

Penalty and smoothing methods for convex semi-infinite programming

Alfred Auslender

Université de Lyon, Université de Lyon 1, CNRS, UMR 5208 Institut Camille Jordan, France, and Department of Economics, Ecole Polytechnique, F-91128 Palaiseau, Cedex, France

email: auslender.alfred@gmail.com

Miguel A. Goberna

Dep. of Statistics and Operations Research, University of Alicante, 03080 Alicante, Spain.

email: mgoberna@ua.es

Marco A. López

Dep. of Statistics and Operations Research, University of Alicante, 03080 Alicante, Spain.

email: marco.antonio@ua.es

In this paper we consider min-max convex semi-infinite programming. In order to solve these problems we introduce a unified framework concerning Remez-type algorithms and integral methods coupled with penalty and smoothing methods. This framework subsumes well-known classical algorithms, but also provides some new methods with interesting properties. Convergence of the primal and dual sequences are proved under minimal assumptions.

Key words: convex semi-infinite programming; asymptotic functions; penalty methods; smoothing methods; duality

MSC2000 Subject Classification: Primary: 90C34, 90C25; Secondary: 90C51

OR/MS subject classification: Primary: programming; Secondary: nonlinear

History: Received: ; Revised: .

1. Introduction In this paper we consider min-max Convex Semi-Infinite Programming (CSIP) problems. More precisely, let T_1 and T_2 be compact metric spaces, and let Q be a closed convex set in \mathbb{R}^n . Furthermore, let $f : T_1 \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g : T_2 \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be finite and continuous functions on $T_1 \times Q$ and $T_2 \times Q$, respectively, and such that for each t the functions $f_t(\cdot) := f(t, \cdot)$ and $g_t(\cdot) := g(t, \cdot)$ are lower semicontinuous (lsc), convex on \mathbb{R}^n , and at least \mathcal{C}^1 on Q .

We consider in this paper the following problem

$$(P) \quad v(P) = \inf\{F(x) \mid x \in C\}, \tag{1}$$

where $F(x) := \sup\{f_t(x) \mid t \in T_1\}$, $G(x) := \sup\{g_t(x) \mid t \in T_2\}$, $C := Q \cap D$, and $D := \{x : G(x) \leq 0\}$. The optimal set of (P) is denoted by S_P .

In the particular case that T_1 is a singleton set, (P) is an ordinary CSIP problem. For solving CSIP, we propose in this paper Remez-type algorithms and integral methods coupled with penalty and smoothing methods.

Remez-type methods (or outer approximations) are inspired by the first algorithm of Remez [23], proposed for approximating functions in the framework of Linear Semi-Infinite Programming (LSIP), that can be described roughly as follows:

Let T_1^k and T_2^k be finite subsets of T_1 and T_2 and denote

$$F^k(x) = \sup\{f_t(x) \mid t \in T_1^k\}, \quad C^k = Q \cap D^k, \quad D^k = \{x : g_t(x) \leq 0 \forall t \in T_2^k\}.$$

Initialization: Set $k = 0$ and start with T_1^0, T_2^0 .

Step 1 Compute $x^k \in \operatorname{argmin}\{F^k(x) \mid x \in C^k\}$.

Step 2 Compute

$$t_1^{k+1} \in \operatorname{argmax}\{f(t, x^k) \mid t \in T_1\}, \quad t_2^{k+1} \in \operatorname{argmax}\{g(t, x^k) \mid t \in T_2\}.$$

Step 3 Choose, for $i = 1, 2$, $T_i^{k+1} \subset T_i$ satisfying $t_i^{k+1} \in T_i^{k+1}$.

Set $k \leftarrow k + 1$; go to Step 1.

This numerical approach requires to solve nonconvex optimization problems in Step 2, which is certainly the main difficulty in the general case. Indeed, from a computational point of view, actually this is only possible for particular cases, mainly when the functions $f(\cdot, x)$ and $g(\cdot, x)$ are polynomial, with low dimensional sets T_1 and T_2 . But in this paper we focus on Step 1 and we try to propose a "good" approximation, (\tilde{P}^k) , of the subproblem

$$(P^k) \quad \inf\{F^k(x) \mid x \in D^k \cap Q\},$$

in the sense that (\tilde{P}^k) can be solved efficiently by a classical gradient or Newton-type method. When Q is polyhedral and when the functions f_t and g_t are affine, then (P^k) is a linear subproblem which is usually solved by the simplex dual method. But when the cardinality of T_i^k , $|T_i^k|$, grows beyond a certain limit, it is well-known that slow convergence arises and one way to overcome this drawback is to control $|T_i^k|$ by some constraint dropping schemes. The reader is referred to Section 3.1 and 3.2 of the survey of Reemtsen-Gorner [22] for a review of the extensive literature on this particular subject.

Concerning CSIP, numerous known methods consist in solving an approximating convex problem (\tilde{P}^k) . Supposing that F is \mathcal{C}^1 (as it is generally the case in ordinary CSIP), we can use cutting plane methods of Cheney-Goldstein [10], Kelley [14], Veinott [31], or Elzinga-Moore [11], and their variants (see, e.g., [22] for more references). Applied to LSIP, especially [10] and [14] turn out to be identical or mere modifications of the dual simplex method discussed above, so that they have similar properties and drawbacks. To avoid slow convergence, again constraint dropping rules are given under some conditions as strict convexity on F for [10] and [14]. We refer again the reader to Section 4 of [22] for more information on this subject. In this paper we consider another type of approximation for (P^k) :

$$(\tilde{P}^k) \quad \inf\{\tilde{F}^k(x) + \tilde{G}^k(x) \mid x \in Q\}.$$

Here \tilde{F}^k approximates F^k and \tilde{G}^k approximates the indicator function of D^k , δ_{D^k} (i.e. $\delta_{D^k}(x) = 0$ if $x \in D^k$, $\delta_{D^k}(x) = +\infty$ otherwise), so that the data which define (\tilde{P}^k) are \mathcal{C}^1 . There are many ways to smooth F^k (see in particular [13] and [21]), but for the sake of simplicity we consider here only the most important, and widely used in the literature in different fields. It is based on the smoothing of $\max\{\lambda_i : i = 1 \dots m\}$ by the function $\log(\sum_{i=1}^m \exp(\lambda_i p))/p$, with $p > 0$. More precisely, this smoothing gives here

$$\tilde{F}^k(x) := \frac{\log(\sum_{t \in T_1^k} \exp(f(t, x)p))}{p}, \quad \text{with } p = [\log |T_1^k|]^2. \quad (2)$$

This type of smoothing has been proposed by many authors for solving convex finite minimax problems, in particular by Bertsekas [7], Ben-Tal and Teboulle [6], Alvarez [1], and Nesterov [18]. This smoothing approach has been also proposed by Polak-Royset-Womersley [20], by Sheu-Wu [27] for finite min-max problems subject to infinitely many linear constraints and, more recently, by Sheu-Lin [26] for continuous min-max problems, motivated by the global approach of Fang-Wu [12] using an integral analog. We must also smooth the function δ_{D^k} and to do that we consider the smoothing approach by penalty and barrier functions introduced, for ordinary convex programs, by Auslender-Cominetti-Haddou [3]. These authors exploited the notion of recession functions to provide a wide class of penalty and barrier methods for usual convex programs, with a finite number of inequalities. In this paper we consider only penalty methods. Indeed there are some drawbacks with barrier methods, in particular the choice at each Step k of an interior point as a starting point. So we consider here two subclasses of penalty functions introduced in [3] (not all can be used). They are composed by those functions $\theta : \mathbb{R} \rightarrow \mathbb{R}_+$ which are \mathcal{C}^1 , convex, nondecreasing, and satisfy some additional properties, and we choose

$$\tilde{G}^k(x) := \frac{\gamma_k \sum_{t \in T_2^k} \theta(g(t, x)\delta_k)}{|T_2^k| \delta_k}, \quad (3)$$

with appropriated sequences of positive scalars $\{\gamma_k\}$ and $\{\delta_k\}$.

To summarize, we propose in Section 3 the Remez-type algorithm described above, where in Step 1 x^k

is an approximate optimal solution of a suitable regularization of (\tilde{P}^k) , with the smoothing and the penalization given by (2) and (3), while in Step 3 we choose $T_i^{k+1} = T_i^k \cup \{t_i^{k+1}\}$. The efficiency of the algorithm will depend on the subroutine used to compute x^k . With these approximations, \tilde{F}^k and \tilde{G}^k , when Q is the whole space, the problem (\tilde{P}^k) becomes an unconstrained convex smooth problem for which gradient or Newton-type methods can be used. The same holds when Q is “simple” (a box, the positive orthant, a ball, a simplex, ...). Convergence is established under the following minimal assumption: “ F is level bounded on the feasible set” and not under the assumption that Q is bounded. Furthermore, in Section 4 we associate with the sequence $\{x^k\}$ generated in Section 3 by the algorithm a dual sequence of measures for which we prove convergence to optimal solutions of the classical dual problem associated with (P) .

In this context, with Remez-type approximations (Step 2), Sheu-Lin [26] proposed the so-called entropic smoothing method for the min-max program where $T_2 = \emptyset$. Concerning ordinary CSIP ($|T_1| = 1$), to the best of our knowledge, Remez-type algorithms coupled with penalty methods have been only introduced by Martinet [17]. Comparisons with these two works are established in Comment 1, Section 3. On the other hand, particular penalty and smoothing functions and methods have been introduced for solving semi-infinite programs in three other contexts. Special penalty functions appear in the context of local reduction methods (see, e.g., Section 5.2 of [22] and references therein). In another context they are coupled with adaptive grid methods (see for example Kaplan-Tichatschke [15], Polak-Royset [19], and references therein) where the parameters of the procedures of discretization, smoothing, regularization, and penalization are adjusted. The third context concerns penalty, barrier and smoothing methods coupled with integral methods, and we investigate this field. This kind of integral methods has been studied by many researchers (see, e.g., Auslender [2], Teboulle [28], Teo-Goh [29], Teo-Rehbock-Jennings [30], Lin-Fang-Wu [16], Schattler [25], Polak-Higgins-Mayne [21]), and Fang-Wu [12]) and has the advantage to avoid nonconvex global optimization in Step 2 of Remez-type methods, via integrals which convexify the approximated functions. In this paper we do not consider barrier methods, and in Section 5 we propose an algorithm for solving (P) which consists of computing at Step k an optimal solution x^k of the convex C^1 approximating problem

$$(P_{ips}^k) \quad \inf\{\tilde{F}^k(x) + \tilde{G}^k(x) \mid x \in Q\}, \tag{4}$$

where

$$\tilde{F}^k(x) = \frac{1}{p_k} \log \left(\int_{T_1} \exp(f(t, x)p_k) dt \right), \quad \tilde{G}^k(x) = \gamma_k \int_{T_2} \frac{\theta(g(t, x)\delta_k)}{\delta_k} dt.$$

In this formula the parameters γ_k , δ_k , and p_k will be adjusted for obtaining convergence. In fact, and also in Section 3, we regularize the objective function by adding a term $\epsilon_k \|x\|^2$ with $\epsilon_k > 0$ and we compute an ϵ_k -optimal solution of the regularized problem. This regularization stabilizes the algorithm and provides an implementable subroutine. Without this regularization ($\epsilon_k = 0 \forall k$), this unified framework contains in particular the classical penalty and smoothing methods introduced in [2], [12], [16], and [30] but also provides new penalty and smoothing methods. Again, convergence is shown under the following minimal assumption: “ F is level bounded on the feasible set” and not under the assumption that Q is bounded. This requires, as for Remez-type algorithms, an analysis more subtle than usual, which is built on the use of the theory of recession functions developed in [4]. Convergence, also for the dual sequence of measures associated to the primal sequence, is established under the additional Slater’s condition. As pointed out in Comment 2, Section 5, our assumptions are weaker than those used in [2], [12], [16], and [30]. Finally, since the algorithms as well as the convergence analysis are built on the use of the theory of the recession functions, we recall in the next section the material from this theory which is needed in the sequel.

