813E-03 | 5.459E+01 | 1.134E-04 |
| maxq | 20 | 20 | 0 | 7190 | 8.713E-03 | 7.813E-03 | 0.000E+00 | 8.713E-03 |
| maxl | 20 | 40 | 0 | 12111 | 3.028E-03 | 7.813E-03 | 0.000E+00 | 3.028E-03 |
| goffin | 50 | 50 | 0 | 2045 | 0.000E+00 | 7.813E-03 | 0.000E+00 | 0.000E+00 |
| polak 6.1 | 2 | 3 | 0 | 92 | 1.973E+00 | 7.906E-03 | 1.952E+00 | 7.087E-03 |
| polak 6.2 | 20 | 20 | 0 | 5174 | 1.553E-03 | 9.244E-03 | 0.000E+00 | 1.553E-03 |
| polak 6.3 | 4 | 50 | 0 | 138 | 5.467E-01 | 7.813E-03 | 2.637E-03 | 5.426E-01 |
| polak 6.4 | 4 | 102 | 0 | 138 | 5.497E-01 | 7.813E-03 | 2.650E-03 | 5.456E-01 |
| polak 6.5 | 4 | 202 | 0 | 139 | 5.495E-01 | 7.813E-03 | 2.650E-03 | 5.454E-01 |
| polak 6.6 | 3 | 50 | 0 | 104 | 5.441E-01 | 7.813E-03 | 4.500E-03 | 5.372E-01 |
| polak 6.7 | 3 | 102 | 0 | 104 | 5.441E-01 | 7.813E-03 | 4.505E-03 | 5.372E-01 |
| polak 6.8 | 3 | 202 | 0 | 104 | 5.441E-01 | 7.813E-03 | 4.505E-03 | 5.372E-01 |
| polak 6.9 | 2 | 2 | 0 | 88 | 1.161E-01 | 7.812E-03 | 0.000E+00 | 1.161E-01 |
| polak 6.10 | 1 | 25 | 0 | 58 | 1.782E-01 | 1.105E-02 | 1.782E-01 | 6.121E-07 |
| polak 6.11 | 1 | 51 | 0 | 60 | 1.783E-01 | 1.105E-02 | 1.783E-01 | 6.630E-08 |
| polak 6.12 | 1 | 101 | 0 | 61 | 1.784E-01 | 1.105E-02 | 1.784E-01 | 5.382E-07 |
| polak 6.13 | 1 | 501 | 0 | 59 | 1.784E-01 | 1.105E-02 | 1.784E-01 | 1.021E-07 |
| polak 6.14 | 100 | 100 | 0 | 44694 | 3.337E-03 | 7.812E-03 | 0.000E+00 | 3.337E-03 |
| polak 6.15 | 200 | 200 | 0 | 50001 | 1.210E-01 | 1.914E-02 | 0.000E+00 | 1.210E-01 |
| polak 6.16 | 100 | 50 | 0 | 50002 | 1.621E-01 | 2.210E-02 | 0.000E+00 | 1.621E-01 |
| polak 6.17 | 200 | 50 | 0 | 50003 | 1.782E+00 | 3.125E-02 | 0.000E+00 | 1.782E+00 |
| mad 1 | 2 | 3 | 1 | 105 | -3.879E-01 | 1.235E-02 | -3.897E-01 | 1.246E-03 |
| mad 2 | 2 | 3 | 1 | 42 | -3.304E-01 | 1.353E-02 | -3.304E-01 | -9.735E-10 |
| mad 4 | 2 | 3 | 2 | 201 | -4.461E-01 | 1.105E-02 | -4.489E-01 | 1.967E-03 |
| wong 2 | 10 | 6 | 3 | 358 | 2.654E+01 | 1.377E-02 | 2.431E+01 | 8.830E-02 |
| wong 3 | 20 | 14 | 4 | 660 | 1.019E+02 | 1.271E-02 | 1.337E+02 | -2.364E-01 |

