Mathematical Programming 60 (1993) 187–214
North-Holland

A smooth method for the finite minimax problem

G. Di Pillo and L. Grippo

Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", Roma, Italy

S. Lucidi

Istituto di Analisi dei Sistemi e Informatica del CNR, Roma, Italy

Received 30 July 1990
Revised manuscript received 13 January 1992

We consider unconstrained minimax problems where the objective function is the maximum of a finite number of smooth functions. We prove that, under usual assumptions, it is possible to construct a continuously differentiable function, whose minimizers yield the minimizers of the max function and the corresponding minimum values. On this basis, we can define implementable algorithms for the solution of the minimax problem, which are globally convergent at a superlinear convergence rate. Preliminary numerical results are reported.

Key words: Nonlinear programming, unconstrained optimization, nondifferentiable optimization, minimax problems.

1. Introduction

Many problems of interest can be formulated as finite minimax problems of the form

$$\min_x \max_i f_i(x), \quad x \in \mathbb{R}^n, \quad i = 1, \ldots, m,$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ are smooth functions. This formulation includes, for instance, L_1 and L_∞ approximation problems, solution of systems of nonlinear equations, problems of finding feasible points of systems of inequalities and solution of nonlinear programming problems by means of nondifferentiable penalty functions. We refer, for example, to [5, 6, 7, 24, 25, 40].

The finite minimax problem has been one of the main motivations for the development

Correspondence to: Prof. G. Di Pillo, Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", Via Buonarrota 12, 00185 Roma, Italy.

This research was partially supported by the National Research Program on "Metodi di ottimizzazione per le decisioni", Ministero dell'Università e della Ricerca Scientifica e Tecnologica, Italy.

of nondifferentiable optimization theory and for the construction of algorithms for general nondifferentiable problems such as subgradient methods, bundle methods and cutting plane methods (see [13, 21, 22, 23, 30, 31, 35, 36, 37]). It is known that these techniques can be proved, at most, to be linearly convergent. A superlinearly convergent technique based on the solution of nonsmooth subproblems for the computation of the search direction has been recently proposed in [32].

On the other hand, if we restrict our attention to the finite minimax problem and take into account the particular structure of its nondifferentiability, it is also suitable to make use of smooth optimization methods. Smooth methods include regularization techniques based on the approximation of the max function by means of a smooth function (see, for instance, [2, 39, 14]) and methods based on the solution of the equivalent nonlinear programming problem

$$\begin{array}{ll} \text{minimize} & z \\ \text{subject to} & f_i(x) - z \leqslant 0, \quad i = 1, \dots, m\,. \end{array}$$

Several smooth methods exploit the Kuhn–Tucker optimality conditions for this problem by solving either unconstrained subproblems [1, 3, 4, 32], or constrained subproblems [19, 11, 12, 34, 38].

The main problem arising in the construction of algorithms for the solution of minimax problems is that of conciliating global convergence with an ultimate superlinear convergence rate. A method that achieves this requirement is that considered in [11, 12, 38], based on a trust region strategy. However, this technique may require the global solution of two (possibly nonconvex) quadratic subproblems at each iteration.

Another approach is that of employing a two-stage procedure that combines a globally convergent algorithm for the minimization of the max function with a Newton-type algorithm exploiting approximate second order information on the active set (see [18, 28]).

The approach followed here is based on a nonlinear programming formulation and leads to the construction of a continuously differentiable exact penalty function, whose minimizers are solutions to the minimax problem, for finite values of the penalty coefficient.

Continuously differentiable exact penalty functions for general nonlinear problems, which allow the construction of globally convergent algorithms with superlinear convergence rate, have been already considered in [8] and [27]. However the results established there are based on the assumption that there exists a compact perturbation of the feasible set, which is not satisfied for the nonlinear programming problem equivalent to the minimax problem.

Therefore we consider a continuously differentiable penalty function, which takes into account the particular structure of the max function, and we prove that the correspondence between local or global minimizers of the two functions can be established under the same assumptions that are commonly used in the literature concerning minimax problems.

In Section 2 we introduce the new function and we prove the existence of a global minimizer; in Section 3 we relate stationary points of this function with critical points of the max function; in Section 4 we study the correspondence between local and global minimizers of the two functions. In Section 5 we describe a globally convergent algorithm, based on an automatic adjustment rule for the penalty parameter. Finally, in Section 6 we report some numerical results obtained for a set of well-known test problems.

2. Problem formulation and assumptions

We consider finite minimax problems of the form

$$\min_{x \in \mathbb{R}^n} \phi(x) \tag{1}$$

where

$$\phi(x) := \max_{1 \leqslant i \leqslant m} f_i(x)$$

and $f_i : \mathbb{R}^n \to \mathbb{R}$ are twice continuously differentiable functions. We define the index sets

$$I_A(x) := \{i : f_i(x) = \phi(x)\}, \qquad I_N(x) := \{i : f_i(x) < \phi(x)\} .$$

We suppose that the following assumptions hold.

Assumption 1. There exists a point $x_a \in \mathbb{R}^n$ such that the level set

$$\mathscr{E} := \{x \in \mathbb{R}^n : \phi(x) \leqslant \phi(x_a)\}$$

is compact, and a point $x_b \in \mathscr{E}$ such that $\phi(x_b) < \phi(x_a)$.

Assumption 2. For each $x \in \mathscr{E}$ the vectors

$$\begin{pmatrix} \nabla f_i(x) \\ -1 \end{pmatrix}, \quad i \in I_A(x) ,$$

are linearly independent.

Assumption 1 is introduced in order to ensure the existence of a solution to problem (1) and is obviously satisfied if $\phi(x)$ is radially unbounded, that is $\lim_{\|x\| \to \infty} \phi(x) = \infty$, and x_a is not a solution of problem (1). Assumption 2 is a common assumption in the literature on the minimax problem.

A *critical point* for problem (1) is a point x^* such that:

$$\max_{i \in I_A(x^*)} \nabla f_i(x^*)' d \geqslant 0 \quad \text{for all } d \in \mathbb{R}^n .$$

We recall the following well-known result on minimax problems (see, for instance, [24]), where we use the notation

$$f(x) := (f_1(x), \ldots, f_m(x))' ,$$

$$\nabla f(x) := (\nabla f_1(x), \ldots, \nabla f_m(x)) ,$$

$$F(x) := \mathrm{Diag}_{1 \leqslant i \leqslant m} f_i(x) ,$$

$$u := (1, \ldots, 1)', \quad u \in \mathbb{R}^m .$$

Proposition 2.1. *A point $x^* \in \mathbb{R}^n$ is a critical point for problem (1) if and only if there exists a vector $\lambda^* \in \mathbb{R}^m$ such that*

$$\nabla f(x^*) \lambda^* = 0 ,$$

$$u'\lambda^* = 1\,,$$
$$(F(x^*) - \phi(x^*)I)\lambda^* = 0\,,$$
$$\lambda^* \geq 0\,. \qquad \square$$

Problem (1) is equivalent to the nonlinear programming problem in the $n+1$ variables $x \in \mathbb{R}^n$ and $z \in \mathbb{R}$:

$$\begin{aligned} &\text{minimize} \quad z \\ &\text{subject to} \quad f_i(x) - z \leq 0, \quad i = 1, \dots, m\,, \end{aligned} \tag{2}$$

and the conditions of Proposition 2.1 can be interpreted as the Kuhn–Tucker conditions for problem (2) at the point $(x^*, \phi(x^*))$.

We shall attempt to find local or global solutions of problem (1) by constructing a suitable continuously differentiable exact penalty function for problem (2). This requires the definition of a continuously differentiable multiplier function that yields an estimate of the multiplier vector associated with problem (2) as a function the variables (x, z). More specifically, we define on the set $\mathscr{C} \times \mathbb{R}$ a function $\lambda : \mathbb{R}^{n+1} \to \mathbb{R}^m$ by minimizing with respect to λ the quadratic function

$$\Psi(\lambda) := \|\nabla f(x)\lambda\|^2 + \|(F(x) + zI)\lambda\|^2 + (u'\lambda - 1)^2 + \gamma(x, z)\|\lambda\|^2\,,$$

where

$$\gamma(x, z) := \sum_{i=1}^{m} [\max(0, f_i(x) - z)]^p, \quad p \geq 2\,.$$

The first three terms of the function $\Psi(\lambda)$ are quadratic penalty terms on the Kuhn–Tucker necessary conditions for Problem 2 that are expressed by equalities [15]. The last term has been introduced in [27] in connection with exact penalty functions for general nonlinear programming problems and ensures that the Hessian matrix of $\Psi(\lambda)$ is positive definite also at points where some of the inequalities $f_i(x) - z \leq 0$ are violated.

Let us denote by $M(x, z)$ the Hessian matrix of $\Psi(\lambda)$:

$$M(x, z) := = \nabla f(x)'\nabla f(x) + (F(x) - zI)^2 + uu' + \gamma(x, z)I\,. \tag{3}$$

Now we prove that $M(x, z)$ is uniformly positive definite on $\mathscr{C} \times \mathbb{R}$. We denote by $\mu(x, z)$ the smallest eigenvalue of the matrix $M(x, z)$, that is

$$\mu(x, z) := \min_{\|v\|=1} v'M(x, z)v, \quad v \in \mathbb{R}^m\,.$$

Moreover, by Assumption 1 and the continuity assumption, we can define the numbers

$$f_{\max} := \max_{i=1,\dots,m} \max_{x \in \mathscr{C}} f_i(x), \qquad f_{\min} := \min_{i=1,\dots,m} \min_{x \in \mathscr{C}} f_i(x)\,.$$

Proposition 2.2. *There exists a positive constant $\tilde{\mu}$ such that*

$$\mu(x, z) \geq \tilde{\mu}$$

for all $x \in \mathscr{C}$ and $z \in \mathbb{R}$.

Proof. We prove first that $\mu(x, z) > 0$ for all $x \in \mathscr{C}$ and $z \in \mathbb{R}$. By construction, the matrix $M(x, z)$ is positive semidefinite and hence $\mu(x, z) \geqslant 0$. Now, by contradiction assume that there exist $\bar{x} \in \mathscr{C}$, $\bar{z} \in \mathbb{R}$ and $\bar{v} \in \mathbb{R}^m$ with $\|\bar{v}\| = 1$ such that $\bar{v}' M(\bar{x}, \bar{z}) \bar{v} = 0$. By (3) this implies

$$\nabla f(\bar{x}) \bar{v} = 0 , \tag{4}$$

$$(F(\bar{x}) - \bar{z} I) \bar{v} = 0 , \tag{5}$$

$$u' \bar{v} = 0 , \tag{6}$$

$$\gamma(\bar{x}, \bar{z}) = 0 . \tag{7}$$

By (7) we have that $f_i(\bar{x}) \leqslant \bar{z}$ for $i = 1, \ldots, m$. By (5) we have that $\bar{v}_i = 0$ for i such that $f_i(\bar{x}) < \bar{z}$. Moreover for every index i such that $f_i(\bar{x}) = \bar{z}$, Assumption 2, (4) and (6) imply that $\bar{v}_i = 0$. Thus we get a contradiction to the assumption that $\|\bar{v}\| = 1$, so that

$$\mu(x, z) > 0 \quad \text{for all } (x, z) \in \mathscr{C} \times \mathbb{R} . \tag{8}$$

Now let δ be any positive number and let $\Delta := [f_{\min} - \delta, f_{\max} + \delta]$. It follows that for $x \in \mathscr{C}$, and for $z \notin \Delta$ we have $|f_i(x) - z| > \delta$ for $i = 1, \ldots, m$ and therefore, by (3), we have

$$\mu(x, z) \geqslant \min_{i = 1, \ldots, m} (f_i(x) - z)^2 > \delta^2 .$$

Letting

$$\eta := \min_{x \in \mathscr{C}, z \in \Delta} \mu(x, z) ,$$

from (8) we get $\eta > 0$ and hence we can conclude that $\mu(x, z) \geqslant \tilde{\mu} := \min[\delta^2, \eta] > 0$. □

By Proposition 2.2 the matrix $M(x, z)$ is nonsingular on $\mathscr{C} \times \mathbb{R}$; therefore the unique minimizer $\lambda(x, z)$ of Ψ is given by

$$\lambda(x, z) = M(x, z)^{-1} u . \tag{9}$$

The main properties of the multiplier function λ are given in the next two propositions.