2. Preliminaries Given a set $Q \subset \mathbb{R}^n$, we denote by $\text{cl } Q$, $\text{int } Q$, $\text{conv } Q$, and $\text{cone } Q$ the closure, the interior, the convex hull, and the conical convex hull of Q , respectively. We associate with $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ its domain $\text{dom } f := \{x : f(x) < +\infty\}$ and its epigraph $\text{epi } f := \{(x, r) : f(x) \leq r\}$.

We recall here some basic notions about asymptotic cones and functions (for more details see, for instance, the books of Auslender and Teboulle [4] and of Rockafellar [24]).

The asymptotic cone of a set $Q \subseteq \mathbb{R}^n$ is defined to be

$$Q_\infty = \left\{ d : \exists \lambda_k \rightarrow +\infty, x_k \in Q \text{ with } d = \lim_{k \rightarrow \infty} \frac{x_k}{\lambda_k} \right\}. \tag{5}$$

When Q is convex and closed, it coincides with its recession cone

$$0^+(Q) := \{d : x + \lambda d \in Q \ \forall \lambda > 0, \forall x \in Q\}. \tag{6}$$

Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be lsc and proper (i.e., $\text{dom } f \neq \emptyset$). We recall that the asymptotic function f_∞ of f is defined through the relation

$$\text{epi } f_\infty = (\text{epi } f)_\infty.$$

As a straightforward consequence, we get (cf. [4, Theorem 2.5.1])

$$f_\infty(d) = \inf \left\{ \liminf_{k \rightarrow +\infty} \frac{f(\lambda_k x^k)}{\lambda_k} : \lambda_k \rightarrow +\infty, x^k \rightarrow d \right\}, \quad (7)$$

where $\{\lambda_k\} \subset \mathbb{R}$ and $\{x^k\} \subset \mathbb{R}^n$. Note that f_∞ is positively homogeneous, that is

$$f_\infty(\lambda d) = \lambda f_\infty(d) \quad \forall d, \forall \lambda > 0. \quad (8)$$

REMARK 2.1 (7) is fundamental in the convergence analysis of unbounded sequences and it is often used in the following way: Let $\{x^k\}$ be a sequence satisfying

$$\lim_{k \rightarrow \infty} \|x^k\| = +\infty, \quad \lim_{k \rightarrow \infty} \frac{x^k}{\|x^k\|} = d,$$

and let $\alpha \in \mathbb{R}$ so that $f_\infty(d) > \alpha$. Then it follows from (7) that for all k sufficiently large we have

$$f(x^k) = f \left(\|x^k\| \frac{x^k}{\|x^k\|} \right) \geq \alpha \|x^k\|.$$

When f is a proper lsc convex function its asymptotic function is also a proper lsc convex function that coincides with the recession function

$$0^+ f(d) = \lim_{\lambda \rightarrow +\infty} \frac{f(x + \lambda d) - f(x)}{\lambda} \quad \forall x \in \text{dom } f, \quad (9)$$

which implies that

$$f_\infty(d) = \lim_{\lambda \rightarrow +\infty} \frac{f(\lambda d)}{\lambda} \quad \forall d \in \text{dom } f. \quad (10)$$

Furthermore,

$$(\delta_Q)_\infty = \delta_{Q_\infty}. \quad (11)$$

If $f, g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper lsc convex functions, and $\text{dom}(f + g) \neq \emptyset$, then

$$(f + g)_\infty(d) = f_\infty(d) + g_\infty(d). \quad (12)$$

Furthermore when $\{f_i\}_{i \in I}$ is a family of proper lsc convex functions defined on \mathbb{R}^n with values in $\mathbb{R} \cup \{+\infty\}$ and the function $f = \sup_{i \in I} f_i$ is proper, then we have

$$f_\infty = \sup_{i \in I} (f_i)_\infty. \quad (13)$$

When f is a proper lsc convex function, a useful consequence of (6) and (9) is the equation

$$\{x : f(x) \leq \lambda\}_\infty = \{d : f_\infty(d) \leq 0\}_\infty, \quad (14)$$

for any λ such that $\{x : f(x) \leq \lambda\} \neq \emptyset$.

The following proposition is crucial in the convergence analysis. The reader can find a proof in [4, Chapter 3].

PROPOSITION 2.1 Let Q be a closed convex set in \mathbb{R}^n and let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper lsc convex function such that $Q \cap \text{dom } f \neq \emptyset$. Consider the optimization problem

$$(P) \quad f^* = \inf \{f(x) \mid x \in Q\}.$$

Then a necessary and sufficient condition for the optimal set of (P) to be nonempty and compact is given by

$$f_\infty(d) \leq 0 \quad \text{and} \quad d \in Q_\infty \Rightarrow d = 0,$$

or equivalently f is level-bounded on Q , i.e., for every λ , $\{x \in Q : f(x) \leq \lambda\}$ is compact when nonempty. This is equivalent to

$$\lim_{\|x\| \rightarrow \infty, x \in Q} f(x) = +\infty.$$

In our analysis, the composite function is of a particular interest. More precisely, we consider the composition between a penalty function $\theta \in \mathcal{F}$ and a convex function f , where

$$\mathcal{F} = \left\{ \begin{array}{l} \theta : \mathbb{R} \rightarrow \mathbb{R}_+ \text{ convex, nondecreasing, nonconstant,} \\ \mathcal{C}^1, \text{ and such that } \lim_{u \rightarrow -\infty} \theta(u) = 0 \end{array} \right\}. \quad (15)$$

Since $\theta \in \mathcal{F}$ takes nonnegative values and it is nondecreasing, we have $\theta_\infty(-1) = 0$. Then, since it is nonconstant, $\theta_\infty(1) > 0$. The following result was proved in [3] in a more general setting.

PROPOSITION 2.2 *Let $\theta \in \mathcal{F}$, and let f be a proper lsc convex function, and consider the composite function*

$$g(x) = \begin{cases} \theta(f(x)), & \text{if } x \in \text{dom } f, \\ +\infty, & \text{otherwise.} \end{cases}$$

Then g is a proper lsc convex function.

In the rest of this paper we consider the following two subsets of \mathcal{F} :

$$\mathcal{F}_1 = \{\theta \in \mathcal{F} : \theta_\infty(1) < +\infty\} \text{ and } \mathcal{F}_2 = \{\theta \in \mathcal{F} : \theta_\infty(1) = +\infty\}. \quad (16)$$

Obviously the function $\theta(u) = u^+ := \max\{u, 0\}$ which has been used in the literature is not \mathcal{C}^1 , but satisfies all the other properties required for \mathcal{F}_1 . However this function, for which our convergence analysis holds, is out of interest for our purpose since it is not smooth.

In [9], Chen and Mangasarian provided a systematic way to generate elements of \mathcal{F}_1 . These are smooth approximations of the function u^+ and are built as follows. Let p be a positive piecewise continuous probability density function, with a finite number of pieces. Let $F(t) = \int_{-\infty}^t p(s)ds$ be the associated distribution function and suppose that $\theta(u) = \int_{-\infty}^u F(t)dt$ is well defined. Then we have ([9, Proposition 2.2]) that θ is a strictly convex \mathcal{C}^1 function from \mathbb{R} to \mathbb{R}_+ , strictly increasing, with

$$0 < \theta'(u) < 1, \quad -M_2 \leq \theta(u) - u^+ \leq M_1 \quad \forall u \in \mathbb{R}, \quad (17)$$

where $M_1 := \int_{-\infty}^0 |s|p(s)ds$ and $M_2 := [\int_{-\infty}^{+\infty} sp(s)ds]^+$, provided that $M_i < +\infty$, $i = 1, 2$. From these inequalities and the definition of θ , it follows that

$$\left| \frac{\theta(\lambda u)}{\lambda} - u^+ \right| \leq \frac{\max\{M_1, M_2\}}{\lambda} \quad \forall \lambda > 0, \forall u \in \mathbb{R}, \quad \theta_\infty(1) = 1, \quad \lim_{u \rightarrow -\infty} \theta(u) = 0, \quad (18)$$

so that $\theta \in \mathcal{F}_1$. Specific cases of interest are

$$\theta_1(u) = \log(1 + \exp(u)), \quad \text{with } p_1(s) = \frac{\exp(-s)}{(1 + \exp(-s))^2},$$

$$\theta_2(u) = 2^{-1}(u + \sqrt{u^2 + 4}), \quad \text{with } p_2(s) = \frac{2}{(s^2 + 4)^{\frac{3}{2}}},$$

and

$$\theta_3(u) = \begin{cases} 0, & u \leq -1, \\ \frac{1}{4}(u+1)^2, & -1 < u < 1, \\ u, & u \geq 1, \end{cases} \quad \text{with } p_3(s) = \begin{cases} \frac{1}{2}, & -1 \leq s \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Finally, as well-known penalty functions which belong to \mathcal{F}_2 we have the classical penalty functions and the exponential function:

$$\theta_4(u) = \frac{1}{2}(u^+)^2, \quad \theta_5(u) = (u^+)^3, \quad \text{and } \theta_6(u) = \exp(u).$$

3. Remez-type algorithm coupled with penalty and smoothing methods In this section we consider the optimization problem (P) described in (1), satisfying the given assumptions on the data, and the Remez-type algorithm described in the introduction. For the sake of simplicity we choose

$$T_i^{k+1} = T_i^k \cup \{t_i^{k+1}\}, \quad i = 1, 2, \quad (19)$$

where $t_1^{k+1} \in T_1$ and $t_2^{k+1} \in T_2$ solve approximately the auxiliary problems in Step 2, i.e.,

$$\begin{cases} f(t_1^{k+1}, x^k) \geq \max\{f(t, x^k) \mid t \in T_1\} - \mu_k, \\ g(t_2^{k+1}, x^k) \geq \max\{g(t, x^k) \mid t \in T_2\} - \mu_k, \end{cases} \quad (20)$$

with

$$\mu_k \geq 0 \quad \forall k \quad \text{and} \quad \lim_{k \rightarrow +\infty} \mu_k = 0. \quad (21)$$

From now on in this section we consider the following assumption:

(A₁) F is level bounded on C ,

and sometimes we shall assume also:

(A₂) Slater's condition holds; i.e., there exists $u \in Q$ such that $G(u) < 0$.

