Penalty and Smoothing Methods for Convex Semi-Infinite Programming

Alfred Auslender
Université de Lyon, CNRS, UMR 5208 Institut Camille Jordan, 69622 Villeurbanne, Cedex, France, and Department of Economics, Ecole Polytechnique, F-91128 Palaiseau, Cedex, France, auslender.alfred@gmail.com

Miguel A. Goberna, Marco A. López
Department of Statistics and Operations Research, University of Alicante, 03080 Alicante, Spain
{mgoberna@ua.es, marco.antonio@ua.es}

In this paper we consider min-max convex semi-infinite programming. 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
ORMS subject classification: Primary: programming; secondary: nonlinear

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 $f : T_1 \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $g : T_2 \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ 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

$$(P) \quad \inf \{ F(x) \mid x \in C \},$$

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 problems, 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 $T_1^0$ and $T_2^0$ 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 $x^k \in \arg \min \{ F^k(x) \mid x \in C^k \}$.

Step 2. Compute

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

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 solving nonconvex optimization problems in Step 2, which is certainly the main difficulty in the general case. Indeed, from a computational point of view, 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, $(\bar{P}^k)$, of the subproblem

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

in the sense that $(\bar{P}^k)$ 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 $T_k^i$, $|T_k^i|$, grows beyond a certain limit, it is well known that slow convergence arises and one way to overcome this drawback is to control $|T_k^i|$ by some constraint dropping schemes. The reader is referred to §§3.1 and 3.2 of the survey of Reemtsen and Görner [22] for a review of the extensive literature on this particular subject.

Concerning CSIP, numerous known methods consist of solving an approximating convex problem ($\tilde{P}^k$). Supposing that $F$ is $\mathcal{C}^1$ (as is generally the case in ordinary CSIP), we can use cutting-plane methods of Cheney and Goldstein [10], Kelley [15], Veinott [31], or Elzinga and Moore [11], and their variants (see, e.g., Reemtsen and Görner [22] for more references). Applied to LSIP, especially Cheney and Goldstein [10] and Kelley [15] 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, constraint dropping rules are again given under some conditions as strict convexity on $F$ for Cheney and Goldstein [10] and Kelley [15]. We again refer the reader to §4 of Reemtsen and Görner [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 $D^k$, $\delta_{D^k}$ (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 Gigola and Gomez [13] and Polak et al. [20]), but for the sake of simplicity we consider here only the most important and widely used in different fields in the literature. It is based on the smoothing of $\max\{\lambda_i : i = 1, \ldots, m\}$ by the function $\log(\sum_{i=1}^m \exp(\lambda_i p))/p$, with $p > 0$. More precisely, this smoothing gives

$$\tilde{F}^k(x) := \frac{\log(\sum_{t \in T^k_2} \exp(f(t, x)p))}{p}, \quad \text{with } p = [\log |T_k^i|]^2. \quad (2)$$

This type of smoothing has been proposed by many authors for solving convex finite min-max problems, in particular by Bertsekas [7], Ben-Tal and Teboulle [6], Alvarez [1], and Nesterov [18]. This smoothing approach has also been proposed by Polak et al. [21], by Sheu and Wu [27] for finite min-max problems subject to infinitely many linear constraints and, more recently, by Sheu and Lin [26] for continuous min-max problems, motivated by the global approach of Fang and Wu [12] using an integral analog. We must also smooth the function $\delta_{D^k}$ and to do that we consider the smoothing approach by penalty and barrier functions introduced, for ordinary convex programs, by Auslender et al. [5]. 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 Auslender et al. [5] (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

$$\tilde{G}^k(x) := \frac{\gamma_k}{|T_k^2|} \frac{\sum_{t \in T^k_2} \theta(g(t, x)\delta_k)}{\delta_k}, \quad (3)$$

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

To summarize, we propose in §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 $T_{k+1} = T_k \cup \{x^{k+1}\}$. 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 §4 we associate with the sequence $\{x^k\}$, generated in §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 and Lin [26] proposed the so-called entropic smoothing method for the min-max program where $T_2 = \emptyset$. Concerning ordinary CSIP ($|T_k^i| = 1$) problems, to the best of our knowledge, Remez-type algorithms coupled with penalty methods have only been introduced by Martinet [17]. Comparisons with these two works are established in Comment 1, §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.,
§5.2 of Reemtsen and Görner [22] and references therein). In another context they are coupled with adaptive grid methods (see, for example, Kaplan and Tichatschke [14], Polak and 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 and Goh [29], Teo et al. [30], Lin et al. [16], Schattler [25], Polak et al. [20], Fang and Wu [12]) and has the advantage of avoiding 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 §5 we propose an algorithm for solving
\[ (P^k_{pi}) \quad \inf \{ \bar{F}^k(x) + \bar{G}^k(x) \mid x \in Q \}, \]
where
\[ \bar{F}^k(x) = \frac{1}{p_k} \log \left( \int_{T_k} \exp(f(t, x)p_k) \, dt \right), \quad \bar{G}^k(x) = \gamma_k \int_{T_k} \frac{\theta(g(t, x)\delta_k)}{\delta_k} \, dt. \]