Table 3: Numerical performance of Algorithm $DF_{\texttt{mod1}}$

| PROBLEM | n | q | m | nF | $f(\bar{x})$ | $\bar{\mu}$ | $f^\star$ | $\Delta$ |
|---|---|---|---|---|---|---|---|---|
| crescent | 2 | 2 | 0 | 78 | 2.418E-01 | 1.000E+00 | 0.000E+00 | 2.418E-01 |
| polak 1 | 2 | 2 | 0 | 106 | 2.718E+00 | 1.000E+00 | 2.718E+00 | 7.654E-09 |
| lq | 2 | 2 | 0 | 95 | -1.274E+00 | 1.000E+00 | -1.414E+00 | 5.796E-02 |
| mifflin 1 | 2 | 2 | 0 | 65 | -1.000E+00 | 1.000E+00 | -1.000E+00 | 0.000E+00 |
| mifflin 2 | 2 | 2 | 0 | 77 | -8.193E-01 | 1.000E+00 | -1.000E+00 | 9.033E-02 |
| charalambous-conn 1 | 2 | 3 | 0 | 94 | 2.041E+00 | 1.000E+00 | 1.952E+00 | 3.017E-02 |
| charalambous-conn 2 | 2 | 3 | 0 | 81 | 2.223E+00 | 1.000E+00 | 2.000E+00 | 7.435E-02 |
| demyanov-malozemov | 2 | 3 | 0 | 84 | -3.000E+00 | 1.000E+00 | -3.000E+00 | 0.000E+00 |
| ql | 2 | 3 | 0 | 156 | 7.473E+00 | 1.000E+00 | 7.200E+00 | 3.332E-02 |
| hald-madsen 1 | 2 | 4 | 0 | 292 | 8.496E-03 | 1.000E+00 | 0.000E+00 | 8.496E-03 |
| rosen | 4 | 4 | 0 | 515 | -4.356E+01 | 1.000E+00 | -4.400E+01 | 9.842E-03 |
| hald-madsen 2 | 5 | 42 | 0 | 299 | 9.496E-03 | 1.000E+00 | 1.220E-04 | 9.372E-03 |
| polak 2 | 10 | 2 | 0 | 285 | 5.460E+01 | 1.000E+00 | 5.459E+01 | 1.134E-04 |
| maxq | 20 | 20 | 0 | 1858 | 0.000E+00 | 1.000E+00 | 0.000E+00 | 0.000E+00 |
| maxl | 20 | 40 | 0 | 891 | 0.000E+00 | 1.000E+00 | 0.000E+00 | 0.000E+00 |
| goffin | 50 | 50 | 0 | 2045 | 0.000E+00 | 1.000E+00 | 0.000E+00 | 0.000E+00 |
| polak 6.1 | 2 | 3 | 0 | 106 | 2.041E+00 | 1.000E+00 | 1.952E+00 | 3.014E-02 |
| polak 6.2 | 20 | 20 | 0 | 692 | 2.384E-09 | 1.000E+00 | 0.000E+00 | 2.384E-09 |
| polak 6.3 | 4 | 50 | 0 | 1527 | 8.864E-03 | 1.000E+00 | 2.637E-03 | 6.211E-03 |
| polak 6.4 | 4 | 102 | 0 | 2260 | 7.785E-03 | 1.000E+00 | 2.650E-03 | 5.122E-03 |
| polak 6.5 | 4 | 202 | 0 | 1428 | 1.106E-02 | 1.000E+00 | 2.650E-03 | 8.388E-03 |
| polak 6.6 | 3 | 50 | 0 | 262 | 6.592E-03 | 1.000E+00 | 4.500E-03 | 2.083E-03 |
| polak 6.7 | 3 | 102 | 0 | 264 | 8.179E-03 | 1.000E+00 | 4.505E-03 | 3.657E-03 |
| polak 6.8 | 3 | 202 | 0 | 400 | 8.545E-03 | 1.000E+00 | 4.505E-03 | 4.022E-03 |
| polak 6.9 | 2 | 2 | 0 | 91 | 1.162E-01 | 1.000E+00 | 0.000E+00 | 1.162E-01 |
| polak 6.10 | 1 | 25 | 0 | 52 | 1.038E+00 | 1.000E+00 | 1.782E-01 | 7.300E-01 |
| polak 6.11 | 1 | 51 | 0 | 53 | 1.105E+00 | 1.000E+00 | 1.783E-01 | 7.866E-01 |
| polak 6.12 | 1 | 101 | 0 | 51 | 1.139E+00 | 1.000E+00 | 1.784E-01 | 8.150E-01 |
| polak 6.13 | 1 | 501 | 0 | 57 | 1.167E+00 | 1.000E+00 | 1.784E-01 | 8.389E-01 |
| polak 6.14 | 100 | 100 | 0 | 3452 | 3.433E-09 | 1.000E+00 | 0.000E+00 | 3.433E-09 |
| polak 6.15 | 200 | 200 | 0 | 6891 | 3.433E-09 | 1.000E+00 | 0.000E+00 | 3.433E-09 |
| polak 6.16 | 100 | 50 | 0 | 3452 | 5.364E-09 | 1.000E+00 | 0.000E+00 | 5.364E-09 |
| polak 6.17 | 200 | 50 | 0 | 7233 | 1.023E-08 | 1.000E+00 | 0.000E+00 | 1.023E-08 |
| mad 1 | 2 | 3 | 1 | 43 | -3.896E-01 | 1.000E+00 | -3.897E-01 | 5.878E-05 |
| mad 2 | 2 | 3 | 1 | 42 | -3.304E-01 | 1.000E+00 | -3.304E-01 | -9.735E-10 |
| mad 4 | 2 | 3 | 2 | 72 | -4.489E-01 | 1.000E+00 | -4.489E-01 | 4.601E-07 |
| wong 2 | 10 | 6 | 3 | 236 | 2.522E+01 | 1.000E+00 | 2.431E+01 | 3.609E-02 |
| wong 3 | 20 | 14 | 4 | 451 | 1.076E+02 | 1.000E+00 | 1.337E+02 | -1.938E-01 |