Following Proposition 2.1 we remark that Assumption (A₁) is equivalent to the implication

$$F_\infty(d) \leq 0, \quad G_\infty(d) \leq 0 \quad \text{and} \quad d \in Q_\infty \Rightarrow d = 0. \quad (22)$$

The following lemma shows that the existence of starting sets for the first algorithm of Remez with nice properties, is a consequence of Assumption (A₁). It was proved in [22, Lemma 2.4] when $|T_1| = 1$ and $Q = \mathbb{R}^n$. Here we give a completely new and different proof for the general case, more concise and based on the properties of the recession functions.

LEMMA 3.1 *Assume that (A₁) holds. Then, there exist finite nonempty subsets $T_1^0 \subset T_1$ and $T_2^0 \subset T_2$ such that F^0 is level bounded on C^0 .*

Proof. Since $F_\infty = \sup_{t \in T_1} (f_t)_\infty$ and $G_\infty = \sup_{t \in T_2} (g_t)_\infty$, by (A₁) and Proposition 2.1,

$$\{d : (f_t)_\infty(d) \leq 0 \quad \forall t \in T_1; (g_t)_\infty(d) \leq 0 \quad \forall t \in T_2; \delta_{Q_\infty}(d) \leq 0\} = \{0\}. \quad (23)$$

Let $d \mapsto \langle a_i, d \rangle$, $i \in I$, be the family of all the linear minorants of all the functions in (23), which are positively homogeneous, proper, lsc, and convex. Then (23) holds if and only if $\{d : \langle a_i, d \rangle \leq 0, i \in I\} = \{0\}$, i.e., $\text{cone}\{a_i, i \in I\} = \mathbb{R}^n$ (or, equivalently, $0 \in \text{int conv}\{a_i, i \in I\}$). This happens if and only if there exists $J \subset I$, $|J| = n + 1$, such that $\text{cone}\{a_i, i \in J\} = \mathbb{R}^n$ (or, equivalently, $0 \in \text{int conv}\{a_i, i \in J\}$). In that case $\{d : \langle a_i, d \rangle \leq 0, i \in J\} = \{0\}$. Replacing in (23) each linear function $d \mapsto \langle a_i, d \rangle$, $i \in J$, by one of the minorized constraint functions, we conclude the existence of $T_1^0 \subset T_1$ and $T_2^0 \subset T_2$, $|T_1^0 \cup T_2^0| \leq n + 1$, such that

$$\{d : (f_t)_\infty(d) \leq 0 \quad \forall t \in T_1^0; (g_t)_\infty(d) \leq 0 \quad \forall t \in T_2^0; \delta_{Q_\infty}(d) \leq 0\} = \{0\}. \quad (24)$$

Thus F^0 is level bounded on C^0 , again by Proposition 2.1. Finally, if $T_i^0 = \emptyset$, replacing it with $T_i^0 = \{t_i\}$ for an arbitrary $t_i \in T_i$, $i = 1, 2$, we get the aimed conclusion. \square

REMARK 3.1 *There are some particular cases where the sets T_i^0 are easily obtainable:*

- (a) *If Q is bounded, we can take $T_1^0 = \{t_1\}$ and $T_2^0 = \{t_2\}$ for any $t_i \in T_i$, $i = 1, 2$.*
- (b) *If for some $t_1^0 \in T_1$, $f_{t_1^0}$ (resp., $t_2^0 \in T_2$, $f_{t_2^0}$) is level bounded on Q , we can take $T_1^0 = \{t_1^0\}$ and $T_2^0 = \{t_2\}$, with t_2 arbitrary in T_2 (resp., $T_2^0 = \{t_2^0\}$ and $T_1^0 = \{t_1\}$, with t_1 arbitrary in T_1).*
- (c) *In LSIP, $Q = \mathbb{R}^n$, $f(t, x) = \langle a(t), x \rangle - b(t) \quad \forall t \in T_1$, and $g(t, x) = \langle a(t), x \rangle - b(t) \quad \forall t \in T_2$. In that case, if $T_i = \text{clint } T_i \subset \mathbb{R}^{n_i}$, $i = 1, 2$, (as it happens in practice) taking a sequence of real numbers $\beta_r \searrow 0$,*

then $\text{dist}(T_i \cap \beta_r \mathbb{Z}^{n_i}, T_i) \rightarrow 0$, $i = 1, 2$. Since (24) becomes $\{d : \langle a(t), d \rangle \leq 0, t \in T_1^0 \cup T_2^0\} = \{0\}$, i.e., $0 \in \text{int conv} \{a(t), t \in T_1^0 \cup T_2^0\}$, we can take the regular grids $T_i^0 = T_i \cap \beta_r \mathbb{Z}^{n_i}$, $i = 1, 2$, for sufficiently large r .

(d) In ordinary CSIP (with $T_1 = \{t_1\}$), if $T_2 = \text{clint } T_2 \subset \mathbb{R}^m$ and $\beta_r \searrow 0$, since $\text{dist}(T_2 \cap \beta_r \mathbb{Z}^m, T_2) \rightarrow 0$, it is possible to take the regular grid $T_2^0 = T_2 \cap \beta_r \mathbb{Z}^m$ for sufficiently large r by the argument of [22, Lemma 2.4].

Denote $r_i = |T_i^0|$, the cardinality of T_i^0 . Then $r_i^k := |T_i^k| \leq r_i + k$, $i = 1, 2$. As it was said in the Introduction, we can use for approximating F^k the function

$$\tilde{F}_p^k(x) := \frac{\log(\sum_{t \in T_1^k} \exp(f(t, x)p))}{p},$$

with $p > 0$. It is well known that this function is convex (sum of log-convex functions) and that we have the uniform estimate (see, for example, [27])

$$0 \leq \tilde{F}_p^k(x) - F^k(x) \leq \frac{\log(|T_1^k|)}{p}, \quad \forall x \in \mathbb{R}^n.$$

If T_1 is reduced to a single point, it is worthwhile to note that $\tilde{F}_p^k(x) = F^k(x) = F(x)$ and that in Step 2 the computation of t_1^{k+1} is unnecessary. From now on for each k we set $p_k = \lceil \log(r_1 + k) \rceil$, and use the approximating function

$$\tilde{F}^k := \tilde{F}_{p_k}^k, \tag{25}$$

so that

$$0 \leq \tilde{F}^k(x) - F^k(x) \leq \frac{1}{\log(r_1 + k)}, \quad \forall x \in \mathbb{R}^n. \tag{26}$$

Let now $\{\epsilon_k\}$ be a sequence of real numbers such that

$$\epsilon_k > 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} \epsilon_k = 0. \tag{27}$$

Let $\theta \in \mathcal{F}$ and let $\{\delta_k\}, \{\gamma_k\}$ be sequences of positive real numbers. Recalling (3), we define for $k = 1, 2, \dots$, the approximating functions

$$\tilde{G}^k(x) := \frac{\gamma_k}{|T_2^k|} \sum_{t \in T_2^k} \frac{\theta(g(t, x)\delta_k)}{\delta_k}, \quad \tilde{H}^k(x) := \tilde{F}^k(x) + \tilde{G}^k(x).$$

which are convex by Proposition 2.2. Associated with these functions we introduce the regularized subproblem

$$(\tilde{P}_{re}^k) \quad \inf \{ \tilde{H}^k(x) + \epsilon_k \|x\|^2 \mid x \in Q \}.$$

This subproblem will be solved in Step 1 of the forthcoming algorithm within an error ϵ_k .

REMARK 3.2 *The objective function $H_{re}^k(\cdot) := \tilde{H}^k(\cdot) + \epsilon_k \|\cdot\|^2$ is strongly convex, so that $\text{argmin}\{\tilde{H}^k(x) + \epsilon_k \|x\|^2 \mid x \in Q\}$ is a single point y^k . As a consequence, there exists at least a point x^k satisfying*

$$x^k \in Q, \quad \tilde{H}^k(x^k) + \epsilon_k \|x^k\|^2 \leq \tilde{H}^k(x) + \epsilon_k \|x\|^2 + \epsilon_k \quad \forall x \in Q. \tag{28}$$

When \tilde{H}^k is \mathcal{C}^1 , and $Q = \mathbb{R}^n$, then it is worthwhile to note that any usual convergent gradient method will provide in a finite number of steps such a point by using the implementable stopping rule

$$\|\nabla \tilde{H}^k(x^k) + 2\epsilon_k x^k\| \leq \sqrt{2}\epsilon_k. \tag{29}$$

Indeed, writing the strong convex inequality $\langle \nabla H_{re}^k(x^k) - \nabla H_{re}^k(y^k), x^k - y^k \rangle \geq 2\epsilon_k \|x^k - y^k\|^2$, it follows from (29) that $\|x^k - y^k\| \leq \frac{1}{\sqrt{2}}$. Since $H_{re}^k(y^k) \geq H_{re}^k(x^k) + \langle \nabla H_{re}^k(x^k), y^k - x^k \rangle$, using again (29), we obtain (28).

Note that this implementable stopping rule does not imply (28) if the original objective function \tilde{H}^k is not regularized by adding $\epsilon_k \|x\|^2$.

Now we can describe our basic algorithm:

The Remez penalty smoothing algorithm-RPSALG

Initialization: Set $k = 0$ and start with T_1^0 and T_2^0 defined in Lemma 3.1.

Step 1 Compute x^k satisfying (28).

Step 2 Compute t_1^{k+1} and t_2^{k+1} satisfying (20) with (21).

Step 3 Set: $T_i^{k+1} = T_i^k \cup \{t_i^{k+1}\}$ $i = 1, 2$.

Set $k \leftarrow k + 1$; go to Step 1.

Each triple $(\theta, \{\gamma_k\}, \{\delta_k\})$ determines a different instance of **RPSALG**. In order to prove its convergence we consider the following conditions involving a sequence $\{m_k\}$ such that $m_k \geq |T_2^k| \forall k$:

(a) $\theta \in \mathcal{F}_1$, $\lim_{k \rightarrow \infty} \frac{\gamma_k}{\delta_k} = 0$, and $\lim_{k \rightarrow \infty} \frac{\gamma_k}{m_k} = +\infty$.

(b) $\theta \in \mathcal{F}_2$, $\lim_{k \rightarrow \infty} \frac{\gamma_k}{\delta_k} = 0$, and $\frac{\gamma_k}{m_k} > \varepsilon \forall k$, for a certain $\varepsilon > 0$.