In this formula the parameters \( \gamma_k, \delta_k, \) and \( p_k \) will be adjusted for obtaining convergence. In fact, and also in §3, we regularize the objective function by adding a term \( \epsilon_k \|x\|^2 \) with \( \epsilon_k > 0 \) and we compute an \( \epsilon_k \)-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 Auslender [2], Fang and Wu [12], Lin et al. [16], and Teo et al. [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 Auslender and Teboulle [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, §5, our assumptions are weaker than those used in Auslender [2], Fang and Wu [12], Lin et al. [16], and Teo et al. [30]. Finally, because 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 \( \text{cl} \, Q \), int \( Q \), conv \( Q \), and cone \( Q \) the closure, the interior, the convex hull, and the conical convex hull of \( Q \), respectively. We associate with \( f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) its domain \( \text{dom} \, f := \{ x \mid f(x) < +\infty \} \) and its epigraph \( \text{epi} \, f := \{(x, r) \mid f(x) \leq r \} \).

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

The asymptotic cone of a set \( Q \subset \mathbb{R}^n \) is defined to be
\[ Q_\infty = \left\{ d \mid \exists \lambda_k \to +\infty, \ x_k \in Q \text{ with } d = \lim_{k \to +\infty} \frac{x_k}{\lambda_k} \right\}. \]

When \( Q \) is convex and closed, it coincides with its recession cone
\[ 0^+(Q) := \{ d \mid x + \lambda d \in Q \text{ for all } \lambda > 0, \forall x \in Q \}. \]

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

As a straightforward consequence, we get (cf. Auslender and Teboulle [4, Theorem 2.5.1])
\[ f_\infty(d) = \inf \left\{ \liminf_{k \to +\infty} \frac{f(\lambda_k x^k)}{\lambda_k} \mid \lambda_k \to +\infty, \ x^k \to d \right\}, \]
where \( \{\lambda_k\} \subset \mathbb{R} \) and \( \{x^k\} \subset \mathbb{R}^n \). Note that \( f_\infty \) is positively homogeneous; that is,
\[ f_\infty(\lambda d) = \lambda f_\infty(d) \quad \forall d, \ \forall \lambda > 0. \]
Remark 2.1. Equation (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
\[
\lim_{k \to \infty} \|x^k\| = +\infty, \quad \lim_{k \to \infty} \frac{x^k}{\|x^k\|} = d,
\]
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
\[
f(x^k) = f\left(\frac{\|x^k\|}{\|x^k\|} \frac{x^k}{\|x^k\|}\right) \geq \alpha \|x^k\|.
\]

When \( f \) is a proper lsc convex function its asymptotic function is also a proper lsc convex function that coincides with the recession function
\[
0^+ f(d) = \lim_{\lambda \to +\infty} \frac{f(x + \lambda d) - f(x)}{\lambda} \quad \forall x \in \text{dom } f,
\]
which implies that
\[
f_\infty(d) = \lim_{\lambda \to +\infty} \frac{f(\lambda d)}{\lambda} \quad \forall d \in \text{dom } f.
\]
Furthermore,
\[
(\delta_Q)_\infty = \delta_{Q_\infty}.
\]
If \( f, g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) are proper lsc convex functions, and \( \text{dom}(f + g) \neq \emptyset \), then
\[
(f + g)_\infty(d) = f_\infty(d) + g_\infty(d).
\]
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
\[
f_\infty = \sup_{i \in I} f_i_\infty.
\]

When \( f \) is a proper lsc convex function, a useful consequence of (6) and (9) is the equation
\[
\{x : f(x) \leq \lambda\}_\infty = \{d : f_\infty(d) \leq 0\}_\infty,
\]
for any \( \lambda \) 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 Auslender and Teboulle [4, Chapter 3].