Table 4: Numerical performance of Algorithm $DF_{\texttt{mod2}}$

| PROBLEM | n | q | m | nF | $f(\bar{x})$ | $\bar{\mu}$ | $f^{\star}$ | $\Delta$ |
|---|---|---|---|---|---|---|---|---|
| crescent | 2 | 2 | 0 | 79 | 2.693E-03 | 1.000E-02 | 0.000E+00 | 2.693E-03 |
| polak 1 | 2 | 2 | 0 | 106 | 2.718E+00 | 1.000E-02 | 2.718E+00 | 7.654E-09 |
| lq | 2 | 2 | 0 | 142 | -1.412E+00 | 1.000E-02 | -1.414E+00 | 1.072E-03 |
| mifflin 1 | 2 | 2 | 0 | 65 | -1.000E+00 | 1.000E-02 | -1.000E+00 | 0.000E+00 |
| mifflin 2 | 2 | 2 | 0 | 74 | -9.982E-01 | 1.000E-02 | -1.000E+00 | 9.172E-04 |
| charalambous-conn 1 | 2 | 3 | 0 | 130 | 1.953E+00 | 1.000E-02 | 1.952E+00 | 4.080E-04 |
| charalambous-conn 2 | 2 | 3 | 0 | 91 | 2.003E+00 | 1.000E-02 | 2.000E+00 | 1.060E-03 |
| demyanov-malozemov | 2 | 3 | 0 | 84 | -3.000E+00 | 1.000E-02 | -3.000E+00 | 0.000E+00 |
| ql | 2 | 3 | 0 | 148 | 7.203E+00 | 1.000E-02 | 7.200E+00 | 3.656E-04 |
| hald-madsen 1 | 2 | 4 | 0 | 165 | 1.270E-03 | 1.000E-02 | 0.000E+00 | 1.270E-03 |
| rosen | 4 | 4 | 0 | 812 | -4.399E+01 | 1.000E-02 | -4.400E+01 | 3.083E-04 |
| hald-madsen 2 | 5 | 42 | 0 | 856 | 6.762E-03 | 1.000E-02 | 1.220E-04 | 6.639E-03 |
| polak 2 | 10 | 2 | 0 | 285 | 5.460E+01 | 1.000E-02 | 5.459E+01 | 1.134E-04 |
| maxq | 20 | 20 | 0 | 7153 | 5.821E-11 | 1.000E-02 | 0.000E+00 | 5.821E-11 |
| maxl | 20 | 40 | 0 | 9663 | 5.913E-05 | 1.000E-02 | 0.000E+00 | 5.913E-05 |
| goffin | 50 | 50 | 0 | 2045 | 0.000E+00 | 1.000E-02 | 0.000E+00 | 0.000E+00 |
| polak 6.1 | 2 | 3 | 0 | 329 | 1.953E+00 | 1.000E-02 | 1.952E+00 | 3.821E-04 |
| polak 6.2 | 20 | 20 | 0 | 1305 | 2.384E-09 | 1.000E-02 | 0.000E+00 | 2.384E-09 |
| polak 6.3 | 4 | 50 | 0 | 1990 | 8.010E-03 | 1.000E-02 | 2.637E-03 | 5.359E-03 |