Proposition 2.3. *Let $\tilde{\mu} > 0$ be the constant introduced in Proposition* 2.2; *then the multiplier function λ satisfies the bound*

$$\|\lambda(x, z)\| \leqslant \sqrt{m} / \tilde{\mu} \tag{10}$$

for all $x \in \mathscr{C}$ and all $z \in \mathbb{R}$.

Proof. Inequality (10) follows immediately from (9) and Proposition 2.2. □

Proposition 2.4. *Let $x^* \in \mathscr{C}$ be a critical point for problem* (1) *and let $z^* = \phi(x^*)$. Then the multiplier function λ satisfies the condition*

$$\lambda(x^*, z^*) = \lambda^* , \tag{11}$$

where λ^* is the multiplier vector introduced in Proposition 2.1.

Proof. By Proposition 2.1, there exists a vector $\lambda^* \in \mathbb{R}^m$ such that $\lambda_i^* = 0$ for $i \notin I_A(x^*)$ and

$$\sum_{i \in I_A(x^*)} \begin{pmatrix} \nabla f_i(x^*) \\ -1 \end{pmatrix} \lambda_i^* = \begin{pmatrix} 0 \\ -1 \end{pmatrix} .$$

As $\gamma(x^*, z^*) = 0$, this implies that $\Psi(\lambda^*) = 0$ and hence that λ^* is a global minimizer of $\Psi(\lambda)$. On the other hand, $\lambda(x^*, z^*)$ is the unique global minimum point of $\Psi(\lambda)$ and thus $\lambda(x^*, z^*) = \lambda^*$. $\square$

Let

$$\mathscr{E} := \{x \in \mathbb{R}^n \colon \phi(x) < \phi(x_a)\} ,$$

We have that $\mathscr{E}$ is contained in the interior $\mathring{\mathscr{E}}$ of $\mathscr{E}$ and that $x_b \in \mathscr{E}$. On the set $\mathscr{E} \times \mathbb{R}$ we can now define the following function:

$$P(x, z; \varepsilon) := z + \lambda(x, z)'c(x, z; \varepsilon) + \frac{1}{\varepsilon} c(x, z; \varepsilon)'B(x)^{-1}c(x, z; \varepsilon) , \tag{12}$$

where $\varepsilon > 0$ and

$$\begin{aligned} &B(x) := \mathrm{Diag}(b_i(x)), \qquad b_i(x) := \phi(x_a) - f_i(x), \\ &c_i(x, z; \varepsilon) := f_i(x) - z + y_i^2(x, z; \varepsilon), \\ &y_i(x, z; \varepsilon) := \{-\min[0, f_i(x) - z + \tfrac{1}{2}\varepsilon b_i(x)\lambda_i(x, z)]\}^{1/2} . \end{aligned} \tag{13}$$

The expression of $P(x, z; \varepsilon)$ can be derived by replacing the multiplier vector λ in the augmented Lagrangian function of Hestenes–Powell–Rockafellar, with the continuously differentiable multiplier function $\lambda(x, z)$ and by weighting the penalty term by means of the matrix $B(x)^{-1}$. The elements $b_i(x)$ of $B(x)$ define shifted barrier terms on the values of the functions $f_i(x)$ and this ensures that the x-component of the points produced in the unconstrained minimization of P remains in the compact level set $\mathscr{E}$. It will be apparent in the sequel that this feature constitutes an essential requirement for establishing properties of exactness of the function P and for proving global convergence of computational algorithms.

A similar approach has been described in [8] with reference to general nonlinear programming problems; the peculiar features of the function P considered here are the different structure of the multiplier function and the special form of the barrier terms. In particular, we note that the function P may have unbounded level sets in $\mathscr{E} \times \mathbb{R}$ for some values of ε and this requires an analysis of the dependence of the level sets of P on the penalty parameter.

It can be verified that the function P can be rewritten into the form

$$P(x, z; \varepsilon) = z + \sum_{i=1}^{m} \Bigg(\lambda_i(x, z)(f_i(x) - z) + \frac{1}{\varepsilon b_i(x)} (f_i(x) - z)^2 - \frac{1}{\varepsilon b_i(x)} [\min(0, f_i(x) - z + \tfrac{1}{2}\varepsilon b_i(x)\lambda_i(x, z))]^2 \Bigg) , \tag{14}$$

which shows that P is continuously differentiable on $\mathscr{E} \times \mathbb{R}$.

Let us define the index sets

$$\begin{aligned} &I_0(x, z; \varepsilon) := \{i \colon f_i(x) - z + \tfrac{1}{2}\varepsilon b_i(x)\lambda_i(x, z) \geqslant 0\} , \\ &I_-(x, z; \varepsilon) := \{i \colon f_i(x) - z + \tfrac{1}{2}\varepsilon b_i(x)\lambda_i(x, z) < 0\} . \end{aligned}$$

Then from (14) it follows that P can also be put into the form

$$P(x, z; \varepsilon) = z + \sum_{i \in I_0(x,z;\varepsilon)} \left(\lambda_i(x, z)(f_i(x) - z) + \frac{1}{\varepsilon b_i(x)} (f_i(x) - z)^2 \right)$$
$$- \tfrac{1}{4}\varepsilon \sum_{i \in I_-(x,z;\varepsilon)} b_i(x) \lambda_i(x, z)^2 . \tag{15}$$

We denote by $\mathscr{F}$ the intersection of the set $\mathscr{E} \times \mathbb{R}$ with the feasible set for problem (2), that is

$$\mathscr{F} := \{(x, z) \in \mathscr{E} \times \mathbb{R}: f_i(x) - z \leqslant 0, i = 1,..., m\} .$$

The next proposition establishes useful upper and lower bounds on P.

Proposition 2.5. *Let ε be any given positive number; then:*
(i) *$P(x, z; \varepsilon) \leqslant z$ for every $(x, z) \in \mathscr{F}$;*
(ii) *$P(x, z; \varepsilon) \geqslant z - (\varepsilon m^2/(4\tilde{\mu}^2))(\phi(x_a) - f_{\min})$ for every $(x, z) \in \mathscr{E} \times \mathbb{R}$.*

Proof. Let $(x, z) \in \mathscr{F}$. For $i \in I_0(x, z; \varepsilon)$ we have that

$$0 \leqslant f_i(x) - z + \tfrac{1}{2}\varepsilon b_i(x)\lambda_i(x, z) \leqslant \tfrac{1}{2}(f_i(x) - z + \varepsilon b_i(x)\lambda_i(x, z)) ,$$

so that, as $f_i(x) - z \leqslant 0$ and $b_i(x) > 0$, for the ith term of the first summation in (15) we obtain

$$\lambda_i(x, z)(f_i(x) - z) + \frac{(f_i(x) - z)^2}{\varepsilon b_i(x)} \leqslant 0 .$$

As the last term of (15) is nonpositive, assertion (i) is proved. By (15) we have also

$$P(x, z; \varepsilon) \geqslant z - \tfrac{1}{4}\varepsilon \sum_{i=1}^{m} b_i(x) \lambda_i(x, z)^2 ,$$

so that assertion (ii) follows from Proposition 2.3. $\square$

Let us define the level set

$$\mathscr{L}_\varepsilon := \{(x, z) \in \mathscr{E} \times \mathbb{R}: P(x, z; \varepsilon) \leqslant \phi(x_b)\} ,$$

where x_b is the point considered in Assumption 1.

As $(x_b, \phi(x_b)) \in \mathscr{F}$, by Proposition 2.5 we have $P(x_b, \phi(x_b); \varepsilon) \leqslant \phi(x_b)$, so that $(x_b, \phi(x_b)) \in \mathscr{L}_\varepsilon$.

Now we prove the following theorem, which is the main result in this section.

Theorem 2.6. *Let ε_0 be a given positive number and let $\hat{\varepsilon} = \min[\varepsilon_0, \varepsilon_1]$ where $\varepsilon_1 > 0$ satisfies the inequality*

$$\varepsilon_1 < \frac{4\tilde{\mu}^2}{m^2} \frac{\phi(x_a) - \phi(x_b)}{\phi(x_a) - f_{\min}} .$$

Then:
(i) *there exist numbers z_L and z_U, with $z_L < z_U$, such that, for every $\varepsilon \in (0, \varepsilon_0]$, we have*

$$\mathscr{L}_\varepsilon \subset \mathscr{E} \times [z_L, z_U] ;$$

(ii) *for every $\varepsilon \in (0, \hat{\varepsilon}]$ the set $\mathscr{L}_\varepsilon$ is compact and the function P admits a global minimum point on $\mathscr{E} \times \mathbb{R}$.*

Proof. By using Proposition 2.5(ii) we have that $P(x, z; \varepsilon) > \phi(x_b)$ for all $x \in \mathscr{E}$, $z > z_U$ and $\varepsilon \in (0, \varepsilon_0]$ where

$$z_U := \phi(x_b) + \frac{m^2 \varepsilon_0}{4\tilde{\mu}^2} (\phi(x_a) - f_{\min}) .$$

Recalling Proposition 2.3 we have that

$$|\lambda_i(x, z)| \leqslant \sqrt{m}/\tilde{\mu}$$

for all $i = 1, \ldots, m$ and for all $(x, z) \in \mathscr{E} \times \mathbb{R}$; moreover by definition of $b_i(x)$ and $f_{\min}$ we have $b_i(x) \leqslant \phi(x_a) - f_{\min}$, for all $i = 1, \ldots, m$ and for all $x \in \mathscr{E}$. Therefore letting

$$\tilde{z} := f_{\min} - \tfrac{1}{2}\varepsilon_0(\phi(x_a) - f_{\min})\sqrt{m}/\tilde{\mu} ,$$

we have that

$$I_-(x, z; \varepsilon) = \emptyset \quad \text{for all } x \in \mathscr{E} \text{ and } z < \tilde{z} .$$

Then, if $z < \tilde{z}$ by (15) we can write

$$P(x, z; \varepsilon) \geqslant z + \sum_{i=1}^{m} \left(\lambda_i(x, z)(f_i(x) - z) + \frac{1}{\varepsilon_0(\phi(x_a) - f_{\min})} (f_i(x) - z)^2 \right),$$

so that, as $f_i(x)$ and $\lambda_i(x, z)$ are uniformly bounded there must exist $z_L \leqslant \tilde{z}$ such $P(x, z; \varepsilon) > \phi(x_b)$ for all $x \in \mathscr{E}$, $z < z_L$ and $\varepsilon \in (0, \varepsilon_0]$.