(c) $\theta \in \mathcal{F}_2$, $\lim_{k \rightarrow \infty} \delta_k = +\infty$, $\frac{\gamma_k}{m_k} > \varepsilon \forall k$ for a certain $\varepsilon > 0$, $\left\{ \frac{\gamma_k}{\delta_k} \right\}$ is bounded, and either $\theta(0) = 0$ or (A_2) holds.

REMARK 3.3 *The natural choice is $m_k = r_2 + k$. However another choice will be proposed at the end of this Section. Furthermore it is worthwhile to note that condition (a) as well as condition (b) implies that $\lim_{k \rightarrow \infty} \delta_k = +\infty$.*

THEOREM 3.1 *Assume that (A_1) holds. If $(\theta, \{\gamma_k\}, \{\delta_k\})$ satisfies at least one of the conditions (a), (b), (c), then the sequence built by **RPSALG** is bounded and each limit point of this sequence is an optimal solution of (P) .*

Proof. Let $u \in Q$ such that $G(u) < 0$ if (A_2) holds and $u \in C$ otherwise.

1) Let $l \leq k$ be fixed nonnegative integers. Since $F^l \leq F^k \leq F$ from the definition, using (26) in the basic inequality (28) we get

$$F^l(x^k) + \frac{\gamma_k}{|T_2^k|} \sum_{t \in T_2^k} \frac{\theta(g(t, x^k)\delta_k)}{\delta_k} \leq F(u) + \frac{\gamma_k}{|T_2^k|} \sum_{t \in T_2^k} \frac{\theta(g(t, u)\delta_k)}{\delta_k} + \nu_k(u),$$

with

$$\nu_k(u) := \frac{1}{\log(r_1 + k)} + \epsilon_k \|u\|^2 + \epsilon_k, \quad \lim_{k \rightarrow \infty} \nu_k(u) = 0. \quad (30)$$

Since θ is nondecreasing and nonnegative it follows that

$$\begin{aligned} F^l(x^k) + \frac{\gamma_k}{|T_2^k|} \sum_{t \in T_2^k} \frac{\theta(g(t, x^k)\delta_k)}{\delta_k} &\leq F(u) + \frac{\gamma_k}{|T_2^k|} \sum_{t \in T_2^k} \frac{\theta(G(u)\delta_k)}{\delta_k} + \nu_k(u) \\ &\leq F(u) + \frac{\gamma_k}{\delta_k} \theta(G(u)\delta_k) + \nu_k(u). \end{aligned} \quad (31)$$

2) Let us prove now that the sequence $\{x^k\}$ is bounded. Suppose the contrary. Then there exists a subsequence $\{x^k\}_{k \in K}$ such that

$$\lim_{\substack{k \rightarrow +\infty \\ k \in K}} \|x^k\| = +\infty, \quad \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \frac{x^k}{\|x^k\|} = d \neq 0, \quad d \in Q_\infty. \quad (32)$$

Let l be arbitrary. Let $\alpha_t^l < g(t, \cdot)_\infty(d) \forall t \in T_2^l$ and $\beta^l < (F^l)_\infty(d)$. Then, as pointed out in Remark 2.1, there exists k_l such that

$$F^l(x^k) \geq \beta^l \|x^k\|, \quad g(t, x^k) \geq \alpha_t^l \|x^k\| \forall t \in T_2^l, \quad \forall k \in K \text{ such that } k \geq \max\{l, k_l\}.$$

Since θ is nonnegative and nondecreasing, dividing both members of inequality (31) by $\|x^k\|$ we deduce

$$\beta^l + \frac{\gamma_k}{|T_2^k|} \frac{\theta(\alpha_t^l \|x^k\| \delta_k)}{\|x^k\| \delta_k} \leq \frac{F(u)}{\|x^k\|} + \frac{\gamma_k \theta(G(u)\delta_k)}{\|x^k\| \delta_k} + \frac{\nu_k(u)}{\|x^k\|}, \quad \forall t \in T_2^l. \quad (33)$$

Since θ is nondecreasing and $\delta_k G(u) \leq 0$, then $\theta(G(u)\delta_k) \leq \theta(0) \forall k$ so that the right hand side of (33) converges to zero as $k \rightarrow \infty$. As a consequence, $\forall \tau^l > 0$ we have for k large enough

$$\beta^l + \frac{\gamma_k}{|T_2^k|} \frac{\theta(\alpha_t^l \|x^k\| \delta_k)}{\|x^k\| \delta_k} \leq \tau^l \quad \forall t \in T_2^l. \quad (34)$$

Let us show now that

$$\alpha_t^l \leq 0 \quad \forall t \in T_2^l. \quad (35)$$

Suppose the contrary, i.e., there exists some $t \in T_2^l$ with $\alpha_t^l > 0$. Consider

$$\lim_{k \rightarrow \infty} \frac{\gamma_k}{|T_2^k|} \frac{\theta(\alpha_t^l \|x^k\| \delta_k)}{\|x^k\| \delta_k} = \lim_{k \rightarrow \infty} \left(\frac{\alpha_t^l \gamma_k}{|T_2^k|} \right) \left[\frac{\theta(\alpha_t^l \|x^k\| \delta_k)}{\alpha_t^l \|x^k\| \delta_k} \right] \geq \lim_{k \rightarrow \infty} \left(\frac{\alpha_t^l \gamma_k}{m_k} \right) \left[\frac{\theta(\alpha_t^l \|x^k\| \delta_k)}{\alpha_t^l \|x^k\| \delta_k} \right], \quad (36)$$

where $\lim_{k \rightarrow \infty} \frac{\theta(\alpha_t^l \|x^k\| \delta_k)}{\alpha_t^l \|x^k\| \delta_k} = \theta_\infty(1)$. It is easy to see that the limit in (36) is $+\infty$ under any of the assumptions (a), (b) and (c), in contradiction with (34). Thus (35) holds.

Furthermore, since θ is nonnegative we deduce from (34) that $\beta^l \leq \tau^l$ for all positive τ^l , i.e., $\beta^l \leq 0$. From (35), letting $\beta^l \rightarrow (F^l)_\infty(d)$, $\alpha_t^l \rightarrow g(t, \cdot)_\infty(d) \forall t \in T_2^l$, it follows that

$$(F^l)_\infty(d) \leq 0, \quad g(t, \cdot)_\infty(d) \leq 0 \quad \forall t \in T_2^l.$$

Therefore, if we set $l = 0$ we get, together with $d \in Q_\infty$, a contradiction with the fact that F^0 is level bounded on C^0 .

3) Now let x_∞ be a limit point of the sequence $\{x^k\}$. Since Q is closed, $x_\infty \in Q$. Furthermore since T_1 and T_2 are compact, there exist $\bar{t}_i \in T_i$, $i = 1, 2$ and subsequences $\{x^k\}_{k \in K}$, $\{t_i^{k+1}\}_{k \in K}$, $i = 1, 2$, such that

$$\lim_{k \in K} x^k = x_\infty, \quad \lim_{k \in K} t_i^{k+1} = \bar{t}_i, \quad i = 1, 2. \quad (37)$$

Let l be arbitrary. Let $\tilde{\alpha}_t^l < g(t, x_\infty)$, $\forall t \in T_2^l$, $\tilde{\beta}^l < F^l(x_\infty)$. Then by continuity there exists k_l such that

$$F^l(x^k) \geq \tilde{\beta}^l, \quad g(t, x^k) \geq \tilde{\alpha}_t^l \quad \forall t \in T_2^l, \quad \forall k \in K \text{ such that } k \geq \max\{l, k_l\}.$$

As a consequence, since θ is nondecreasing, we deduce from inequality (31) that

$$\tilde{\beta}^l + \sum_{t \in T_2^l} \frac{\gamma_k}{|T_2^k|} \frac{\theta(\tilde{\alpha}_t^l \delta_k)}{\delta_k} \leq F(u) + \gamma_k \frac{\theta(G(u)\delta_k)}{\delta_k} + \nu_k(u), \quad \text{with } \lim_{k \rightarrow \infty} \nu_k(u) = 0. \quad (38)$$

Since $0 \leq \theta(G(u)\delta_k) \leq \theta(0)$ and $\lim_{u \rightarrow -\infty} \theta(u) = 0$, then

$$\lim_{k \rightarrow \infty} \gamma_k \frac{\theta(G(u)\delta_k)}{\delta_k} = 0 \quad (39)$$

under one of the conditions (a), (b), (c).

It follows that the right hand side of (38) converges to $F(u)$ as $k \rightarrow \infty$. As a consequence, since θ is nonnegative, we get that

$$\tilde{\beta}^l \leq F(u), \quad \tilde{\beta}^l + \frac{\gamma_k}{|T_2^k|} \frac{\theta(\tilde{\alpha}_t^l \delta_k)}{\delta_k} \leq F(u) + \kappa_k(u) \quad \forall t \in T_2^l, \quad \text{where } \lim_{k \rightarrow \infty} \kappa_k(u) = 0. \quad (40)$$

Repeating the same arguments as in part 2), we deduce from (40) that

$$\tilde{\alpha}_t^l \leq 0 \quad \forall t \in T_2^l. \quad (41)$$

Letting $\tilde{\beta}^l \rightarrow F^l(x_\infty)$, $\tilde{\alpha}_t^l \rightarrow g(t, x_\infty) \forall t \in T_2^l$ it follows from (40) and (41) that

$$F^l(x_\infty) \leq F(u), \quad (42)$$

and

$$g(t, x_\infty) \leq 0 \quad \forall t \in T_2^l \forall l. \quad (43)$$

Now passing to the limit as $l \rightarrow +\infty$ we get

$$g(\bar{t}_2, x_\infty) \leq 0. \quad (44)$$

Finally

$$g(t_2^{k+1}, x_\infty) = g(t_2^{k+1}, x^k) + [g(t_2^{k+1}, x_\infty) - g(t_2^{k+1}, x^k)],$$

so that, according to (20),

$$g(t_2^{k+1}, x_\infty) \geq G(x^k) - \mu_k + [g(t_2^{k+1}, x_\infty) - g(t_2^{k+1}, x^k)]. \quad (45)$$

Passing to the limit, using (21), (37), (44), and the fact that G and g are continuous we get $G(x_\infty) \leq 0$, so that $x_\infty \in C$.