| polak 6.4 | 4 | 102 | 0 | 865 | 9.830E-03 | 1.000E-02 | 2.650E-03 | 7.162E-03 |
| polak 6.5 | 4 | 202 | 0 | 2284 | 1.063E-02 | 1.000E-02 | 2.650E-03 | 7.963E-03 |
| polak 6.6 | 3 | 50 | 0 | 590 | 6.429E-03 | 1.000E-02 | 4.500E-03 | 1.921E-03 |
| polak 6.7 | 3 | 102 | 0 | 589 | 7.040E-03 | 1.000E-02 | 4.505E-03 | 2.524E-03 |
| polak 6.8 | 3 | 202 | 0 | 365 | 7.446E-03 | 1.000E-02 | 4.505E-03 | 2.928E-03 |
| polak 6.9 | 2 | 2 | 0 | 88 | 1.161E-01 | 1.000E-02 | 0.000E+00 | 1.161E-01 |
| polak 6.10 | 1 | 25 | 0 | 62 | 1.784E-01 | 1.000E-02 | 1.782E-01 | 1.625E-04 |
| polak 6.11 | 1 | 51 | 0 | 60 | 1.784E-01 | 1.000E-02 | 1.783E-01 | 5.924E-05 |
| polak 6.12 | 1 | 101 | 0 | 61 | 1.784E-01 | 1.000E-02 | 1.784E-01 | 2.368E-05 |
| polak 6.13 | 1 | 501 | 0 | 60 | 1.784E-01 | 1.000E-02 | 1.784E-01 | 1.464E-05 |
| polak 6.14 | 100 | 100 | 0 | 50005 | 3.713E-02 | 1.000E-02 | 0.000E+00 | 3.713E-02 |
| polak 6.15 | 200 | 200 | 0 | 50002 | 8.690E-02 | 1.000E-02 | 0.000E+00 | 8.690E-02 |
| polak 6.16 | 100 | 50 | 0 | 50001 | 1.617E-01 | 1.000E-02 | 0.000E+00 | 1.617E-01 |
| polak 6.17 | 200 | 50 | 0 | 50001 | 6.276E-01 | 1.000E-02 | 0.000E+00 | 6.276E-01 |
| mad 1 | 2 | 3 | 1 | 43 | -3.896E-01 | 1.000E-02 | -3.897E-01 | 5.878E-05 |
| mad 2 | 2 | 3 | 1 | 42 | -3.304E-01 | 1.000E-02 | -3.304E-01 | -9.735E-10 |
| mad 4 | 2 | 3 | 2 | 72 | -4.489E-01 | 1.000E-02 | -4.489E-01 | 4.601E-07 |
| wong 2 | 10 | 6 | 3 | 236 | 2.522E+01 | 1.000E-02 | 2.431E+01 | 3.609E-02 |
| wong 3 | 20 | 14 | 4 | 451 | 1.076E+02 | 1.000E-02 | 1.337E+02 | -1.938E-01 |

Table 5: Numerical performance of Algorithm $\mathrm{DF}_{\texttt{mod3}}$

## References


[1] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, New York, 1999.