It can be concluded that $z \in [z_L, z_U]$ for all $(x, z) \in \mathscr{E} \times \mathbb{R}$ such that $P(x, z; \varepsilon) \leqslant \phi(x_b)$, and this proves that assertion (i) holds. In order to establish assertion (ii), we shall prove that the set $\mathscr{L}_\varepsilon$ is compact for sufficiently small values of ε. Let us suppose that ε is in $(0, \hat{\varepsilon}]$. As $\mathscr{E}$ is bounded, assertion (i) implies that $\mathscr{L}_\varepsilon$ is bounded and therefore we must prove that, for each $\varepsilon \in (0, \hat{\varepsilon}]$, the set $\mathscr{L}_\varepsilon$ is also closed.

Let $\{(x_k, z_k)\}$ be a sequence such that $(x_k, z_k) \in \mathscr{L}_\varepsilon$ and let $(\bar{x}, \bar{z})$ be a limit point of this sequence. By definition of $\mathscr{L}_\varepsilon$ we have

$$P(x_k, z_k; \varepsilon) \leqslant \phi(x_b) \quad \text{for all } k . \tag{16}$$

If $\bar{x} \in \mathscr{E}$, this implies that $(\bar{x}, \bar{z}) \in \mathscr{L}_\varepsilon$ by the continuity assumptions. Therefore, let us assume, by contradiction, that $\bar{x} \in \partial\mathscr{E}$. Define the index set

$$J := \{i \in \{1, \ldots, m\} : f_i(\bar{x}) = \phi(x_a)\} ;$$

then if $\bar{x} \in \partial\mathscr{E}$ we have

$$f_i(\bar{x}) = \phi(x_a) \quad \text{for } i \in J ,$$
$$f_i(\bar{x}) < \phi(x_a) \quad \text{for } i \notin J ,$$

and hence $\phi(\bar{x}) = \phi(x_a)$. Now, by (16) we have

$$\limsup_{k\to\infty} P(x_k, z_k; \varepsilon) \leqslant \phi(x_b) .$$

This implies, by (12), $c_i(\bar{x}, \bar{z}; \varepsilon) = 0$, for $i \in J$, and hence, recalling (13), we have $f_i(\bar{x}) - \bar{z} \leqslant 0$, for all $i \in J$, which yields, in turn, that

$$f_i(\bar{x}) - \bar{z} \leqslant 0 \quad \text{for all } i \in \{1, ..., m\} ,$$

so that $\bar{z} \geqslant \phi(\bar{x}) = \phi(x_a)$. By Proposition 2.5(ii) we have

$$P(x_k, z_k; \varepsilon) \geqslant z_k - \frac{\varepsilon m^2}{4\tilde{\mu}^2} (\phi(x_a) - f_{\min}) .$$

By definition of $\hat{\varepsilon}$, this yields in the limit

$$\liminf_{k\to\infty} P(x_k, z_k; \varepsilon) > \bar{z} - (\phi(x_a) - \phi(x_b)) .$$

Then, using (16) and recalling that $\bar{z} \geqslant \phi(x_a)$, we obtain

$$\phi(x_b) \geqslant \liminf_{k\to\infty} P(x_k, z_k; \varepsilon) > \phi(x_b) ,$$

which establishes the contradiction. □

3. Stationary points of P

In this section we relate critical points of problem (1) with stationary points of P. First of all, we give an expression for the gradient ∇P. Noting that we can treat formally the vector function $y(x, z; \varepsilon)$ as a constant vector in the evaluation of the derivatives of P, the components $\nabla_x P$ and $\nabla_z P$ of the gradient ∇P are obtained as

$$\begin{aligned} \nabla_x P(x, z; \varepsilon) = {} & \nabla f(x)\lambda(x, z) + \nabla_x \lambda(x, z) c(x, z; \varepsilon) \\ & + \frac{2}{\varepsilon} \nabla f(x) B(x)^{-1} c(x, z; \varepsilon) \\ & + \frac{1}{\varepsilon} \nabla f(x) C(x, z; \varepsilon) B(x)^{-2} c(x, z; \varepsilon) , \end{aligned} \tag{17}$$

$$\begin{aligned} \nabla_z P(x, z; \varepsilon) = {} & 1 - \lambda(x, z)' u + \nabla_z \lambda(x, z) c(x, z; \varepsilon) \\ & - \frac{2}{\varepsilon} u' B(x)^{-1} c(x, z; \varepsilon) , \end{aligned} \tag{18}$$

where

$$C(x, z; \varepsilon) := \mathrm{Diag}(c_i(x, z; \varepsilon)) .$$

It can be verified that the derivatives of the multiplier function are given by

$$\nabla_x \lambda(x, z)' = -M(x, z)^{-1} \Bigg[\nabla f(x)' \sum_{i=1}^{m} \lambda_i(x, z) \nabla^2 f_i(x)$$

$$+ \sum_{i=1}^{m} e_i \lambda(x, z)' \nabla f(x)' \nabla^2 f_i(x)$$

$$+ 2\Lambda(x, z)(F(x) - zI) \nabla f(x)'$$

$$+ p\lambda(x, z) \sum_{i=1}^{m} [\max(0, f_i(x) - z)]^{p-1} \nabla f_i(x)' \Bigg],$$

$$\nabla_z \lambda(x, z)' = M(x, z)^{-1} \left[F(x) - \left(2z - p \sum_{i=1}^{m} [\max(0, f_i(x) - z)]^{p-1} \right) I \right] \lambda(x, z),$$

where e_i denotes the ith column of the identity matrix and

$$\Lambda(x, z) := \mathrm{Diag}(\lambda_i(x, z)).$$

Now we can prove the following theorem.

Theorem 3.1. *Let $x^* \in \mathscr{E}$ be a critical point for problem* (1), *and let $z^* = \phi(x^*)$. Then for every $\varepsilon > 0$ we have*:

(a) $c(x^*, z^*; \varepsilon) = 0$;

(b) $\nabla P(x^*, z^*; \varepsilon) = 0$;

(c) $P(x^*, z^*; \varepsilon) = \phi(x^*)$.

Proof. Recalling Proposition 2.4 we have that $\lambda^* = \lambda(x^*, z^*)$ satisfies the conditions stated in Proposition 2.1. Now, by assumption we have $f_i(x^*) - z^* \leqslant 0$. If $f_i(x^*) - z^* = 0$ then from Proposition 2.1, the definition of y_i and the assumption that $x^* \in \mathscr{E}$ we have $y_i(x^*, z^*; \varepsilon) = 0$ and, hence, $c_i(x^*, z^*; \varepsilon) = 0$; if $f_i(x^*) - z^* < 0$ by Proposition 2.1 we have $\lambda_i(x^*, z^*) = 0$ so that $y_i^2(x^*, z^*; \varepsilon) = -f_i(x^*) + z^*$ and again $c_i(x^*, z^*; \varepsilon) = 0$. This proves (a); (b) is a consequence of (a) and Proposition 2.1; finally (c) follows from (a) and the definition of P. □

In order to prove a converse of Theorem 3.1 we need some preliminary results that are stated in the next propositions.

Proposition 3.2. *Let $(x^*, z^*) \in \mathscr{E} \times \mathbb{R}$ be a stationary point of P and assume that $c(x^*, z^*; \varepsilon) = 0$. Then x^* is a critical point for problem* (1) *and $z^* = \phi(x^*)$.*

Proof. By (17) and (18) and the assumptions made we have

$$\nabla f(x^*) \lambda(x^*, z^*) = 0, \qquad u' \lambda(x^*, z^*) = 1.$$

Now, by definition of $\lambda(x^*, z^*)$, noting that $c(x^*, z^*; \varepsilon)=0$ implies that $\gamma(x^*, z^*)=0$, we have

$$(F(x^*)-z^*I)\lambda(x^*, z^*)=0\,. \tag{19}$$

Then, if $f_i(x^*)-z^*=0$ for some i, we have $y_i^2(x^*, z^*; \varepsilon)=0$ and, hence, by definition of y_i and (19) we get $\lambda_i(x^*, z^*)\geqslant 0$ for all i. Moreover, $c(x^*, z^*; \varepsilon)=0$ implies $f_i(x^*)\leqslant z^*$, for all i. Then, as $\lambda(x^*, z^*)\geqslant 0$ and $u'\lambda(x^*, z^*)=1$, by (19) there must exist an index i such that $z^*=f_i(x^*)$. Hence we have $z^*=\phi(x^*)$, so that, letting $\lambda^*=\lambda(x^*, z^*)$, the assertion follows from Proposition 2.1. □

Proposition 3.3. *Let ε_0 be a given positive number, then*:

(i) *there exists a constant $K_0\geqslant 0$ such that, if $\varepsilon\in(0, \varepsilon_0]$ and $(x, z)\in\mathscr{L}_\varepsilon$, we have*

$$z\geqslant\phi(x)-\varepsilon^{1/2}K_0\,; \tag{20}$$

(ii) *if $\{\varepsilon_k\}$ is a sequence of numbers in $(0, \varepsilon_0]$ converging to zero and $\{(x_k, z_k)\}$ is a sequence of points in $\mathscr{L}_{\varepsilon_k}$ then $\{(x_k, z_k)\}$ admits a limit point and every limit point belongs to $\mathscr{F}$.*

Proof. Recalling (12), since $(x, z)\in\mathscr{L}_\varepsilon$, we have

$$\begin{aligned}(\phi(x_a)-f_{\min})^{-1}\|c(x, z; \varepsilon)\|^2 &\leqslant c(x, z; \varepsilon)B(x)^{-1}c(x, z; \varepsilon)\\ &\leqslant\varepsilon[\phi(x_b)-z-\lambda(x, z)'c(x, z; \varepsilon)]\,.\end{aligned}$$

Therefore, by Theorem 2.6 (i), from (21) we obtain

$$\|c(x, z; \varepsilon)\|\leqslant\varepsilon^{1/2}K_0\,, \tag{22}$$

where

$$K_0 := [M_0(\phi(x_a)-f_{\min})]^{1/2}\,,$$
$$M_0 := \max_{\mathscr{C}\times[z_L, z_U]\times[0, \varepsilon_0]}[\phi(x_b)-z-\lambda(x, z)'c(x, z; \varepsilon)]\,.$$

By (22) we have, for $i=1, \ldots, m$,

$$f_i(x)-z\leqslant f_i(x)-z+y_i(x, z; \varepsilon)^2\leqslant\|c(x, z; \varepsilon)\|\leqslant\varepsilon^{1/2}K_0\,,$$

which proves part (i).