Proposition 2.1. Let \( Q \) be a closed convex set in \( \mathbb{R}^n \) and let \( f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) 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 \( \lambda \), \( \{x \in Q : f(x) \leq \lambda\} \) is compact when nonempty. This is equivalent to
\[
\lim_{\|x\| \to \infty, x \in Q} f(x) = +\infty.
\]

In our analysis, the composite function is of particular interest. More precisely, we consider the composition between a penalty function \( \theta \in \mathcal{F} \) and a convex function \( f \), where
\[
\mathcal{F} = \left\{ \theta : \mathbb{R} \to \mathbb{R}^+_\infty \text{ convex, nondecreasing, nonconstant, } \mathcal{C}^1, \text{ and such that } \lim_{u \to +\infty} \theta(u) = 0 \right\}.
\]
Because \( \theta \in \mathcal{F} \) takes nonnegative values and it is nondecreasing, we have \( \theta_\infty(-1) = 0 \). Then, because it is nonconstant, \( \theta_\infty(1) > 0 \). The following result was proved in Auslender et al. [5] 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 $\mathcal{F}$:
\[
\mathcal{F}_1 = \{ \theta \in \mathcal{F} : \theta_{\infty}(1) < +\infty \} \quad \text{and} \quad \mathcal{F}_2 = \{ \theta \in \mathcal{F} : \theta_{\infty}(1) = +\infty \}.
\]
(16)

Obviously the function $\theta(u) = u^+ := \max\{u, 0\}$ which has been used in the literature is not $C^1$, but satisfies all the other properties required for $\mathcal{F}_1$. However, this function, for which our convergence analysis holds, is not of interest for our purpose because 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 $\theta(u) = \int_{-\infty}^{\infty} F(t) dt$ is well defined. Then we have (Chen and Mangasarian [9, Proposition 2.2]) that $\theta$ is a strictly convex $C^1$ function from $\mathbb{R}$ to $\mathbb{R}$, strictly increasing, with
\[
0 < \theta'(u) < 1, \quad -M_2 \leq \theta(u) - u^+ \leq M_1 \quad \forall \ u \in \mathbb{R},
\]
(17)
where $M_i := \int_{-\infty}^{0} |s| p(s) ds$ and $M_i := \int_{-\infty}^{\infty} s p(s) ds$, provided that $M_i < +\infty$, $i = 1, 2$. From these inequalities and the definition of $\theta$, it follows that
\[
\left| \frac{\theta(\lambda u)}{\lambda} - u^+ \right| \leq \max\{M_1, M_2\} \forall \lambda > 0, \forall \ u \in \mathbb{R}, \quad \theta_{\infty}(1) = 1, \quad \lim_{u \to +\infty} \theta(u) = 0,
\]
(18)
so that $\theta \in \mathcal{F}_1$. Specific cases of interest are
\[
\theta_1(u) = \log(1 + \exp(u)), \quad \text{with } p_1(s) = \frac{\exp(-s)}{(1 + \exp(-s))^2},
\]
\[
\theta_2(u) = 2^{-1}(u + \sqrt{u^2 + 4}), \quad \text{with } p_2(s) = \frac{2}{(s^2 + 4)^{1/2}},
\]
and
\[
\theta_3(u) = \begin{cases} 
0, & u \leq -1, \\
\frac{1}{4}(u + 1)^2, & -1 < u < 1, \\
u, & u \geq 1,
\end{cases}
\]
\[
\text{with } \text{p}_3(s) = \begin{cases} 
\frac{1}{2}, & -1 \leq s \leq 1, \\
0, & \text{otherwise}.
\end{cases}
\]

Finally, as well known penalty functions which belong to $\mathcal{F}_2$ we have the classical penalty functions and the exponential function:
\[
\theta_4(u) = \frac{1}{2}(u^+)^2, \quad \theta_5(u) = (u^+)^3, \quad \text{and} \quad \theta_6(u) = \exp(u).
\]

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
\[
T_i^{k+1} = T_i \cup \{t_i^{k+1}\}, \quad i = 1, 2,
\]
(19)
where $t_i^{k+1} \in T_i$ and $t_2^{k+1} \in T_2$ solve approximately the auxiliary problems in Step 2, i.e.,
\[
\begin{cases} 
f(t_i^{k+1}, x_i) \geq \max\{f(t, x_i) \mid t \in T_i\} - \mu_k, \\
g(t_2^{k+1}, x_i) \geq \max\{g(t, x_i) \mid t \in T_2\} - \mu_k,
\end{cases}
\]
(20)
with
\[
\mu_k \geq 0 \quad \forall \ k \quad \text{and} \quad \lim_{k \to +\infty} \mu_k = 0.
\]
(21)
From now on in this section we consider the following assumption.