Coming back to inequality (42), with $u \in C$ ($u \in Q$ and $G(u) < 0$ if (A_2) holds), by continuity we get

$$F^l(x_\infty) \leq v(P) \quad \forall l. \quad (46)$$

Now we define $j(l) = \max\{j \in K : j < l\}$. Then

$$F^l(x_\infty) \geq f(t_1^{j(l)+1}, x_\infty) = f(t_1^{j(l)+1}, x^{j(l)}) + [f(t_1^{j(l)+1}, x_\infty) - f(t_1^{j(l)+1}, x^{j(l)})].$$

According to (20), $F(x^{j(l)}) \leq f(t_1^{j(l)+1}, x^{j(l)}) + \mu_{j(l)}$. Passing to the limit in these inequalities, and using (46), we get $F(x_\infty) \leq v(P)$, which proves that $x_\infty \in S_P$. \square

REMARK 3.4 *The functions θ_4 and θ_5 satisfy the assumption $\theta(0) = 0$, but not θ_6 .*

REMARK 3.5 *When D is defined with a finite number of inequalities q , then we can take $T_2^0 = \{1, \dots, q\}$, and we do not need to compute in **RPSALG** the element t_2^{k+1} . Obviously the convergence proof remains valid. Furthermore in that case we can choose $m_k = q$ which leads to parameters γ_k smaller than for $m_k = r_2 + k$.*

REMARK 3.6 *A unified framework for penalty and barrier methods was developed in [4] for nonconvex programs containing a finite number of inequalities and semi-definite constraints. The convergence results given in Theorem 3.1 can be extended to the nonconvex setting in a similar way, but using much more sophisticated results on asymptotic functions (observe that some results of Section 2 are only valid for convex functions (e.g., (12), (13) or Proposition 2.1).*

Comment 1

In the min-max case ($T_2 = \emptyset$), **RPSALG** coincides with the entropic smoothing method proposed by Sheu-Lin [26], where convergence was obtained under the stronger condition: Q is compact. For ordinary CSIP ($|T_1| = 1$), Martinet proposed in [17] an algorithm similar to **RPSALG**, the difference being the formula giving the approximating penalized term. In fact, Martinet chose $\tilde{G}^k(x) = \alpha_k \sum_{t \in T_2^k} \theta(g(t, x))$, with $\alpha_k \geq 1$, instead of (3). The class of penalty functions considered in [17] consists of continuous functions $\theta : \mathbb{R} \rightarrow \mathbf{R}_+$ such that $\theta(t) = 0$ if $t \leq 0$. This is a very restrictive condition which is violated in particular by $\theta_1, \theta_2, \theta_3$ and by the exponential function θ_6 . Actually, this condition concerns essentially functions as θ_4 or θ_5 , for which the two frameworks coincide. With a completely different proof, convergence in [17] was obtained in the nonconvex case, but under the stronger assumption which imposes to F to be level bounded on Q instead on the feasible set $Q \cap D$, as in Theorem 3.1. Furthermore in [17] there is no duality analysis as in the following section.

It should also be noted that both schemes require summing up over T_2^k to evaluate the values of the penalized function and of its gradient. In this case deletion rules can be helpful to improve the models. Such a rule has been proposed in [17], where convergence is proved in the convex case, with the assumption just cited above but imposing the additional one that F is uniformly strictly convex.

4. Duality results In this section we assume, for the sake of simplicity, that $Q = \mathbb{R}^n$, that T_1 is reduced to a single point, so that F is \mathcal{C}^1 on the whole space \mathbb{R}^n , and we suppose that $\nabla_x g(\cdot, \cdot)$ exists and is continuous on $T_2 \times \mathbb{R}^n$. We use the following notation:

a) $\mathcal{C}(T_2)$ is the Banach space of real-valued continuous functions on T_2 , equipped with the maximum norm

$$\|h\| = \max\{|h(t)| : t \in T_2\}.$$

By $\mathcal{C}_+(T_2)$ we denote the cone of nonnegative valued functions in $\mathcal{C}(T_2)$.

b) $M(T_2)$ is its topological dual, i.e., the space of all the finite signed Borel measures on T_2 , embedded with the total variation norm. We have

$$\langle h, \sigma \rangle := \int_{T_2} h(t)\sigma(dt) \quad \forall \sigma \in M(T_2), \quad \forall h \in C(T_2).$$

Since T_2 is a metric space, $\mathcal{C}(T_2)$ is separable and every finite signed Borel measure on T_2 is regular (see, for instance, [8, Example 2.37]).

By $M_+(T_2)$ we represent the positive cone of $M(T_2)$, i.e., the subset of $M(T_2)$ composed by the finite Borel measures on T_2 . For $\sigma \in M_+(T_2)$ we have $\|\sigma\| = \int_{T_2} \sigma(dt)$.

c) $L(x, \sigma)$ is the usual Lagrangian function associated with (P) , i.e.,

$$L(x, \sigma) := F(x) + \langle g_x, \sigma \rangle = F(x) + \int_{T_2} g(t, x)\sigma(dt),$$

with $x \in \mathbb{R}^n$, $\sigma \in M_+(T_2)$, and $g_x(t) := g(t, x)$, for all $t \in T_2$.

d) Associated with the Lagrangian function we consider the function

$$\psi(\sigma) := \inf\{L(x, \sigma) \mid x \in \mathbb{R}^n\}.$$

e) Associated with our primal problem (P) we define the usual *Lagrangian dual problem*

$$(D) \quad v(D) = \sup_{\sigma \in M_+(T_2)} \inf_{x \in \mathbb{R}^n} L(x, \sigma) \equiv \sup_{\sigma \in M_+(T_2)} \psi(\sigma). \quad (47)$$

g) The so-called *weak duality inequality* $v(P) \geq v(D)$ always holds. The optimal set of (D) is denoted by S_D .

The following theorem gathers the most relevant properties of the dual pair.

THEOREM 4.1 *Assume that assumptions (A_1) and (A_2) are satisfied. Then, the following statements hold:*

(i) *The strong duality $v(D) = v(P)$ is satisfied, and the dual optimal set S_D is nonempty and bounded for the total variation norm.*

(ii) *If $\bar{\sigma} \in M_+(T_2)$ and $\bar{x} \in \operatorname{argmin} \psi(\bar{\sigma})$ are such that*

$$g(t, \bar{x}) \leq 0 \quad \forall t \in T_2, \quad \text{and} \quad \langle g_{\bar{x}}, \bar{\sigma} \rangle = 0,$$

then \bar{x} and $\bar{\sigma}$ are optimal for (P) and (D) , respectively. Moreover, under the current assumptions we have

$$\bar{x} \in \operatorname{argmin} \psi(\bar{\sigma}) \iff \nabla_x L(\bar{x}, \bar{\sigma}) = \nabla F(\bar{x}) + \int_{T_2} \nabla_x g(t, \bar{x})\bar{\sigma}(dt) = 0. \quad (48)$$

The proof comes straightforwardly from Theorems 5.97 and 5.98, Corollary 5.109, and (5.278) in [8].

Let us come back to algorithm **RPSALG** in which we compute a point $x^k \in \mathbb{R}^n$ satisfying the stopping rule

$$\left\| \nabla F(x^k) + \frac{\gamma_k}{|T_2^k|} \sum_{t \in T_2^k} \theta'(g(t, x^k)\delta_k) \nabla_x g(t, x^k) + 2\epsilon_k x^k \right\| \leq \sqrt{2}\epsilon_k. \quad (49)$$

For the rest of this section we suppose that $(\theta, \{\gamma_k\}, \{\delta_k\})$ satisfies at least one of the conditions (a), (b), (c'), where (a) and (b) are defined in Section 3, and

(c') $\theta \in \mathcal{F}_2$, $\lim_{k \rightarrow \infty} \delta_k = +\infty$, $\frac{\gamma_k}{m_k} > \epsilon \quad \forall k$ for a certain $\epsilon > 0$, $\left\{ \frac{\gamma_k}{\delta_k} \right\}$ is bounded, and $\theta(0) = 0$.

Thanks to Remark 3.2 the point x^k satisfies (28) so that Theorem 3.1 holds. Furthermore this inequality leads us to introduce the sequence of discrete measures $\{\sigma^k\}$ associated with the sequence $\{x^k\}$ by

$$\sigma^k := \frac{\gamma_k}{|T_2^k|} \sum_{t \in T_2^k} \theta'(g(t, x^k)\delta_k) \alpha_t, \quad (50)$$

where α_t is the Dirac distribution concentrated at point t .

Using Theorem 4.1 we get the following dual convergence theorem, in which we prove the weak*-convergence of a sequence $\{\sigma^k\} \in M(T_2)$, to some element $\sigma \in M(T_2)$, i.e.,

$$\lim_{k \rightarrow \infty} \langle h, \sigma^k \rangle = \langle h, \sigma \rangle, \quad \forall h \in \mathcal{C}(T_2).$$

THEOREM 4.2 Assume that (A_1) and (A_2) are satisfied, and suppose that $(\theta, \{\gamma_k\}, \{\delta_k\})$ satisfies at least one of the conditions (a), (b), (c'). Then, the following statements hold:

(i) The sequence $\{\sigma^k\}$ given in (50) is strongly bounded.

(ii) There exists at least a weak*-limit point of this sequence, and each weak*-limit point of this sequence belongs to S_D .

Proof. i-1) Let us consider the (possibly empty) set $I_k := \{t \in T_2 : g(t, x^k) \leq 0\}$. Since θ is nonnegative and convex we get

$$\forall t \in T_2^k \cap I_k : \frac{\gamma_k}{|T_2^k|} \theta'(g(t, x^k) \delta_k) (0 - \delta_k g(t, x^k)) \leq \frac{\gamma_k}{|T_2^k|} [\theta(0) - \theta(g(t, x^k) \delta_k)] \leq \frac{\gamma_k}{|T_2^k|} \theta(0),$$

and since θ' is nonnegative it follows, from the definition of σ^k , that

$$\langle g_{x^k}, \sigma^k \rangle \geq \sum_{t \in T_2^k \cap I_k} \frac{\gamma_k}{|T_2^k|} \theta'(g(t, x^k) \delta_k) g(t, x^k) \geq -\frac{\gamma_k}{\delta_k} \theta(0). \quad (51)$$

i-2) Let us prove now that the sequence $\{\sigma^k\}$ is strongly bounded. If not, there will exist a subsequence $\{\sigma^k\}_{k \in K}$ such that $\lim_{k \rightarrow \infty, k \in K} \|\sigma^k\| = \infty$, and we define the measures

$$\hat{\sigma}^k := \sigma^k / \|\sigma^k\|, \quad k \in K.$$

Then recall that the separability of $\mathcal{C}(T_2)$ entails that the ball $B^* := \{\sigma \in M(T_2) : \|\sigma\| \leq 1\}$ is weak* sequentially compact. As a consequence of that, and since the sequence $\{x^k\}_{k \in K}$ is bounded with limit points in S_P (according to Theorem 3.1), there must exist a subsequence $\{\hat{\sigma}^k\}_{k \in K'}$, with $K' \subset K$, such that

$$\lim_{k \rightarrow \infty, k \in K'} x^k = x_\infty \in S_P, \quad w^* - \lim_{k \rightarrow \infty, k \in K'} \hat{\sigma}^k = \hat{\sigma} \in M_+(T_2), \quad \|\hat{\sigma}\| = 1. \quad (52)$$

Now from (49) we obtain

$$\left\| \frac{\nabla F(x^k)}{\|\sigma^k\|} + \langle \nabla_x g_{x^k}, \hat{\sigma}^k \rangle + \frac{2\epsilon_k x^k}{\|\sigma^k\|} \right\| \leq \frac{\sqrt{2}\epsilon_k}{\|\sigma^k\|}, \quad k \in K'. \quad (53)$$