By Theorem 2.6(i) we have that $\mathscr{L}_{\varepsilon_k}$ is contained in the compact set $\mathscr{C}\times[z_L, z_U]$ and therefore the sequence $\{(x_k, z_k)\}$ admits a limit point in $\mathscr{C}\times[z_L, z_U]$. Let now $(\hat{x}, \hat{z})$ be a limit point of $\{(x_k, z_k)\}$. By (20) we have

$$z_k\geqslant\phi(x_k)-\varepsilon_k^{1/2}K_0\,,$$

and, hence, taking limits over the subsequence converging to $(\hat{x}, \hat{z})$ we obtain $\hat{z}\geqslant\phi(\hat{x})$ which establishes the result. □

Proposition 3.4. *Let $(\hat{x}, \hat{z}) \in \mathscr{F}$; then there exist numbers $\varepsilon(\hat{x}, \hat{z}) > 0$, $p(\hat{x}, \hat{z}) > 0$ and $\sigma(\hat{x}, \hat{z}) > 0$ such that*

$$\varepsilon \| \nabla f(x)' \nabla_x P(x, z; \varepsilon) - u \nabla_z P(x, z; \varepsilon) \| \geqslant \sigma(\hat{x}, \hat{z}) \| c(x, z; \varepsilon) \| \tag{23}$$

for all $\varepsilon \in (0, \varepsilon(\hat{x}, \hat{z})]$ and all (x, z) such that $\| (x, z) - (\hat{x}, \hat{z}) \| \leqslant \rho(\hat{x}, \hat{z})$.

Proof. By definition of $y(x, z; \varepsilon)$ and $c(x, z; \varepsilon)$, we have

$$Y^2(x, z; \varepsilon) \lambda(x, z) = -\frac{2}{\varepsilon} Y^2(x, z; \varepsilon) B(x)^{-1} c(x, z; \varepsilon) , \tag{24}$$

where

$$Y(x, z; \varepsilon) := \mathrm{Diag}(y_i(x, z; \varepsilon)) .$$

Recalling the definition of $\lambda(x, z)$ and using (24) we can write

$$\begin{aligned} &(\nabla f(x)' \nabla f(x) + uu') \lambda(x, z) - u \\ &\quad = -((F(x) - zI)^2 + \gamma(x, z) I) \lambda(x, z) \\ &\quad = -(F(x) - zI)(F(x) - zI + Y^2(x, z; \varepsilon)) \lambda(x, z) \\ &\qquad + (F(x) - zI) Y^2(x, z; \varepsilon) \lambda(x, z) - \gamma(x, z) \lambda(x, z) \\ &\quad = -(F(x) - zI)\left(\Lambda(x, z) + \frac{2}{\varepsilon} Y^2(x, z; \varepsilon) B(x)^{-1} \right) c(x, z; \varepsilon) \\ &\qquad - \gamma(x, z) \lambda(x, z) \end{aligned} \tag{25}$$

Now, taking into account (17) and (18), we obtain

$$\begin{aligned} &\nabla f(x)' \nabla_x P(x, z; \varepsilon) - u \nabla_z P(x, z; \varepsilon) \\ &\quad = (\nabla f(x)' \nabla f(x) + uu') \lambda(x, z) - u \\ &\qquad + (\nabla f(x)' \nabla_x \lambda(x, z) - u \nabla_z \lambda(x, z)) c(x, z; \varepsilon) \\ &\qquad + \frac{1}{\varepsilon} \nabla f(x)' \nabla f(x) (2B(x)^{-1} + C(x, z; y\ \varepsilon) B(x)^{-2}) c(x, z; \varepsilon) \\ &\qquad + \frac{2}{\varepsilon} uu' B(x)^{-1} c(x, z; \varepsilon) . \end{aligned} \tag{26}$$

Then, by substituting (25) into (26), we have

$$\begin{aligned} &\varepsilon [\nabla f(x)' \nabla_x P(x, z; \varepsilon) - u \nabla_z P(x, z; \varepsilon)] \\ &\quad = [\varepsilon H(x, z; \varepsilon) + K(x, z; \varepsilon)] c(x, z; \varepsilon) - \varepsilon \gamma(x, z) \lambda(x, z) , \end{aligned} \tag{27}$$

where

$$H(x, z; \varepsilon) := -(F(x) - zI)\Lambda(x, z) + \nabla f(x)'\nabla_x \lambda(x, z) - u\nabla_z \lambda(x, z) ,$$

$$\begin{aligned} K(x, z; \varepsilon) := & -2(F(x) - zI)Y^2(x, z; \varepsilon)B(x)^{-1} \\ & + \nabla f(x)'\nabla f(x)(2B(x)^{-1} + C(x, z; \varepsilon)B(x)^{-2}) \\ & + 2uu'B(x)^{-1} . \end{aligned}$$

As $(\hat{x}, \hat{z}) \in \mathscr{F}$ we have

$$Y^2(\hat{x}, \hat{z}); 0) = -(F(\hat{x}) - \hat{z}I), \qquad c(\hat{x}, \hat{z}; 0) = 0, \qquad \gamma(\hat{x}, \hat{z}) = 0 ,$$

so that

$$K(\hat{x}, \hat{z}; 0) = 2M(\hat{x}, \hat{z})B(\hat{x})^{-1} .$$

Since the matrix $K(\hat{x}, \hat{z}; 0)$ is positive definite, by continuity there exist numbers $\varepsilon_1(\hat{x}, \hat{z}) > 0$, $\rho(\hat{x}, \hat{z}) > 0$ and $\sigma_1(\hat{x}, \hat{z}) > 0$ such that

$$\| [\varepsilon H(x, z; \varepsilon) + K(x, z; \varepsilon)] c(x, z; \varepsilon) \| \geqslant \sigma_1(\hat{x}, \hat{z}) \| c(x, z; \varepsilon) \| \tag{28}$$

for all $\varepsilon \in [0, \varepsilon_1(\hat{x}, \hat{z})]$ and all (x, z) such that $\| (x, z) - (\hat{x}, \hat{z}) \| \geqslant \rho(\hat{x}, \hat{z})$.

Recalling the definitions of $\gamma(x, z)$ and $c(x, z; \varepsilon)$ and taking into account the equivalence of norms on $\mathbb{R}^n$ we have for some positive constant θ,

$$\gamma(x, z) \leqslant \theta \left[\sum_{i=1}^{m} \max(0, f_i(x) - z)^2 \right]^{p/2} \leqslant \theta \| c(x, z; \varepsilon) \|^p . \tag{29}$$

Now, by (27), (28) and (29) the assertion of the proposition follows. □

We can now prove the following theorem which can be considered a converse of Theorem 3.1.

Theorem 3.5. *There exists an $\varepsilon^* > 0$ such that, for all $\varepsilon \in (0, \varepsilon^*]$, if $(x_\varepsilon, z_\varepsilon) \in \mathscr{L}_\varepsilon$ is a stationary point of $P(x, z; \varepsilon)$ then x_ε is a critical point of problem (1) and $z_\varepsilon = \phi(x_\varepsilon)$.*

Proof. The proof is by contradiction. Assume that the assertion is false. Then, for every integer k, there exists an $\varepsilon_k \geqslant 1/k$ and a point $(x_k, z_k) \in \mathscr{L}_{\varepsilon_k}$ such that $\nabla P(x_k, z_k; \varepsilon_k) = 0$ and $P(x_k, z_k; \varepsilon_k) \leqslant \phi(x_b)$, but x_k is not a critical point of problem (1) or $z_k \neq \phi(x_k)$.

By Proposition 3.3(ii) we obtain that the sequence $\{(x_k, z_k)\}$ admits a limit point $(\hat{x}, \hat{z}) \in \mathscr{F}$. Therefore, by Proposition 3.4, for sufficiently large values of k there exists (x_k, z_k) satisfying $c(x_k, z_k; \varepsilon_k) = 0$ so that, by Proposition 3.2, x_k is a critical point of problem (1) and $z_k = \phi(x_k)$; thus we get a contradiction. □

4. Minimum points of P

In this section we analyze the correspondence between minimizers of P and minimizers of ϕ. We assume that ε^* is the threshold value of the penalty parameter considered in Theorem 3.5, that $\hat{\varepsilon}$ is the value defined in Theorem 2.6 and we define $\bar{\varepsilon} := \min[\hat{\varepsilon}, \varepsilon^*]$.

In the next theorem we give a result pertaining to global minimizers.

Theorem 4.1. *For every $\varepsilon \in (0, \bar{\varepsilon}]$ we have that:*

(a) *if $(x_\varepsilon, z_\varepsilon)$ is a global minimum point of $P(x, z; \varepsilon)$ on $\mathscr{L}_\varepsilon$, then x_ε is a global minimum point of problem (1) and $z_\varepsilon = \phi(x_\varepsilon)$;*

(b) *if x^* is a global minimum point of $\phi(x)$ then $(x^*, \phi(x^*))$ is a global minimum point of P on $\mathscr{E} \times \mathbb{R}$.*

Proof. By Theorem 2.6, for every $\varepsilon \in (0, \hat{\varepsilon}]$, the function $P(x, z; \varepsilon)$ admits a global minimum point x_ε, z_ε on $\mathscr{L}_\varepsilon$. Therefore we have $\nabla P(x_\varepsilon, z_\varepsilon; \varepsilon) = 0$ and, hence, Theorem 3.5 ensures that, for every $\varepsilon \in (0, \varepsilon^*]$, x_ε is a critical point of ϕ and that $z_\varepsilon = \phi(x_\varepsilon)$. Then, by Theorem 3.1, we have $P(x_\varepsilon, z_\varepsilon; \varepsilon) = \phi(x_\varepsilon)$. On the other hand, if x^* is a global minimum point of $\phi(x)$ then x^* is a critical point of ϕ and $\phi(x^*) \leqslant \phi(x_b)$, so that x^* belongs to $\mathscr{E}$. Therefore by Theorem 3.1 we have $\phi(x^*) = P(x^*, \phi(x^*); \varepsilon)$.

It can be concluded that for every $\varepsilon \in (0, \bar{\varepsilon}]$ the functions ϕ and P take the same value in correspondence to their respective global minimizers and this establishes (a) and (b). $\square$

A straightforward consequence of Theorems 3.5 and 4.1 is that, if problem (1) is convex, then for $\varepsilon \in (0, \bar{\varepsilon}]$ every stationary point $(x_\varepsilon, z_\varepsilon)$ of P is a global minimizer of P on $\mathscr{E} \times \mathbb{R}$ and x_ε is a global minimizer of ϕ.

In the non convex case it is important to show that local minimizers of P correspond to local minimizers of ϕ. This is proved in the following proposition.