**Assumption (A₁).** $F$ is level bounded on $C$. Sometimes, we shall also assume as follows.

**Assumption (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 the existence of starting sets for the first algorithm of Remez with nice properties is a consequence of Assumption (A₁). It was proved in Reemtsen and Görner [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 $T^0_1 \subset T_1$ and $T^0_2 \subset T_2$ such that $F^0$ is level bounded on $C^0$.

**Proof.** Because $F^*_\infty = \sup_{t \in T_1^0} (f_i^*_\infty)$ and $G^*_\infty = \sup_{t \in T_2^0} (g_j^*_\infty)$, by (A₁) and Proposition 2.1,

$$\{ d : (f_i^*_\infty(d) \leq 0 \forall t \in T_1^0 ; (g_j^*_\infty(d) \leq 0 \forall t \in T_2^0 ; \delta_{Q_\infty}(d) \leq 0 \} = \{0\}.$$

(23)

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

$$\{ d : (f_i^*_\infty(d) \leq 0 \forall t \in T^0_1 ; (g_j^*_\infty(d) \leq 0 \forall t \in T^0_2 ; \delta_{Q_\infty}(d) \leq 0 \} = \{0\}.$$

(24)

Thus $F^0$ is level bounded on $C^0$, again by Proposition 2.1. Finally, if $T^0_1 = \emptyset$, replacing it with $T^0_1 = \{ t_1 \}$ for an arbitrary $t_1 \in T_1$, $i = 1, 2$, we get the aimed conclusion. □

**Remark 3.1.** There are some particular cases where the sets $T^0_i$ are easily obtainable:

(a) If $Q$ is bounded, we can take $T^0_1 = \{ t_1 \}$ and $T^0_2 = \{ t_2 \}$ for any $t \in T_1, i = 1, 2$. (b) If for some $t_1^0 \in T_1, f(t, x) \in T_2^0 = \{ t_2 \}$ level bounded on $Q$, we can take $T^0_1 = \{ t_1^0 \}$ and $T^0_2 = \{ t_2 \}$, with $t_2$ arbitrary in $T_2$ (resp., $T^0_2 = \{ t_2 \}$ and $T^0_1 = \{ t_1 \}$, with $t_1$ arbitrary in $T_1$).

(c) In LSIP, $Q = \mathbb{R}_n^+$, $f(t, x) = \langle a(t), x \rangle - b(t) \forall t \in T_1$, and $g(t, x) = \langle a(t), x \rangle - b(t) \forall t \in T_2$. In that case, if $T_1 = \text{cl int} T_1 \subset \mathbb{R}^n$ and $i = 1, 2$ (as it happens in practice) taking a sequence of real numbers $\beta_+, \text{and} \ 0$, then dist$\{ T \cap \beta, Z^n, T \} \to 0$, $i = 1, 2$. Because $Q$ becomes $d : \langle a, d \rangle \leq 0, t \in T^0_1 \cup T^0_2 \Rightarrow 0$ and, in int conv$\{a(t), t \in T^0_1 \cup T^0_2 \}$, we can take the regular grids $T^0_1 = T_1 \cap \beta, Z^n, i = 1, 2$, for sufficiently large $r$.

(d) In ordinary CSIP (with $T_1 = \{ t_1 \}$), if $T_2 = \text{cl int} T_2 \subset \mathbb{R}^n$ and $\beta_+, \text{and} \ 0$, because dist$\{ T \cap \beta, Z^n, T \} \to 0$, it is possible to take the regular grid $T_2 = \{ t_2 \}$ for sufficiently large $r$ by the argument of Reemtsen and Görner [22, Lemma 2.4].

Denote $r_i = |T^0_i|$, the cardinality of $T^0_i$. Then $r_i := |T^i| \leq r_i + k, i = 1, 2$. As it was said in the introduction, we can use the function

$$F^k_p(x) := \frac{\log(\sum_{t_i \in T^i} \exp(f(t, x)p))}{p},$$

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

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

If $T_1$ is reduced to a single point, it is worthwhile to note that $F^k(x) = F^k(x) = F(x)$ and that in Step 2 the computation of $r_{i+1}$ is unnecessary. From now on for each set we put $p_k = [\log(r_i + k)]^2$, and use the approximating function

$$F^k := F^k_{p_k},$$

so that

$$0 \leq F^k(x) - F^k(x) \leq \frac{1}{\log(r_i + k)}, \quad \forall x \in \mathbb{R}^n.$$
Now let \( \{ \epsilon_k \} \) be a sequence of real numbers such that

\[
\epsilon_k > 0 \quad \text{and} \quad \lim_{k \to \infty} \epsilon_k = 0.
\]