Before taking limits for $k \rightarrow \infty$, $k \in K'$, we write

$$\langle \nabla_x g_{x^k}, \hat{\sigma}^k \rangle = \langle \nabla_x g_{x^k} - \nabla_x g_{x_\infty}, \hat{\sigma}^k \rangle + \langle \nabla_x g_{x_\infty}, \hat{\sigma}^k \rangle. \quad (54)$$

Using the uniform convergence over the compact set T_2 , we get

$$\lim_{k \rightarrow \infty, k \in K'} \|\nabla_x g_{x^k} - \nabla_x g_{x_\infty}\|_\infty = 0. \quad (55)$$

So, since the sequences $\{x^k\}_{k \in K'}$ and $\{\nabla F(x^k)\}_{k \in K'}$ are bounded, from (54) and the weak* convergence of the bounded sequence $\hat{\sigma}^k$ to $\hat{\sigma}$, taking limits in (53) we conclude

$$\langle \nabla_x g_{x_\infty}, \hat{\sigma} \rangle = 0. \quad (56)$$

Let us write now

$$\langle g_{x^k}, \hat{\sigma}^k \rangle = \langle g_{x^k} - g_{x_\infty}, \hat{\sigma}^k \rangle + \langle g_{x_\infty}, \hat{\sigma}^k \rangle. \quad (57)$$

With the same arguments as above, taking limits for $k \rightarrow \infty$, $k \in K'$, we get

$$\lim_{k \rightarrow \infty, k \in K'} \langle g_{x^k}, \hat{\sigma}^k \rangle = \langle g_{x_\infty}, \hat{\sigma} \rangle \leq 0, \quad (58)$$

since $\hat{\sigma} \in M_+(T_2)$ and $g(t, x_\infty) \leq 0 \forall t \in T_2$.

Now, dividing both members of (51) by $\|\sigma^k\|$, we get

$$\langle g_{x^k}, \hat{\sigma}^k \rangle \geq -\frac{\gamma_k}{\delta_k \|\sigma^k\|} \theta(0). \quad (59)$$

Since $\lim_{k \rightarrow \infty} \frac{\gamma_k}{\delta_k \|\sigma^k\|} = 0$, passing to the limit in this inequality, we get with (58)

$$\langle g_{x_\infty}, \hat{\sigma} \rangle = 0. \quad (60)$$

Let us consider now u satisfying Slater's condition. Since $\|\widehat{\sigma}\| = 1$ and $\widehat{\sigma} \in M_+(T_2)$, it follows that $\langle g_u, \widehat{\sigma} \rangle < 0$. Since $g(t, \cdot)$ is convex, we get $h(t) := \langle \nabla_x g(t, x_\infty), u - x_\infty \rangle \leq g(t, u) - g(t, x_\infty)$ and from (56), and (60) it follows

$$0 = \langle h, \widehat{\sigma} \rangle = \langle \langle \nabla_x g_{x_\infty}, \widehat{\sigma} \rangle, u - x_\infty \rangle \leq \langle g_u - g_{x_\infty}, \widehat{\sigma} \rangle < 0,$$

a contradiction.

ii) Since the sequence $\{\sigma^k\}$ is bounded, and applying again that B^* is weak* sequentially compact, there will exist at least a w^* -limit point. Let σ_∞ be an arbitrary w^* -limit point of this sequence. Since $\{x^k\}$ is bounded with limit points in S_P there exists a subsequence $\{\sigma^k\}_{k \in K}$ such that

$$\lim_{k \rightarrow \infty, k \in K} x^k = x_\infty \in S_P, \quad w^* - \lim_{k \rightarrow \infty, k \in K} \sigma^k = \sigma_\infty \in M_+(T_2). \quad (61)$$

Using the same arguments as for (58) we get $\langle g_{x_\infty}, \sigma_\infty \rangle \leq 0$. Now, since either $\theta(0) = 0$ or $\lim_{k \rightarrow \infty} \frac{\gamma_k}{\delta_k} = 0$, and passing to the limit in (51), we get

$$\langle g_{x_\infty}, \sigma_\infty \rangle = 0. \quad (62)$$

Now coming back to (49), passing to the limit and using the same arguments as in part i-2) we get

$$\nabla F(x_\infty) + \langle \nabla_x g_{x_\infty}, \sigma_\infty \rangle = 0.$$

Then applying Theorem 4.1, it follows that $\sigma_\infty \in S_D$. \square

5. Integral-type Algorithm coupled with Penalty and Smoothing Methods RPSALG, as all Remez-type methods, requires to solve nonconvex optimization problems in Step 2. From a computational point of view, actually this is only possible for particular cases, for instance when the functions $f(\cdot, x)$, $g(\cdot, x)$ are polynomial, with low dimensional sets T_1 and T_2 . An alternative strategy can be to consider a global smoothing and penalization via integrals, which convexifies these functions. In this section we suppose that T_i , $i = 1, 2$, is a compact set in some finite-dimensional Euclidean space, with a nonempty interior, and that

$$f(t, x) = h(x) + \langle a(t), x \rangle - b(t), \quad \forall t \in T_1, \quad g(t, x) = \langle a(t), x \rangle - b(t), \quad \forall t \in T_2,$$

where $h : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex, lsc, and \mathcal{C}^1 on Q . For more general cases we refer to the forthcoming Remark 5.2.

For $\delta_k > 0$, $p_k > 0$, $\gamma_k > 0$, and $\theta \in \mathcal{F}$ we set

$$I_k(x) = \frac{1}{p_k} \log \left(\int_{T_1} \exp[\langle a(t), x \rangle - b(t)] p_k dt \right), \quad E_k(x) = \gamma_k \int_{T_2} \frac{\theta(g(t, x) \delta_k)}{\delta_k} dt,$$

where dt is the Lebesgue measure. Then we consider

$$J_k(x) = h(x) + I_k(x), \quad R_k(x) = J_k(x) + E_k(x),$$

and with $\epsilon_k > 0$, we introduce the associated regularized subproblem

$$(P_{irps}^k) \quad \inf \{ R_k(x) + \epsilon_k \|x\|^2 \mid x \in Q \}. \quad (63)$$

Observe that I_k is convex ([12, Lemma 1]) and, obviously, E_k is also convex, so that R_k is convex, lsc, and \mathcal{C}^1 on Q . Consequently, the objective function of (P_{irps}^k) is strongly convex on Q and there exists at least a point x^k satisfying

$$x^k \in Q, \quad h(x^k) + I_k(x^k) + E_k(x^k) + \epsilon_k \|x^k\|^2 \leq h(u) + I_k(u) + E_k(u) + \epsilon_k \|u\|^2 + \epsilon_k \quad \forall u \in Q. \quad (64)$$

REMARK 5.1 *Since h and θ are \mathcal{C}^1 , the objective function of (P_{irps}^k) is also \mathcal{C}^1 , with*

$$\nabla I_k(x) = \frac{\int_{T_1} \exp[\langle a(t), x \rangle - b(t)] p_k a(t) dt}{\int_{T_1} \exp[\langle a(t), x \rangle - b(t)] p_k dt}, \quad \nabla E_k(x) = \gamma_k \int_{T_2} \theta'(g(t, x) \delta_k) a(t) dt. \quad (65)$$

Then it is worthwhile to note as in Remark 3.2 that, when Q is the whole space, any usual gradient method will provide in a finite number of steps such a point by using the implementable stopping rule

$$\|\nabla R_k(x^k) + 2\epsilon_k x^k\| \leq \sqrt{2}\epsilon_k. \quad (66)$$

We suppose now for the rest of this section that $\epsilon_k > 0$, $\forall k$, $\lim_{k \rightarrow \infty} \epsilon_k = 0$, $\lim_{k \rightarrow \infty} p_k = +\infty$, and we introduce the following conditions:

$$(a') \theta \in \mathcal{F}_1, \lim_{k \rightarrow \infty} \frac{\gamma_k}{\delta_k} = 0, \text{ and } \lim_{k \rightarrow \infty} \gamma_k = +\infty.$$

$$(b') \theta \in \mathcal{F}_2, \lim_{k \rightarrow \infty} \frac{\gamma_k}{\delta_k} = 0, \text{ and } \gamma_k > \varepsilon \forall k, \text{ for a certain } \varepsilon > 0.$$

Now we describe our second algorithm as follows:

Integral Penalty Smoothing Algorithm - IPSALG:

Compute, at each Step k , x^k satisfying (64).

From now on, for each $V_i \subset T_i$ we set $\nu(V_i) = \int_{V_i} dt$. Then we have the following lemma:

LEMMA 5.1 *Let $u \in Q$, $\tau(u) := \max\{\langle a(t), u \rangle - b(t) : t \in T_1\}$, and let $\{u^k\}$ be a sequence in Q converging to u . Then*

$$I_k(u) \leq \tau(u) + \frac{\log(\nu(T_1))}{p_k} \quad \text{and} \quad \lim_{k \rightarrow \infty} I_k(u^k) = \tau(u). \quad (67)$$

Proof. Set

$$l(t) := \langle a(t), u \rangle - b(t), \quad l_k(t) = \langle a(t), u^k \rangle - b(t), \text{ and } t_1^* \in \operatorname{argmax}\{l(t) : t \in T_1\}.$$

Then the first inequality in (67) is a direct consequence of the inequality $\exp[l(t)p_k] \leq \exp[\tau(u)p_k] \forall t \in T_1$. Now we claim that for each $\beta > 0$ there exist $\epsilon_1 > 0$ and n_1 such that

$$l_k(t) \geq \tau(u^k) - \beta \quad \forall t \in B(t_1^*; \epsilon_1) := \{t \in T_1 : d(t_1^*, t) \leq \epsilon_1\} \text{ and } \forall k \geq n_1.$$

Otherwise, there would exist $\beta > 0$ such that, for each positive integer r , there exist $t_r \in T_1$ and $k_r \geq r$ verifying $d(t_1^*, t_r) \leq \frac{1}{r}$ and $l_{k_r}(t_r) < \tau(u^{k_r}) - \beta$. The sequence $\{t_r\}$ converges to t_1^* , and since the function τ is continuous, passing to the limit, we get $\tau(u) \leq \tau(u) - \beta$, a contradiction. As a consequence we get, denoting $B_1 := B(t_1^*; \epsilon_1)$,

$$\begin{aligned} I_k(u^k) &\geq \frac{1}{p_k} \log \left(\int_{B_1} \exp[l_k(t)p_k] dt \right) \geq \frac{1}{p_k} \log \left(\int_{B_1} \exp[(\tau(u^k) - \beta)p_k] dt \right) \\ &= (\tau(u^k) - \beta) + \frac{\log(\nu(B_1))}{p_k}. \end{aligned}$$

Passing to the limit, since τ is continuous we obtain $\liminf_{k \rightarrow \infty} I_k(u^k) \geq \tau(u) - \beta$, and then, with $\beta \rightarrow 0^+$, we obtain $\liminf_{k \rightarrow \infty} I_k(u^k) \geq \tau(u)$.