Theorem 4.2. *For every $\varepsilon \in (0, \varepsilon^*]$, if $(x_\varepsilon, z_\varepsilon) \in \mathscr{L}_\varepsilon$ is a local minimum point of $P(x, z; \varepsilon)$, then x_ε is a local minimum point of problem (1) and $z_\varepsilon = \phi(x_\varepsilon)$.*

Proof. Let $(x_\varepsilon, z_\varepsilon) \in \mathscr{L}_\varepsilon$ be a local minimum point of $P(x, z; \varepsilon)$, then, by Theorem 3.5, we have that x_ε is a critical point of problem (1) and that $z_\varepsilon = \phi(x_\varepsilon)$. By Theorem 3.1 we have $P(x_\varepsilon, z_\varepsilon; \varepsilon) = \phi(x_\varepsilon)$, so that, as $(x_\varepsilon, z_\varepsilon)$ is a local minimizer of P, there exists a neighborhood $U(x_\varepsilon, z_\varepsilon)$ such that $\phi(x_\varepsilon) \leqslant P(x, z; \varepsilon)$ for all $(x, z) \in U(x_\varepsilon, z_\varepsilon)$. By Proposition 2.5(i) this implies

$$\phi(x_\varepsilon) \leqslant P(x, z; \varepsilon) \leqslant z \quad \text{for all } (x, z) \in U(x_\varepsilon, z_\varepsilon) \cap \mathscr{F} . \tag{30}$$

As $z_\varepsilon = \phi(x_\varepsilon)$, by continuity we can find a neighborhood $V(x_\varepsilon) \subset \mathscr{E}$ such that $(x,$

$\phi(x)) \in U(x_\varepsilon, z_\varepsilon)$ for all $x \in V(x_\varepsilon)$. Moreover, by definition of ϕ, we have $f_i(x) - \phi(x) \leqslant 0$ for $i = 1, ..., m$, and hence we have $(x, \phi(x)) \in \mathscr{F}$ for all $x \in V(x_\varepsilon)$. Therefore, from (30) and Proposition 2.5(i), we get

$$\phi(x_\varepsilon) \leqslant P(x, \phi(x); \varepsilon) \leqslant \phi(x) \quad \text{for all } x \in V(x_\varepsilon) ,$$

which concludes the proof. □

If the functions f_i are three times continuously differentiable and $p \geqslant 3$, second order optimality results can be established under the following assumption.

Assumption 3. If x^* is a critical point for problem (1) and λ^* is the corresponding multiplier vector introduced in Proposition 2.1, then $\lambda_i^* > 0$ for all i such that $f_i(x^*) = \phi(x^*)$.

Assumption 3 can be interpreted as the *strict complementarity condition* for problem (2), provided that $\phi(x^*) = z^*$.

If $x^* \in \mathscr{E}$ is a critical point for problem (1) and Assumption 3 holds at x^* then, for every given $\varepsilon_0 > 0$ there exists a neighbourhood $U^* \subset \mathscr{E} \times \mathbb{R}$ of $(x^*, \phi(x^*))$ such that for all $(x, z) \in U^*$ and for all $\varepsilon \in (0, \varepsilon_0]$ we have

$$I_0(x, z; \varepsilon) = I_A(x^*), \qquad I_-(x, z; \varepsilon) = I_N(x^*) .$$

Recalling (15), it can be easily verified that the function $P(x, z; \varepsilon)$ is twice continuously differentiable in a neighbourhood of (x^*, z^*), with $z^* = \phi(x^*)$, and its Hessian matrix at (x^*, z^*) is given by

$$\nabla^2 P(x^*, z^*; \varepsilon) := \begin{bmatrix} \nabla_x^2 P(x^*, z^*; \varepsilon) & \nabla_{xz}^2 P(x^*, z^*; \varepsilon) \\ \nabla_{zx}^2 P(x^*, z^*; \varepsilon) & \nabla_z^2 P(x^*, z^*; \varepsilon) \end{bmatrix}, \tag{31}$$

where

$$\begin{aligned}
\nabla_x^2 P(x^*, z^*; \varepsilon) = {} & \sum_{i=1}^{m} \lambda_i(x^*, z^*) \nabla^2 f_i(x^*) \\
& + \nabla f_A(x^*) \nabla_x \lambda_A(x^*, z^*)' + \nabla_x \lambda_A(x^*, z^*) \nabla f_A(x^*)' \\
& + \frac{2}{\varepsilon} \nabla f_A(x^*) B_A(x^*)^{-1} \nabla f_A(x^*)' \\
& - \tfrac{1}{2} \varepsilon \nabla_x \lambda_N(x^*, z^*) B_N(x^*) \nabla_x \lambda_N(x^*, z^*)' , \\
\nabla_z^2 P(x^*, z^*; \varepsilon) = {} & -2 \nabla_z \lambda_A(x^*, z^*) u_A + \frac{2}{\varepsilon} u_A' B_A(x^*)^{-1} u_A \\
& - \tfrac{1}{2} \varepsilon \nabla_z \lambda_N(x^*, z^*) B_N(x^*) \nabla_z \lambda_N(x^*, z^*)' , \\
\nabla_{xz}^2 P(x^*, z^*; \varepsilon) = {} & \nabla f_A(x^*) \nabla_z \lambda_A(x^*, z^*)' - \nabla_x \lambda_A(x^*, z^*) u_A \\
& - \frac{2}{\varepsilon} \nabla f_A(x^*) B_A(x^*)^{-1} u_A \\
& - \tfrac{1}{2} \varepsilon \nabla_x \lambda_N(x^*, z^*) B_N(x^*) \nabla_z \lambda_N(x^*, z^*)' ,
\end{aligned}$$

where the subscripts A and N refer to the partitions induced by the index sets $I_A(x^*)$ and $I_N(x^*)$.

We recall from [3] the following second order sufficiency optimality condition for problem (1).

Proposition 4.3. *Let $x^* \in \mathscr{E}$ be a critical point for problem (1), let λ^* be the corresponding multiplier introduced in Proposition 2.1, and suppose that:*

(i) *Assumption 3 holds at x^*;*

(ii) *(x^*, λ^*) satisfies the condition*

$$x' \sum_{i=1}^{m} \lambda_i^* \nabla^2 f_i(x^*) x > 0$$

for all x such that $x'\nabla f_i(x^*) = 0$, $i \in I_A(x^*)$ and $x \neq 0$.

Then x^* is an isolated local minimum point of problem (1). □

Then we can state the following theorems.

Theorem 4.4. *Let $x^* \in \mathscr{E}$ be a local minimum point for problem (1) satisfying the assumptions of Proposition 4.3. Then there exists an $\tilde{\varepsilon} > 0$ such that for all $\varepsilon \in (0, \tilde{\varepsilon}]$, $(x^*, \phi(x^*))$ is a local minimum point for $P(x, z; \varepsilon)$, and $\nabla^2 P(x^*, \phi(x^*); \varepsilon)$ is positive definite.*

Proof. By Theorem 3.1 we have that $(x^*, \phi(x^*))$ is a stationary point of $P(x, z; \varepsilon)$ for all $\varepsilon > 0$; moreover, recalling Proposition 2.4, we have that $\lambda(x^*, \phi(x^*)) = \lambda^*$ where λ^* is the multiplier vector introduced in Proposition 2.1. By the assumptions made we have that the Hessian matrix $\nabla^2 P$ exists and it is given by (31).

Consider now the quadratic forms on $\mathbb{R}^n \times \mathbb{R}$,

$$\begin{aligned} Q(x, z) :=& x' \sum_{i=1}^{m} \lambda_i^* \nabla^2 f_i(x^*) x + 2x' \nabla f_A(x^*) \nabla_x \lambda_A(x^*, z^*)' x \\ &+ 2x' (\nabla f_A(x^*) \nabla_z \lambda_A(x^*, z^*)' - \nabla_x \lambda_A(x^*, z^*) u_A) z \\ &- 2z^2 \nabla_z \lambda_A(x^*, z^*) u_A , \\ R(x, z) :=& x' \nabla f_A(x^*) B_A(x^*)^{-1} \nabla f_A(x^*)' x \\ &- 2x' \nabla f_A(x^*) B_A(x^*)^{-1} u_A z + z^2 u_A' B_A(x^*)^{-1} u_A , \\ S(x, z) :=& x' \nabla_x \lambda_N(x^*, z^*) B_N(x^*) \nabla_x \lambda_N(x^*, z^*)' x \\ &+ 2x' \nabla_x \lambda_N(x^*, z^*) B_N(x^*) \nabla_z \lambda_N(x^*, z^*)' z \\ &+ z^2 \nabla_z \lambda_N(x^*, z^*) B_N(x^*) \nabla_z \lambda_N(x^*, z^*)' , \end{aligned} \tag{32}$$

so that we can write

$$\begin{pmatrix} x \\ z \end{pmatrix}' \nabla^2 P(x^*, z^*) \begin{pmatrix} x \\ z \end{pmatrix} = Q(x, z) + \frac{2}{\varepsilon} R(x, z) - \tfrac{1}{2} \varepsilon S(x, z) . \tag{33}$$

It can be verified that the quadratic form $R(x, z)$ can be put in the form

$$R(x,z)=\begin{pmatrix}x\\z\end{pmatrix}'\begin{pmatrix}\nabla f_A(x^*)\\-u'_A\end{pmatrix}B_A(x^*)^{-1}\begin{pmatrix}\nabla f_A(x^*)\\-u'_A\end{pmatrix}'\begin{pmatrix}x\\z\end{pmatrix}.$$

This shows that $R(x, z)\geqslant 0$ for all (x, z). Let (x, z) be such that $R(x, z)=0$; then we have

$$\nabla f_A(x^*)'x-u_A z=0\,. \tag{34}$$

Premultiplying (34) by $\lambda_A^{*\prime}$ we get

$$\lambda_A^{*\prime}\,\nabla f_A(x^*)'x-\lambda_A^{*\prime}u_A z=0\,. \tag{35}$$

Noting that, by Proposition 2.1, we have $\lambda_A^{*\prime}u_A=1$ and $\nabla f_A(x^*)\lambda_A^*=0$, from (35) we obtain $z=0$. Putting $z=0$ in (34) we have $\nabla f_A(x^*)'x=0$.

Therefore by (32) and the second order sufficiency assumption we can assert that $Q(x, z)>0$ for all $(x, z)\neq 0$ such that $R(x, z)=0$. Hence recalling known results on pairs of quadratic forms [20], by (33) we have that there exists an $\tilde{\varepsilon}>0$ such that for all $\varepsilon\in(0, \tilde{\varepsilon}]$, the Hessian matrix $\nabla P(x^*, z^*; \varepsilon)$ is positive definite. □

Theorem 4.5. *Suppose that Assumption* 3 *holds at every critical point* $x^*\in\mathscr{E}$ *of problem* (1). *Then, there exists and* $\varepsilon^*>0$ *such that, for all* $\varepsilon\in(0, \varepsilon^*]$, *if* $(x_\varepsilon, z_\varepsilon)\in\mathscr{L}_\varepsilon$ *is an isolated local unconstrained minimum point of* $P(x, z; \varepsilon)$ *and the Hessian matrix* $\nabla^2P(x_\varepsilon, z_\varepsilon; \varepsilon)$ *is positive definite, the point* x_ε *is a local minimum point of problem* (1) *satisfying the second order sufficiency condition, and* $\phi(x_\varepsilon)=z_\varepsilon$.