[2] D. P. Bertsekas, *Constrained Optimization and Lagrange Multipliers Methods*, Academic Press, New York, 1982.

[3] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983.

[4] E. Polak, J. O. Royset and R. S. Womersley, *Algorithms with Adaptive Smoothing for finite Minimax Problems*, J. of Optim. Theory and Appl., 119 (2003), pp. 459–484.

[5] X. Li, *An entropy-based aggregate method for minimax optimization*, Engineering Optimization, 18 (1997), pp. 277–285.

[6] T. G. Kolda, R. M. Lewis and V. Torczon, *Stationarity results for generating set search for linearly constrained optimization*, Sandia Tech. rep. SAND2003-8550 (2003).

[7] S. Lucidi and M. Sciandrone, *On the global convergence of derivative free methods for unconstrained optimization*, SIAM J. Optimization, 13 (2002), pp. 97–116.

[8] S. Lucidi, M. Sciandrone and P. Tseng, *Objective-derivative-free methods for constrained optimization*, Math. Programming Ser. A, 92 (2002), pp. 37–59.

[9] E. Polak, *Optimization: Algorithms and consistent approximations*, Springer-Verlag, New York, Heidelberg, Berlin, 1997.

[10] S. Xu, *Smoothing method for minimax problems*, Computational Optimization and Appl., 20 (2001), pp. 267–279.

[11] J. W. Bandler and C. Charalambous, *Practical Least pth Optimization of Networks*, IEEE Transactions on Microwave Theory and Thecniques, 20 (1972), pp. 834–840.

[12] J. W. Bandler and C. Charalambous, *Nonlinear Minimax Optimization as a Sequence of Least pth Optimization with Finite Values of p*, International Journal of Systems Sciences, 7 (1976), pp. 377–391.

[13] C. Charalambous, *Acceleration of the Least pth Algorithm for Minimax Optimization with Engineering Applications*, Math. Programming, 17 (1979), pp. 270–297.

[14] G. Di Pillo, L. Grippo and S. Lucidi, *A smooth method for the finite minimax problem*, Math. Programming, 60 (1993), pp. 187–214.

[15] C. Gigola and S. Gomez, *A Regularization Method for Solving Finite Convex Min-Max Problems*, SIAM Journal on Numerical Analysis, 27 (1990), pp. 1621–1634.

[16] R. M. Lewis and V. Torczon, *Pattern search methods for linearly constrained minimization*, SIAM Journal on Optimization, 9 (2000), pp. 917–941.

[17] L. Luksan and J. Vlcek, *Test proplems for nonsmooth unconstrained and linearly constrained optimization*, Tech. Report No. 798, Institute of Computer Science, Academy of Sciences of the Czech Republic (2000).

[18] D. Q. Mayne and E. Polak, *Nondifferentiable Optimization via Adaptive Smoothing*, Journal of Optimization Theory and Appl., 43 (1984), pp. 601–614.

[19] R. A. Polyak, *Smooth Optimization Method for Minimax Problems*, SIAM Journal on Control and Optimization, 26 (1988), pp. 1274–1286.

[20] F. Guerra Vazquez, H. Gunzel and H. T. Jongen, *On Logarithmic Smoothing of the Maximum Function*, Annals of Operations Research, 101 (2001), pp. 209–220.

[21] I. Zang, *A Smoothing Technique for Min-Max Optimization*, Math. Programming, 19 (1980), pp. 61–77.