Now using the first inequality in (67) and the continuity of τ we get $\limsup_{k \rightarrow \infty} I_k(u^k) \leq \tau(u)$ so that $\lim_{k \rightarrow \infty} I_k(u^k) = \tau(u)$. \square

THEOREM 5.1 *Suppose that Assumption (A_1) holds and that $(\theta, \{\gamma_k\}, \{\delta_k\})$ satisfies at least one of the conditions (a'), (b'). Then the sequence $\{x^k\}$ built by **IPSALG** is bounded and all its limit points are in S_P .*

Proof. 1) Let $u \in C$ and set

$$\eta_k(u) := \gamma_k \nu(T_2) \frac{\theta(G(u)\delta_k)}{\delta_k} + \frac{\log(\nu(T_1))}{p_k} + \epsilon_k. \quad (68)$$

Since $0 \leq \theta(G(u)\delta_k) \leq \theta(0)$, it follows from (68), and at least one of the two conditions (a') and (b') that

$$\lim_{k \rightarrow \infty} \eta_k(u) = 0. \quad (69)$$

Now since θ is nondecreasing, by definition of $E_k(x^k)$, and from (64) and (67)

$$x^k \in Q, \quad h(x^k) + I_k(x^k) + E_k(x^k) \leq F(u) + \eta_k(u) + \epsilon_k \|u\|^2 \quad \forall u \in C. \quad (70)$$

As a consequence, since $E_k(x^k)$ is nonnegative there exists some $\alpha \in \mathbb{R}$ such that, for k sufficiently large, we have

$$h(x^k) + I_k(x^k) + E_k(x^k) \leq \alpha \text{ and } h(x^k) + I_k(x^k) \leq \alpha. \quad (71)$$

2) Now let us prove that the sequence $\{x^k\}$ is bounded. Suppose the contrary. Then there exists a subsequence $\{x^k\}_{k \in K}$ such that

$$\lim_{k \rightarrow +\infty, k \in K} \|x^k\| = +\infty, \quad \lim_{k \rightarrow +\infty, k \in K} \frac{x^k}{\|x^k\|} = d \neq 0, \quad d \in Q_\infty. \quad (72)$$

Let $t_1^* \in \operatorname{argmax}\{\langle a(t), d \rangle \mid t \in T_1\}$. Then (see Section 2) $F_\infty(d) = h_\infty(d) + \langle a(t_1^*), d \rangle$. Now let $\tau < \langle a(t_1^*), d \rangle$. Then, since $a(\cdot)$ is continuous, for each $r_1 > 0$ with $\tau + 4r_1 \leq \langle a(t_1^*), d \rangle$ there exists $\epsilon_1 > 0$ such that

$$\langle a(t), d \rangle > 2r_1 + \tau \quad \forall t \in B_1 := B(t_1^*; \epsilon_1). \quad (73)$$

Set $u_k(t) := \left\langle a(t), \frac{x^k}{\|x^k\|} \right\rangle - \frac{b(t)}{\|x^k\|}$. Since $a(\cdot), b(\cdot)$ are continuous, the sequence $\{u_k\}_{k \in K}$ converges uniformly on B_1 to $\langle a(\cdot), d \rangle$, and there exists k_0 such that

$$u_k(t) \geq r_1 + \tau, \quad \forall t \in B_1, \text{ and } \forall k \geq k_0, k \in K.$$

Then $\forall k \geq k_0$ we have

$$\begin{aligned} I_k(x^k) &\geq \frac{1}{p_k} \log \left(\int_{B_1} \exp[p_k u_k(t) \|x^k\|] dt \right) \geq \frac{1}{p_k} \log \left(\int_{B_1} \exp[(\tau + r_1) \|x^k\| p_k] dt \right) \\ &= (\tau + r_1) \|x^k\| + \frac{\log(\nu(B_1))}{p_k}, \end{aligned}$$

and this entails

$$\liminf_{k \rightarrow \infty, k \in K} \frac{I_k(x^k)}{\|x^k\|} \geq \tau + r_1.$$

Then, taking $\tau \rightarrow \langle a(t_1^*), d \rangle$, which implies $r_1 \rightarrow 0^+$, we get

$$\liminf_{k \rightarrow \infty, k \in K} \frac{I_k(x^k)}{\|x^k\|} \geq \langle a(t_1^*), d \rangle. \quad (74)$$

Now dividing both members of the second inequality in (71) by $\|x^k\|$ and passing to the limit, we get from inequality (74) and from the definition of h_∞ :

$$\begin{aligned} 0 &= \lim_{k \rightarrow \infty, k \in K} \frac{\alpha}{\|x^k\|} \geq \liminf_{k \rightarrow \infty, k \in K} \left(\frac{h(x^k)}{\|x^k\|} + \frac{I_k(x^k)}{\|x^k\|} \right) \\ &\geq \liminf_{k \rightarrow \infty, k \in K} \frac{h(x^k)}{\|x^k\|} + \liminf_{k \rightarrow \infty, k \in K} \frac{I_k(x^k)}{\|x^k\|} \\ &\geq h_\infty(d) + \langle a(t_1^*), d \rangle = F_\infty(d). \end{aligned} \quad (75)$$

Now we proceed by dividing both members of the first inequality (71) by $\|x^k\|$, passing to the limit and using (75):

$$\begin{aligned} 0 &= \lim_{k \rightarrow \infty, k \in K} \frac{\alpha}{\|x^k\|} \geq \liminf_{k \rightarrow \infty, k \in K} \left(\frac{h(x^k)}{\|x^k\|} + \frac{I_k(x^k)}{\|x^k\|} + \frac{E_k(x^k)}{\|x^k\|} \right) \\ &\geq h_\infty(d) + \langle a(t_1^*), d \rangle + \liminf_{k \rightarrow \infty, k \in K} \frac{E_k(x^k)}{\|x^k\|}. \end{aligned} \quad (76)$$

Now we prove that

$$(g_t)_\infty(d) = \langle a(t), d \rangle \leq 0 \quad \forall t \in T_2, \quad (77)$$

in which case, using (22), relations (72), (75), (77) would imply that assumption (A_1) is not satisfied, and this is a contradiction. So, suppose that (77) does not hold. Then, since $a(\cdot)$ is continuous, there exist $t^* \in T_2$, $r > 0$ and $\epsilon_2 > 0$ such that

$$\langle a(t), d \rangle > 2r \quad \forall t \in B_2 := B(t^*; r) := \{t \in T_2 : d(t^*, t) \leq r\}. \quad (78)$$

Set $u_k(t) := \left\langle a(t), \frac{x^k}{\|x^k\|} \right\rangle - \frac{b(t)}{\|x^k\|}$. Since $a(\cdot), b(\cdot)$ are continuous, the sequence $\{u_k\}_{k \in K}$ converges uniformly on B_2 to $\langle a(\cdot), d \rangle$. Hence, there exists k_1 such that

$$u_k(t) \geq r, \quad \forall t \in B_2, \text{ and } \forall k \geq k_1, k \in K.$$

Since θ is nondecreasing, it follows that

$$\theta(g(t, x^k) \delta_k) \geq \theta(r \|x^k\| \delta_k), \quad \forall t \in B_2, \text{ and } \forall k \geq k_1, k \in K,$$

so that

$$\frac{E_k(x^k)}{\|x^k\|} = \frac{\gamma_k}{\delta_k \|x^k\|} \int_{B_2} \theta(g(t, x^k) \delta_k) dt \geq \nu(B_2) \gamma_k \frac{\theta(r \|x^k\| \delta_k)}{\|x^k\| \delta_k} \quad \forall k \geq k_1, k \in K. \quad (79)$$

Taking limits we get

$$\liminf_{k \rightarrow \infty, k \in K} \frac{E_k(x^k)}{\|x^k\|} \geq \nu(B_2) \liminf_{k \rightarrow \infty, k \in K} \left\{ \gamma_k \frac{\theta(r \|x^k\| \delta_k)}{\|x^k\| \delta_k} \right\}. \quad (80)$$

We shall make the following discussion:

(a') Since

$$\liminf_{k \rightarrow \infty, k \in K} \left\{ \frac{\theta(r \|x^k\| \delta_k)}{\|x^k\| \delta_k} \right\} \geq \theta_\infty(r) = r \theta_\infty(1) > 0$$

and $\lim_{k \rightarrow \infty} \gamma_k = +\infty$, we get $\lim_{k \rightarrow \infty, k \in K} \frac{E_k(x^k)}{\|x^k\|} = +\infty$ and this contradicts (76).

(b') Since $\gamma_k > \varepsilon > 0, \forall k$, and $\theta_\infty(r) = r \theta_\infty(1) = +\infty$ we reach a similar contradiction, and we have to conclude that the sequence $\{x^k\}$ is bounded.

3) Now let $\{x^k\}_{k \in K}$ be a subsequence such that $\lim_{k \rightarrow \infty, k \in K} x^k = x_\infty$ and, in order to finish the proof, let us show that x_∞ is an optimal solution. Since Q is closed it follows that $x_\infty \in Q$. Furthermore, since E_k is nonnegative and h is continuous at x_∞ , passing to the limit in (70), and thanks to (69) and Lemma 5.1, we get

$$F(x_\infty) \leq v(P). \quad (81)$$

Coming back to (71), and since from Lemma 5.1 the sequence $\{h(x^k) + I_k(x^k)\}$ is bounded, then there exists a scalar β such that

$$E_k(x^k) \leq \beta, \quad k \in K. \quad (82)$$

To prove that x_∞ is an optimal solution, since $x_\infty \in Q$, and thanks to (81), we have only to show that $\langle a(t), x_\infty \rangle - b(t) \leq 0 \quad \forall t \in T_2$.

Suppose the contrary. Then, since $a(\cdot), b(\cdot)$ are continuous, there exist $t^* \in T_2$ and $r > 0$ such that

$$\langle a(t), x_\infty \rangle - b(t) > 2r \quad \forall t \in B(t^*; r).$$

Set $u_k(t) := \langle a(t), x^k \rangle - b(t)$. Again, since $a(\cdot), b(\cdot)$ are continuous, the sequence $\{u_k\}_{k \in K}$ converges uniformly on $B(t^*; r)$ to $\langle a(\cdot), x_\infty \rangle - b(\cdot)$. As a consequence of that it follows, for k sufficiently large, that $u_k(t) \geq r \quad \forall t \in B(t^*; r)$. Then since θ is nondecreasing and nonnegative we get from (82)

$$\nu(B_2) \gamma_k \frac{\theta(r \delta_k)}{\delta_k} \leq \beta.$$

Then taking the same arguments given at the end of part 2, and passing to the limit in this inequality, we obtain a contradiction, which finishes the proof. \square

REMARK 5.2 When Q is bounded it is worthwhile to note that the proof of Theorem 5.1 remains valid for functions f_t and g_t not necessarily affine. Indeed the proofs of Lemma 5.1 and parts 1 and 3 of Theorem 5.1 remain valid, by word by word, meanwhile part 2 becomes unnecessary because Q is bounded.