Proof. By Theorem 4.2 there exists an ε^* such that for all $\varepsilon\in(0, \varepsilon^*]$ the point x_ε is a local minimum point for problem (1) and $\phi(x_\varepsilon)=z_\varepsilon$. By assumption, the matrix $\nabla^2P(x_\varepsilon, z_\varepsilon; \varepsilon)$ is positive definite, so that, in particular, $\nabla_x^2P(x_\varepsilon, z_\varepsilon; \varepsilon)$ is positive definite. Hence, recalling the expression of $\nabla_x^2P(x_\varepsilon, z_\varepsilon; \varepsilon)$ given in (31), we have that $x'\sum_{i=1}^m\lambda_i^*\,\nabla^2 f_i(x^*)x>0$ for all x such that $x'\nabla f_i(x^*)=0$, $i\in I_A(x^*)$ and $x\neq 0$, and this completes the proof. □

5. A globally convergent algorithm

As in [8], we can define an automatic adjustment rule for the penalty coefficient that appears in the function $P(x, z; \varepsilon)$. This rule allows us to construct and implementable algorithm which is proved to be globally convergent towards critical points of problem (1).

For a given value of ε, the algorithm makes use of an unconstrained minimization method, which is defined through an iteration map A_ε: $\mathbb{R}^n\times\mathbb{R}\to 2^{\mathbb{R}^n\times\mathbb{R}}$ satisfying the following conditions.

Assumption 4. For each $\varepsilon>0$ we have:

(i) if $(x, z)\in\mathscr{L}_\varepsilon$ and $(y, r)\in A_\varepsilon(x, z)$, then $(y, r)\in\mathscr{L}_\varepsilon$;

(ii) if $\{(x_k, z_k)\}$ is a sequence produced by A_ε, then, every limit point of $\{(x_k, z_k)\}$ in $\mathscr{L}_\varepsilon$ is a stationary point of $P(x, z; \varepsilon)$.

The preceding conditions can be fulfilled by means of simple modifications of any globally convergent unconstrained method. As an example, we may think that the iteration map is realized by employing a gradient related algorithm, with an Armijo-type line search where suitable safeguards are used in order to take into account the presence of barrier terms.

Algorithm model.

Data: $w=x_b$, $v=\phi(x_b)$, $\sigma>0$ and $\phi\in(0,1)$.

Step 1. Set $k=0$. If $\phi(x_b)\leqslant\phi(w)$, then set $x_0=x_b$ and $z_0=\phi(x_b)$; else set $x_0=w$ and $z_0=\phi(w)$.

Step 2. If $\nabla P(x_k, z_k; \varepsilon)=0$; then: if $c(x_k, z_k; \varepsilon)=0$ stop; else set $\varepsilon=\theta\varepsilon$, $w=x_k$, $v=z_k$ and go to Step 1.

Step 3. If

$$\varepsilon\| \nabla f(x_k)'\nabla P(x_k,z_k; \varepsilon) - u\nabla_z P(x_k, z_k; \varepsilon)\|^2 \geqslant \sigma\|c(x_k, z_k; \varepsilon)\|^2,$$

then set $\varepsilon=\theta\varepsilon$, $w=x_k$, $v=z_k$ and go to Step 1.

Step 4. Compute $(y, r)\in A_\varepsilon(x_k, z_k)$; if $\phi(y)\leqslant\phi(x_a)-\varepsilon$ then set $x_{k+1}=y$, $z_{k+1}=r$, $k=k+1$ and go to Step 2; else set $\varepsilon=\theta\varepsilon$, $w=x_k$, $v=z_k$ and go to Step 1.

The following theorem states the convergence of the algorithm. In the proof we denote by $\{\varepsilon_j\}$ the sequence of values of the penalty parameter and by $\{(w_j,v_j)\}$ the sequence of points defined at each updating of ε.

Theorem 5.1. *Suppose that the iteration map A_ε satisfies the conditions stated in Assumption 4. Then, either the algorithm terminates at some $(x_\nu, z_\nu)\in\mathscr{E}\times\mathbb{R}$ and x_ν is a critical point for problem (1), or the algorithm produces a bounded sequence $\{(x_k, z_k)\}\subseteq\mathscr{E}\times\mathbb{R}$ and every limit point (x^*, z^*) is such that x^* is a critical point for problem (1) with $\phi(x^*)=z^*$.*

Proof. We note first that the instructions at Step 1 imply that the iteration map A_ε uses, for each j, a starting point $(x_0, z_0)\in\mathscr{E}\times\mathbb{R}$ satisfying $z_0=\phi(x_0)\leqslant\phi(x_b)$. Therefore, by Proposition 2.5(i), we have $P(x_0, z_0; \varepsilon_j)\leqslant\phi(x_0)\leqslant\phi(x_b)$ and so $(x_0, z_0)\in\mathscr{L}_{\varepsilon_j}$. By Assumption 4(i) and the instructions at Step 4, this implies that the sequence $\{(x_k, z_k)\}$ produced in correspondence to a fixed value of j remains in $\mathscr{L}_{\varepsilon_j}$; thus we have $(x_j, v_j)\in\mathscr{L}_{\varepsilon_j}$ for all j.

Now we prove, by contradiction, that the sequence $\{(w_j, v_j)\}$ produced by the algorithm is finite. Assume the contrary; then by Theorem 2.6(ii), the sets $\mathscr{L}_{\varepsilon_j}$ are compact for sufficiently large values of j, say $j\geqslant j_1$. By Assumption 4(i), we have $A_{\varepsilon_j}(w_j, v_j)\subseteq\mathscr{L}_{\varepsilon_j}$ for all $j\geqslant j_1$.

Consider first the test at Step 4 and let $(y_j, r_j)\in A_{\varepsilon_j}(w_j, v_j)$. Suppose that

$$\phi(y_j)>\phi(x_a)-\varepsilon_j \tag{36}$$

for an infinite subsequence that we relabel $\{(y_j, r_j)\}$. By Proposition 2.5(ii) and Proposition 3.3(i), we have

$$\phi(x_b)\geqslant P(y_j, r_j; \varepsilon_j)\geqslant\phi(y_j)-\varepsilon_j^{1/2}K_1>\phi(x_a)-\varepsilon_j-\varepsilon_j^{1/2}K_1,$$

where

$$K_1=K_0+\varepsilon_0^{1/2}\left(\frac{m^2}{4\tilde{\mu}^2}(\phi(x_a)-f_{\min})\right).$$

Therefore, there exists an index $j_2\geqslant j_1$, such that for $j\geqslant j_2$ we get a contradiction to the assumption $\phi(x_b)<\phi(x_a)$.

As regards Step 2, by Theorem 3.1 and Theorem 3.5, there exists an index $j_3 \geqslant 0$ such that the algorithm cannot construct any point (w_{j+1}, v_{j+1}) with $j \geqslant j_3$ at Step 2.

It follows that the point $(w_{j+1}), v_{j+1}, j \geqslant j_4 = \max[j_2, j_3]$ should have been produced because of a failure to satisfy the test at Step 3. Namely, for $j \geqslant j_4$ we should have

$$\begin{aligned} &\varepsilon_j \| \nabla f(w_{j+1})' \nabla_x P(w_{j+1}, v_{j+1}; \varepsilon_j) - u \nabla_z P(w_{j+1}, v_{j+1}; \varepsilon_j) \|^2 \\ &\quad < \sigma \| c(w_{j+1}, v_{j+1}; \varepsilon_j) \|^2 . \end{aligned} \tag{37}$$

From the instructions at Step 1 and the assumptions made on the iteration map A_ε it follows that $(w_{j+1}, v_{j+1}) \in \mathscr{L}_{\varepsilon_j}$. Now, since $\varepsilon_j \to 0$, as $j \to \infty$, using (37) and Proposition 3.3(ii), we obtain that the sequence $\{w_j, v_j\}$ admits a limit point $(\hat{w}, \hat{v}) \in \mathscr{F}$. Then, recalling Proposition 3.4, we have that for sufficiently large values of j there exists a constant $\sigma(\hat{w}, \hat{v}) > 0$ such that

$$\begin{aligned} &\varepsilon_j \| \nabla f(w_{j+1})' \nabla_x P(w_{j+1}, v_{j+1}; \varepsilon_j) - u \nabla_z P(w_{j+1}, v_{j+1}; \varepsilon_j) \|^2 \\ &\quad \geqslant \frac{1}{\varepsilon_j} \sigma(\hat{w}, \hat{v}) \| c(w_{j+1}, v_{j+1}; \varepsilon_j) \|^2 , \end{aligned}$$

which, for sufficiently small values of ε_j, contradicts (36).

Therefore we can conclude that ε_j is updated only a finite number of times. Denote by ε_L the last value of the penalty parameter and consider the points (x_k, z_k) produced at Step 4, in correspondence to ε_L. As observed at the beginning of the proof, we have, by construction, that $(x_k, z_k) \in \mathscr{L}_{\varepsilon_L}$ for all k. Then, if the algorithm terminates at (x_ν, z_ν), the assertion follows from Proposition 3.2. If the algorithm produces an infinite sequence, by Theorem 2.6(i), we have that the sequence $\{x_k, z_k\}$ is bounded. Moreover, as the test at Step 4 is satisfied, we have also $\phi(x_k) \leqslant \phi(x_a) - \varepsilon_L$, so that the sequence $\{x_k, z_k\}$ remains in a compact set contained in $\mathscr{L}_{\varepsilon_L}$ and admits a limit point in $\mathscr{L}_{\varepsilon_L}$. By Assumption 4(ii) every limit point (x^*, z^*) is a stationary point of P in $\mathscr{E} \times \mathbb{R}$, and by Step 3 we have also:

$$c(x^*, z^*; \varepsilon_L) = 0$$

so that, again by Proposition 3.2, the theorem is proved. □

6. Numerical results

In order to validate the proposed approach from a computational point of view, some preliminary numerical experiments have been carried out. The main objective of our experiments was not that of evaluating a new computational code for minimax problems, but rather that of obtaining some feel for potentialities and limitations of an algorithm using the differentiable penalty function defined in this paper. Therefore, the computational experiments were carried out by employing standard library software for unconstrained minimization of differentiable functions. A computational code, which incorporates the automatic adjustment rule for the penalty parameter given in Section 5, is currently being developed [9].

The results reported here were obtained by means of a two-phase algorithm, employing, in both phases, a Newton-type minimization routine from the NAG library. In the first phase we start from an estimate z_0 of the z-variable and we minimize P with respect to the x-component using the standard initial point x_0. In the second phase, we minimize P with respect to both x and z, using as starting point the solution $\bar{x}$ obtained in the first phase and a new initial estimates z_1 of z. This allows to reduce, when far from the solution, the ill conditioning due to the different scale of the variables x and z. We used the starting values of z given by

$$z_0 = -c_1 \max(\phi_{\min}(x_0), 10^{-3}), \qquad z_1 = \phi_{\min}(\bar{x}) - c_2 \max(\phi_{\min}(\bar{x}), 10^{-3}),$$

where

$$\phi_{\min}(x) = \min_{1 \leqslant i \leqslant m} f_i(x)$$

and c_1, c_2 are positive constants. These choices have the effect that the starting points (x_0, z_0) and $(\bar{x}, z_1)$ are not feasible for every constraint of problem (2). It was experienced that this produces a better conditioning of the penalty terms.

As regard the construction of P, and, in particular, the definition of the barrier term, we note that there is no need to generate explicitly a point x_a and we set

$$\phi(x_a) = c_3 \max(1, \phi(x_0)),$$

where c_3 is a positive constant. This is justified as long as we suppose that $\phi(x) \to \infty$ for $\|x\| \to \infty$.

The multiplier function is evaluated by means of the F04ABF routine of the NAG library, which makes use of the Cholesky factorization of the positive definite matrix $M(x, z)$.

In both phases we used fixed values of ε that were determined according to the rules

$$\varepsilon_0 = \frac{c_4}{\phi(x_a)} \min(1, n/m), \qquad \varepsilon_1 = c_5 \varepsilon_0,$$

for given positive constants c_4 and c_5.

The unconstrained minimization of function P has been performed by employing the E04EBF routine of the NAG library that implements a Newton-type method based on a modified Cholesky factorization of the Hessian matrix. The Hessian matrix of P has been replaced by the consistent approximation obtained by deleting all terms containing the third order derivatives of the problem functions.

The presence of the barrier term has been taken into account by means of the (rather crude) device of defining $P = 10^{20}$ whenever $\phi(x) \geqslant \phi(x_a)$. In the two stages we used the termination conditions

$$\|\nabla P(x, z_0; \varepsilon_0)\| \leqslant \delta_0, \qquad \|\nabla P(x, z; \varepsilon_1)\| \leqslant \delta_1.$$

We considered a set of standard test problems, which are collected in the Appendix, with the following values of the parameters

$$c_1 = 10^{-3}, \quad c_2 = 10^{-1}, \quad c_3 = 10^6, \quad c_4 = 10^3, \quad c_5 = 10^{-3}, \quad \delta_0 = 1, \quad \delta_1 = 10^{-6}.$$

The numerical results are shown in Table 1, where n is the number of variables and m is the number of functions f_i in the original minimax problem, n_i denotes the number of iterations, n_P is the number of evaluations of the function P and ϕ^* is the best value of the objective function obtained. The symbol F_1 denotes a failure of the F04ABF routine during the computation of the multiplier function; the symbol F_2 indicates convergence to a stationary point of P that is not a critical point of ϕ.

In problems Hald–Madsen 1 and Maxl the failure in the F04ABF routine occurs at points that are very close to the optimal solution. In problems Hald–Madsen 2 and El Attar the failure is due to an inappropriate value of the penalty parameter. For these problems convergence to the optimal solution was achieved by setting $c_4 = 10^2$, $c_5 = 10^{-7}$ for the Hald–Madsen 2 test function and $c_4 = 1$, $c_5 = 10^{-5}$ for the El Attar problem. In Table 2 are reported the results obtained in correspondence to the choice $z_0 = z_1 = 0$, $\delta_1 = 10^{-6}$, which gave better results.

Table 1
Numerical results obtained with the E04EBF routine

Problems	n	m	n_i	n_P	ϕ^*
Crescent	2	2	13	19	3 E−9
Polak 1	2	2	9	11	2.7182818
LQ	2	2	7	10	−1.41421356
Mifflin 1	2	2	19	37	−1
Mifflin 2	2	2	7	9	−1
QL	2	3	15	26	7.200000
Charalambous–Conn 1	2	3	6	9	1.95222
Charalambous–Conn 2	2	3	11	16	2.00000
Demyanov–Malozemov	2	3	4	6	−3
Hald–Madsen 1	2	4	24	54	1 E−7 (F_1)
Rosen (x_0')	4	4	15	25	−44
Rosen (x_0'')	4	4	15	25	−44
Shor	5	10	11	15	22.60016
Hald–Madsen 2	5	42	—	—	F_2
El Attar	6	102	—	—	F_2
Polak 2	10	2	6	8	54.598150
Maxquad (x_0')	10	5	18	28	−0.841408
Maxquad (x_0'')	10	5	19	45	−0.841408
Polak 3	11	10	13	16	3.703482
Maxq	20	20	3	4	3 E−7
Maxl	20	40	1	2	1 E−11 (F_1)
Goffin	50	50	4	13	1 E−13

Table 2
Numerical results obtained with the E04EBF routine

Problems	n	m	n_i	n_P	ϕ^*
Hald–Madsen 2	5	42	18	35	0.000122
El Attar	6	102	28	56	0.0349

It is quite difficult to compare the results given here with those obtained in the literature by using methods of nonsmooth optimization (see, for example, [29, 35]). We note that each evaluation of P requires the evaluation of first order derivatives of the problem functions and the solution of a linear system for computing the multiplier function. Thus, an evaluation of P is definitely more expensive than a function and subgradient call in nonsmooth methods. As regards the cost of computing a search direction in the minimization of P, we note that each evaluation of ∇P requires, in addition, the evaluation of the second order derivatives $\nabla^2 f_i(x)$. The same information can be used for obtaining a consistent approximation of $\nabla^2 P$, so that the computation of a Newton-type direction requires only the additional solution of an $(n+1)$-dimensional linear system. On the other hand, the evaluation of the search direction in nonsmooth methods requires the solution of a quadratic programming subproblem whose dimension depends on the number of subgradients in the bundle. Therefore it is difficult to compare the burden of each iteration in our smooth approach and in the nonsmooth approach.

Nevertheless, some indication can be obtained by comparing the results of Table 1 for a set of 15 test problems with the results given in [29] and [35] with reference to the well-known code M1FC1 and reported in Table 3. Here the symbol f/g denotes the number of function/subgradient evaluations.

It can be observed that in almost all cases the number of iterations needed for minimizing P is considerably smaller than that needed by M1FC1 and that a greater accuracy in the optimal value is obtained. In order to have some rough comparative figure on the cost of function evaluations, we could assume that each evaluation of P is equivalent to m function/subgradient evaluations in M1FC1. Using this criterion it would appear that our results are definitely better in 8 cases; in the remaining cases the performance of M1FC1 is superior in terms of computational cost, but the solution is less accurate. This indicates that a code performing the minimization of P can be competitive with the available codes of nonsmooth optimization, at least for problems where the number m is not much larger than the number n. For problems where m is much larger than n, as in Hald–Madsen 2 and in El Attar

Table 3
Numerical results obtained with the M1FC1 code

Problems	n	m	n_i	f/g	ϕ^*
Crescent	2	2	31	93	2 E−6
LQ	2	2	16	52	−1.41420
Mifflin 1	2	2	143	281	−0.999967
Mifflin 2	2	2	30	71	−0.999993
QL	2	3	12	30	7.200018
Charalambous–Conn 1	2	3	11	31	1.952253
Charalambous–Conn 2	2	3	12	44	2.001415
Demyanov–Malozemov	2	3	10	33	−3
Rosen (x_0')	4	4	22	61	−43.9998
Shor	5	10	21	71	22.60018
Maxquad (x_0')	10	5	29	69	−0.8413589
Maxquad (x_0'')	10	5	20	54	−0.841359
Maxq	20	20	144	207	0.0
Maxl	20	40	138	213	0.0
Goffin	50	50	72	194	0.00010

problems, the choice of the penalty parameter appear to be more critical and the results appear to be inferior to those given in the literature (see, for instance, [18, 28]).

The preceding results motivate a further study of computational aspects with the objective of producing an efficient code. In particular, it appears that the implementation of an automatic adjustment rule for the penalty parameter could be useful. Furthermore, the choice of the search direction can be substantially improved by taking into account some of the ideas of [18]. Finally, the adoption of nonmonotone global stabilization algorithm of the kind proposed in [17] could be beneficial for improving the speed of convergence and reducing the number of function evaluations.

7. Conclusions

We have shown that, under mild assumptions, the finite minimax problem is equivalent to the unconstrained minimization of a smooth function and this allows the construction of globally and superlinearly convergent methods. The preliminary numerical experiments reported in this paper demonstrate that satisfactory results, both in terms of convergence rate and accuracy of the solution, can be obtained by means of standard library routines, even for fixed values of the penalty parameter.

It appears, however, that further work is needed in order to develop an efficient computational code by exploiting the problem structure and employing an automatic adjustment rule of the penalty parameter. A Newton-type algorithm that incorporates some of these features is currently being developed [9]. Finally we mention that our approach could greatly benefit from the adoption of techniques of automatic differentiation for the evaluation of first and second order derivatives (see, e.g. [16]). It is also worth noticing that the exact penalty approach considered here can easily be extended to the case of constrained minimax problems.

Appendix

Crescent [21].

$$\phi(x) = \max\{x_1^2 + (x_2-1)^2 + x_2 - 1,\ -x_1^2 - (x_2-1)^2 + x_2 + 1\},$$

$n=2$, $\phi(x^*)=0.0$, $x_0=(-1.5, 2)$.

Polak 1 [32].

$$\phi(x) = \max\{e^{x_1^2/1000 + (x_2-1)^2},\ e^{x_1^2/1000 + (x_2+1)^2}\},$$

$n=2$, $\phi(x^*)=2.7182818$, $x_0=(50, 0.05)$.

LQ [29].

$$\phi(x) = \max\{-x_1 - x_2,\ -x_1 - x_2 + (x_1^2 + x_2^2 - 1)\},$$

$n=2$, $\phi(x^*) = -\sqrt{2}$, $x_0=(-0.5, -0.5)$.

Mifflin 1 [29].

$$\phi(x) = -x_1 + \max\{x_1^2 + x_2^2 - 1, 0\},$$

$n=2$, $\phi(x^*) = -1$, $x_0 = (0.8, 0.6)$.

Mifflin 2 [29].

$$\phi(x) = -x_1 + 2(x_1^2 + x_2^2 - 1) + 1.75 \max\{\pm(x_1^2 + x_2^2 - 1)\},$$

$n=2$, $\phi(x^*) = -1$, $x_0 = (-1, -1)$.

Charalambous–Conn 1 [4].

$$\phi(x) = \max\{x_1^2 + x_2^4, (2-x_1)^2 + (2-x_2)^2, 2\,\mathrm{e}^{-x_1+x_2}\},$$

$n=2$, $\phi(x^*) = 1.95222$, $x_0 = (1, -0.1)$.

Charalambous–Conn 2 [4].

$$\phi(x) = \max\{x_1^4 + x_2^2, (2-x_1)^2 + (2-x_2)^2, 2\,\mathrm{e}^{-x_1+x_2}\},$$

$n=2$, $\phi(x^*) = 2$, $x_0 = (2, 2)$.

Demyanov–Malozemov [6].

$$\phi(x) = \max\{5x_1 + x_2, -5x_1 + x_2, x_1^2 + x_2^2 + 4x_2\},$$

$n=2$, $\phi(x^*) = -3$, $x_0 = (1, 1)$.

QL [29].

$$\phi(x) = \max_{1 \leqslant i \leqslant 3} \{f_i(x)\},$$

where

$$f_1(x) = x_1^2 + x_2^2,$$

$$f_2(x) = x_1^2 + x_2^2 + 10(-4x_1 - x_2 + 4),$$

$$f_3(x) = x_1^2 + x_2^2 + 10(-x_1 - 2x_2 + 6),$$

$n=2$, $\phi(x^*) = 7.2$, $x_0 = (-1, 5)$.