REMARK 5.3 When $T_1 = \emptyset$ and $T_2 \neq \emptyset$ we define $F := h, I_k := 0$, while if $T_2 = \emptyset$ and $T_1 \neq \emptyset$ then $D = \mathcal{R}^n$ and we define $E_k := 0$. Then in both cases IPSALG is well defined and Theorem 5.1 obviously remains valid with a proof which becomes simpler.

Duality results

For the sake of simplicity we suppose here that Q is the whole space, that $\{T_1\}$ is reduced to a single element and that x^k satisfies (66). As pointed out in Remark 5.1, x^k satisfies (64), so that Theorem 5.1 holds.

Let us introduce a linear map $J_0 : \mathcal{C}_+(T_2) \rightarrow M_+(T_2)$ as follows

$$\langle J_0 f, h \rangle := \int_{T_2} h f dt \quad \forall f \in \mathcal{C}_+(T_2), \forall h \in \mathcal{C}(T_2).$$

Now we associate with the sequence $\{x^k\}$ the sequence $\{\sigma^k\}$ of measures given by

$$\sigma^k := \gamma_k J_0(\theta'(g(\cdot, x^k)\delta_k)).$$

Using the same techniques as for **RPSALG** we can obtain the following convergence theorem, whose proof is only sketched here.

THEOREM 5.2 *Assume that (A_1) and (A_2) are satisfied, and suppose that $(\theta, \{\gamma_k\}, \{\delta_k\})$ satisfies at least one of the conditions (a'), (b'). Then, statements (i) and (ii) in Theorem 4.2 also hold in this setting.*

Proof. i-1) If we consider again the set $I_k := \{t \in T_2 : g(t, x^k) \leq 0\}$, we have this time

$$\langle g_{x^k}, \sigma^k \rangle \geq \int_{I_k} \gamma_k \theta'(g(t, x^k)\delta_k) g(t, x^k) dt \geq -\frac{\gamma_k}{\delta_k} \nu(T_2)\theta(0). \quad (83)$$

i-2) To prove that the sequence $\{\sigma^k\}$ is strongly bounded, again we suppose the contrary. Then there exists a subsequence $\{\sigma^k\}_{k \in K}$ such that $\lim_{k \rightarrow \infty, k \in K} \|\sigma^k\| = \infty$, and we define the measures $\hat{\sigma}^k := \sigma^k / \|\sigma^k\|$, $k \in K$. Following the same arguments as in Theorem 4.2, now we conclude from (83) that

$$0 \geq \langle g_{x_\infty}, \hat{\sigma} \rangle = \lim_{k \rightarrow \infty, k \in K} \langle g_{x^k}, \hat{\sigma}^k \rangle \geq \lim_{k \rightarrow \infty, k \in K} -\frac{\gamma_k}{\delta_k \|\sigma^k\|} \nu(T_2)\theta(0) = 0.$$

In other words, $\langle g_{x_\infty}, \hat{\sigma} \rangle = 0$, which yields a contradiction with Slater's condition. The rest of the proof is as in Theorem 4.2. \square

To put in perspective the results obtained for IPSALG with respect to related works we end this section with two comments.

Comment 2

IPSALG is a family of methods concerning three types of problems. The first type corresponds to those problems where $T_1 = \emptyset$, $T_2 \neq \emptyset$ (here $F = h$ is C^1 and $I_k = 0$); the second type concerns problems where $T_1 \neq \emptyset$, $T_2 = \emptyset$ (now $C = Q$ and $E_k = 0$); and the third class is the most general with both T_1 and T_2 nonempty.

The references [2], [16], and [30] deal with problems of first type where F is C^1 and $\epsilon_k = 0 \forall k$. In these three papers, as we shall see below, the conditions which are needed for primal convergence are stronger than the unique condition (A_1) required for IPSALG.

In [2] $\theta = \theta_4$, while in [16] $\theta = \theta_6$, so that in both cases $\theta \in \mathcal{F}_2$. Furthermore in both cases $\gamma_k = 1$ and $\lim_{k \rightarrow \infty} \delta_k = +\infty$, so that condition (b') holds and they coincide with IPSALG when $\epsilon_k = 0 \forall k$. In [2] Q is supposed to be compact. In [16] it is assumed that (A_1) and Slater condition hold as well as two other technical conditions. In both cases, the duality results also require stronger assumptions than IPSALG. The method proposed in [30], appears without conditions on $\gamma_k > 0, \delta_k > 0$ as a particular case of IPSALG with $\epsilon_k = 0 \forall k$ and $\theta = \theta_3 \in \mathcal{F}_1$. However these two algorithms are different since the parameters are chosen differently. Indeed in [30] $\delta_k \rightarrow \infty$ and for δ_k fixed γ_k is chosen such that x^k is in addition feasible, while for IPSALG they must satisfy condition (a'). Moreover, in [30] Q is supposed to be compact, the Slater condition is also assumed and no duality result is provided.

In [12] the problem is of second type, a pure min-max problem where $T_2 = \emptyset$ and the proposed algorithm coincides with IPSALG when $\epsilon_k = 0 \forall k$ and $\lim_{k \rightarrow \infty} p_k = \infty$. In [12] Q is supposed to be compact, while for IPSALG we only require (A_1) .

Comment 3

Obviously if we want to use IPSALG, T_i must have low dimension. Then in this case, a natural question raised by the study of IPSALG will be to ask if it has some advantage over RAPSALG, but unfortunately we cannot give a theoretical response to such a question. In fact it would be also interesting to compare problems of the same nature as for example, RAPSALG with Remez cutting plane methods, or IPSALG with integral barrier methods. To gain better understanding of their numerical behavior, future works on implementing and testing these methods should be conducted.

Acknowledgments. M.A. Goberna and M.A. López are supported by DGES of Spain and FEDER of EU, Grant MTM2005-08572-C03-01

References

- [1] F. Alvarez, *Absolute minimizer and exponential penalty*, Ph. D. Thesis, Montpellier 2, France, 1998.
- [2] A. Auslender, *Méthodes et théorèmes de dualité*, Rev. Française Informat. Recherche Opérationnelle **4** (1970), 9-45.
- [3] A. Auslender, R. Cominetti and M. Haddou, *Asymptotic analysis for penalty and barrier methods in convex and linear programming*, Math. Oper. Res. **22** (1997), 43-62.
- [4] A. Auslender and M. Teboulle, *Asymptotic cones and functions in optimization and variational inequalities*, Springer-Verlag, NY, 2003.
- [5] A. Auslender, *Penalty and barrier methods: a unified framework*, SIAM J. Optim. **10** (1999), 653-671.
- [6] A. Ben-Tal and M. Teboulle, *A smoothing technique for nondifferentiable optimization problems*, Lecture Notes in Math. **1405** (1989), 1-11.
- [7] D. P. Bertsekas, *Approximation procedures based on the method of multipliers*, J. Optimization Theory Appl. **23** (1977), 487-510.
- [8] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, Springer-Verlag, NY, 2000.
- [9] C. Chen and O. L. Mangasarian, *A class of smoothing functions for nonlinear and mixed complementarity problems*, Comput. Optim. Appl. **5** (1996), 97-138.
- [10] E. Cheney and A. A. Goldstein, *Newton's method for convex programming and Tchebycheff approximation*. Numer. Math. **1** (1959), 253-268.
- [11] J. Elzinga and T. G. Moore, *A central cutting plane algorithm for the convex programming problem*, Math. Programming **8** (1975), 34-145.
- [12] S. C. Fang, and S. Y. Wu, *Solving min-max problems and linear semi-infinite programs*, Comput. Math. Appl. **32** (1996), 87-93.
- [13] C. Gigola, and S. Gomez, *A regularization method for solving finite convex min-max problems*, SIAM J. Numer. Anal. **27** (1990), 621-1634.
- [14] J. E. Kelley, *The cutting-plane method for solving convex programs*, SIAM J. Control Optim. **8** (1960), 703-712.
- [15] A. A. Kaplan and R. Tichatschke, *Adaptative methods of solving ill-posed semi-infinite convex optimization problems*, Soviet Math. Dokl. **45** (1992), 119-123.
- [16] C. J. Lin, S. C. Fang, and S. Y. Wu, *An unconstrained convex programming approach to linear semi-infinite programming*, SIAM J. Optim. **8** (1998), 443-456.
- [17] B. Martinet, *Algorithmes pour la resolution des problèmes d'optimisation et de minimax*, These Universite Scientifique et Medicale de Grenoble, France, 1972.
- [18] Y. Nesterov, *Smooth minimization of nonsmooth functions*, Math. Programming **103A** (2005), 127-152.
- [19] E. Polak and J. O. Royset, *Algorithms for finite and semi-infinite min-max-min problems using adaptative smoothing techniques*, J. Optim. Theory Appl. **119** (2003), 421-457.
- [20] E. Polak, J. O. Royset, and R.S. Womersley, *Algorithms with adaptative smoothing for finite and semi-infinite min-max problems*, J. Optim. Theory Appl. **119** (2003), 459-484.
- [21] E. Polak, J. E. Higgins, and D. Q. Mayne, *A barrier function method for minimax problems*, Math. Programming **54A** (1992), 155-176.
- [22] R. Reemtsen and S. Görner, *Numerical methods for semi-infinite programming: a survey*, Nonconvex Optim. Appl. **25** (1998), 195-275.
- [23] E. Remez, *Sur la détermination des polynômes d'approximation de degré donne*, Commun. Soc. Math. Kharkoff et Inst. Sci. Math. et Mecan. **10** (1934), 41-63.
- [24] R. T. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, NJ, 1970.
- [25] U. Schattler, *An interior-point method for semi-infinite programming problems*, Annals Oper. Res. **62** (1996), 277-301.

- [26] R. L. Sheu and J. Y. Lin, *Solving continuous min-max problems by an iterative entropic regularization method*, J. Optim. Theory Appl. **121** (2004), 597–612.
- [27] R. L. Sheu and S. Y. Wu, *Combined entropic regularization and path-following method for solving finite convex min-max problems subject to infinitely many linear constraints*, J. Optim. Theory Appl. **101** (1999), 167–190.
- [28] M. Teboulle, *Nonlinear Perturbation for linear Semi-Infinite Optimization Problems*, Proc. 29th IEEE Conference on Decision and Control **4** (1990), 2477-2478.
- [29] K. L. Teo and C. J. Goh, *A simple computational procedure for optimization problems with functional inequality constraints*, IEEE Trans. Automat. Contr. **32** (1987), 940-941.
- [30] K. L. Teo, V. Rehbock and L. S. Jennings, *A new computational algorithm for functional inequality constrained optimization problems*, Automatica J. IFAC **29** (1993) 789–792.
- [31] A. F. Veinott, *The supporting hyperplane method for unimodal programming*, Operations Res. **15** (1967), 47–152.