Hald–Madsen 1 [18].

$$\phi(x) = \max\{\pm 10(x_2 - x_1^2), \pm(1 - x_1)\},$$

$n=2$, $\phi(x^*) = 0.0$, $x_0 = (1.2, 1)$.

Rosen [29].

$$\phi(x) = \max\{f_1(x), f_1(x) + 10f_2(x), f_1(x) + 10f_3(x), f_1(x) + 10f_4(x)\},$$

where

$$f_1(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4 ,$$

$$f_2(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8 ,$$

$$f_3(x) = x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10 ,$$

$$f_4(x) = x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5 ,$$

$n=4$, $\phi(x^*) = -44$, $x_0' = (0, 0, 0, 0)$, $x_0'' = (2, 2, 5, 0)$.

Shor [36].

$$\phi(x) = \max_{1 \leqslant i \leqslant 0} \left\{ b_i \sum_{j=1}^{5} (x_j + a_{ij})^2 \right\} ,$$

$n=5$, $\phi(x^*) = 22.60016$, $x_0 = (0, 0, 0, 0, 1)$.

Hald–Madsen 2 [18].

$$\phi(x) = \max_{1 \leqslant i \leqslant 21} \{ \pm f_i(x) \} ,$$

where

$$f_i(x) = \frac{x_1 + x_2 y_i}{1 + x_3 y_i + x_4 y_i^2 + x_5 y_i^3} - e^{y_i} ,$$

$$y_i = -1 + 0.1(i-1) ,$$

$n=5$, $\phi(x^*) = 0.000122$, $x_0 = (0.5, 0, 0, 0, 0)$.

El Attar [10].

$$\phi(x) = \max_{1 \leqslant i \leqslant 51} \{ \pm f_i(x) \} ,$$

where

$$f_i(x) = x_1 \, e^{-x_2 t_i} \cos(x_3 t_i + x_4) + x_5 \, e^{-x_6 t_i} - y_i ,$$

$$y_i = \tfrac{1}{2} e^{t_i} - e^{-2t_i} + \tfrac{1}{2} e^{-3t_i} + \tfrac{3}{2} e^{-3t_i/2} \sin(7t_i) + e^{-5t_i/2} \sin(5t_i),$$

$$t_i = \tfrac{1}{10}(i-1) ,$$

$n=6$, $\phi(x^*) = 0.0349$, $x_0 = (2, 2, 7, 0, -2, 1)$.

Polak 2 [32].

$$\phi(x) = \max\{ f(x+2a), f(x-2a) \},$$

where

$$f(x) = e^{(0.0001x_2)^2 + x_2^2 + x_3^2 + (2x_4)^2 + x_5^2 + \ldots + x_{10}^2},$$

$$a = (0, 1, 0, 0, 0, 0, 0, 0, 0, 0),$$

$n=10$, $\phi(x^*) = 54.591846$, $x_0 = (100, 0.1, \ldots, 0.1)$.

Maxquad [26].

$$\phi(x) = \max_{1 \leq i \leq 5} \{x'A_i x - b_i' x\},$$

$n=10$, $\phi(x^*) = -0.841408$, $x_0' = (1,\ldots, 1)$, $x_0'' = (0,\ldots, 0)$.

Polak 3 [32].

$$\phi(x) = \max_{1 \leq i \leq 10} \left\{ \sum_{j=1}^{11} \frac{1}{j-i+1} e^{[(x_j - \sin(i-1+2(j-1)))^2]} \right\},$$

$n=11$, $\phi(x^*) = 3.703483$, $x_0 = (1,\ldots, 1)$.

Maxq [29].

$$\phi(x) = \max_{1 \leq i \leq 20} \{x_i^2\},$$

$n=20$, $\phi(x^*) = 0.0$, $x_0 = (1,\ldots, 10, -11,\ldots, -20)$.

Maxl [29].

$$\phi(x) = \max_{1 \leq i \leq 20} \{\pm x_i\},$$

$n=20$, $\phi(x^*) = 0.0$, $x_0 = (1,\ldots, 10, -11,\ldots, -20)$.

Goffin [29].

$$\phi(x) = 50 \max_{1 \leq i \leq 50} \{x_i\} - \sum_{i=1}^{50} x_i,$$

$n=50$, $\phi(x^*) = 0.0$, $x_0 = (1-25.5,\ldots, i-25.5,\ldots, 50-25.5)$.

References

[1] J.W. Bandler and C. Charalambous, "Practical least pth optimization of networks," *IEEE Transactions on Microwave Theory and Techniques (1972 Symposium Issue)* MTT-20 (1972) 834–840.

[2] D.P. Bertsekas, "Nondifferentiable optimization via approximation," *Mathematical Programming Study* 3 (1975) 1–25.

[3] C. Charalambous, "Acceleration of the least pth algorithm for minimax optimization with engineering applications," *Mathematical Programming* 17 (1979) 270–297.

[4] C. Charalambous and A.R. Conn, "An efficient method to solve the minimax problem directly," *SIAM Journal on Numerical Analysis* 15 (1978) 162–187.

[5] F.H. Clarke, V.F. Dem'yanov and F. Giannessi, eds., *Nonsmooth Optimization and Related Topics* (Plenum, New York, 1989).

[6] V.F. Dem'yanov and V.N. Malozemov, *Introduction to Minimax* (Wiley, New York, 1974).

[7] V.F. Dem'yanov and V. Vasil'ev, *Nondifferentiable Optimization* (Springer, New York, 1986).

[8] G. Di Pillo and L. Grippo, "An exact penalty method with global convergence properties for nonlinear programming problems," *Mathematical Programming* 36 (1986) 1–18.

[9] G. Di Pillo, L. Grippo and S. Lucidi, ''Superlinearly convergent algorithms for finite minimax problems,'' Technical Report, IASI, National Research Council (Rome), to appear.

[10] R.A. El-Attar, M. Vidyasagar and S.R.K. Dutta, ''An algorithm for l_1-approximation,'' *SIAM Journal on Numerical Analysis* 16 (1979) 70–86.

[11] R. Fletcher, ''A model algorithm for composite nondifferentiable optimization problems,'' *Mathematical Programming Study* 17 (1982) 67–76.

[12] R. Fletcher, ''Second order correction for nondifferentiable optimization,'' in: G.A. Watson, ed., *Numerical Analysis* (Springer, Berlin, 1982) pp. 85–114.

[13] M. Gaudioso and M.F. Monaco, ''A bundle type approach to the unconstrained minimization of convex nonsmooth functions,'' *Mathematical Programming* 23 (1982) 216–226.

[14] C. Gigola and S. Gomez, ''A regularization method for solving the finite convex min–max problem,'' *SIAM Journal and Numerical Analysis* 27 (1990) 1621–1634.

[15] T. Glad and E. Polak, ''A multiplier method with automatic limitation of penalty growth,'' *Mathematical Programming* 17 (1979) 140–155.

[16] A. Griewank, ''On automatic differentiation,'' in: M. Iri and K. Tanabe, eds., *Mathematical programming* (Kluwer Academic Publishers, Tokyo) pp. 83–107.

[17] L. Grippo, F. Lampariello and S. Lucidi, ''A class of nonmonotone stabilization methods in unconstrained optimization,'' *Numerische Mathematik* 59 (1991) 779–805.

[18] J. Hald and K. Madsen, ''Combined LP and quasi-Newton methods for minimax optimization,'' *Mathematical Programming* 20 (1981) 49–62.

[19] S.P. Han, ''Variable metric methods for minimizing a class of nondifferentiable functions,'' *Mathematical Programming* 20 (1981) 1–13.

[20] M.R. Hestenes, *Optimization Theory. The Finite Dimensional Case* (Wiley, New York, 1975).

[21] K.C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics No. 1133 (Springer, Berlin, 1985).

[22] C. Lemaréchal, ''An extension of Davidson method to nondifferentiable problems,'' *Mathematical Programming Study* 3 (1975) 95–109.

[23] C. Lemaréchal, ''Numerical experiments in nonsmooth optimization,'' in: E.A. Nurminskii, ed., *Progress in Nondifferentiable Optimization* (IIASA Proceedings, Laxenburg, 1982).

[24] C. Lemaréchal, ''Nondifferentiable optimization,'' in: G.L. Nemhauser, A.H.G. Rinnooy Kan and M.J. Todd, eds., *Handbooks in Operations Research and Management Science, Vol. 1: Optimization* (North-Holland, Amsterdam, 1989) pp. 529–572.

[25] C. Lemaréchal and R. Mifflin, eds., *Nonsmooth Optimization* (Pergamon, Oxford, 1978).

[26] C. Lemaréchal and R. Mifflin, ''A set of nonsmooth optimization test problems,'' in: C. Lemaréchal and R. Mifflin, eds., *Nonsmooth Optimization* (Pergamon, Oxford, 1978).

[27] S. Lucidi, ''New results on a continuously differentiable exact penalty function,'' Report R-277, IASI, National Research Council (Rome, 1989).

[28] K. Madsen, ''Minimization of non-linear approximation functions,'' Ph.D. Thesis, Technical University of Denmark (Lyngby, 1985).

[29] M.M. Mäkelä, ''Nonsmooth optimization,'' Ph.D. Thesis, University of Jyväskylä (Jyväskylä, 1990).

[30] R. Mifflin, ''An algorithm for constrained optimization with semismooth functions,'' *Mathematics of Operations Research* 2 (1977) 191–207.

[31] E. Polak, ''Basics of minimax algorithms,'' in: F.H. Clarke, V.F. Dem'yanov and F. Giannessi, eds., *Nonsmooth Optimization and Related Topics* (Plenum, New York, 1989) pp. 343–369.

[32] E. Polak, D.H. Mayne and J.E. Higgins, ''Superlinearly convergent algorithm for min–max problems,'' *Journal of Optimization Theory and Applications* 69 (1991) 407–439.

[33] R.A. Polyak, ''Smooth optimization methods for minimax problems,'' *SIAM Journal on Control and Optimization* 26 (1988) 1274–1286.

[34] B.N. Pshenichny and Y.M. Danilin, *Numerical Methods in Extremal Problems* (Mir, Moscow, 1978).

[35] H. Schramm and J. Zowe, ''A version of the bundle ideal for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results,'' Report N.206, DFG-Schwerpunktprogramms ''Anwendungsbezogene Optimierung und Steuerung,'' Universität Bayreuth (Bayreuth, 1990).

[36] N.Z. Shor, *Minimization Methods for Nondifferentiable Functions* (Springer, Berlin, 1985).

[37] P. Wolfe, ''A Method of conjugate subgradients for minimizing nondifferentiable convex functions,'' *Mathematical Programming Study* 3 (1975) 145–173.

[38] Y. Yuan, ''On the superlinear convergence of a trust region algorithm for nonsmooth optimization,'' *Mathematical Programming* 31 (1985) 269–285.

[39] I. Zang, ''A smoothing-out technique for min–max optimization,'' *Mathematical Programming* 19 (1980) 61–77.

[40] J. Zowe, ''Nondifferentiable optimization,'' in: K. Schittkowski, eds., *Computational Mathematical Programming* (Springer, Berlin, 1985).