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

\[
\tilde{G}^k(x) := \frac{\gamma_k}{T^2_k} \sum_{t \in \mathcal{I}} \frac{\theta(g(t, x) \delta_k)}{\delta_k}, \quad \tilde{H}^k(x) := \tilde{F}^k(x) + \tilde{G}^k(x),
\]

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

\[
(\tilde{P}^k_{\epsilon_k}) \quad \inf \{ \tilde{H}^k(x) + \epsilon_k \| x \|_2^2 \mid x \in \mathcal{Q} \}.
\]

This subproblem will be solved in Step 1 of the forthcoming algorithm within an error \( \epsilon_k \).

**Remark 3.2.** The objective function \( H^k_r(\cdot) := \tilde{H}(\cdot) + \epsilon_k \| \cdot \|_2^2 \) is strongly convex, so that \( \arg \min \{ \tilde{H}(x) + \epsilon_k \| x \|_2^2 \mid x \in \mathcal{Q} \} \) is a single point \( y^k \). As a consequence, there exists at least a point \( x^k \) satisfying

\[
x^k \in \mathcal{Q}, \quad \tilde{H}^k(x^k) + \epsilon_k \| x^k \|_2^2 \leq \tilde{H}^k(x) + \epsilon_k \| x \|_2^2 + \epsilon_k \quad \forall x \in \mathcal{Q}.
\]

When \( \tilde{H}^k \) is \( \mathcal{C}^1 \) and \( \mathcal{Q} = \mathbb{R}^n \), 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

\[
\| \nabla \tilde{H}^k(x^k) + 2 \epsilon_k x^k \|_2 \leq \sqrt{2} \epsilon_k.
\]

Indeed, writing the strong convex inequality \( \langle \nabla H^k_r(x^k) - \nabla H^k_r(y^k), x^k - y^k \rangle \geq 2 \epsilon_k \| x^k - y^k \|_2^2 \), it follows from (29) that \( \| x^k - y^k \|_2 \leq 1/\sqrt{2} \) because \( H^k_r(x^k) \geq H^k_r(y^k) + \langle \nabla H^k_r(x^k), y^k - x^k \rangle \), using (29) again, 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 \( \epsilon_k \| x \|_2^2 \).

Now we can describe our basic algorithm:

**The Remez penalty smoothing algorithm—RPSALG**

Initialization: Set \( k = 0 \) and start with \( T^0_1 \) and \( T^0_2 \) defined in Lemma 3.1.

**Step 1.** Compute \( x^k \) satisfying (28).

**Step 2.** Compute \( t_{k+1}^1 \) and \( t_{k+1}^2 \) satisfying (20) with (21).

**Step 3.** Set: \( T_{k+1}^1 = T_k^1 \cup \{ t_{k+1}^1 \}, \quad i = 1, 2 \).

Set \( k \leftarrow k + 1 \); go to Step 1.

Each triple \( (\theta, \{ \gamma_k \}, \{ \delta_k \}) \) determines a different instance of RPSALG. 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) \( \theta \in \mathcal{T}_1, \lim_{k \to \infty} (\gamma_k / \delta_k) = 0, \) and \( \lim_{k \to \infty} (\gamma_k / m_k) = +\infty \).

(b) \( \theta \in \mathcal{T}_2, \lim_{k \to \infty} (\gamma_k / \delta_k) = 0, \) and \( \gamma_k / m_k > 0 \) \( \forall k \), for a certain \( \epsilon > 0 \).

(c) \( \theta \in \mathcal{T}_3, \lim_{k \to \infty} \delta_k = +\infty, \) \( \gamma_k / m_k > 0 \) \( \forall k \) for a certain \( \epsilon > 0, \) \( \{ \gamma_k / \delta_k \} \) is bounded, and either \( \theta(0) = 0 \) or \( (A_3) \) holds.

**Remark 3.3.** The natural choice is \( m_k = r_1 + k \). 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 \( \lim_{k \to \infty} \delta_k = +\infty \).

**Theorem 3.1.** Assume that \( (A_3) \) 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 \( u \in \mathcal{Q} \) such that \( G(u) < 0 \) if \( (A_3) \) holds and let \( u \in \mathcal{C} \) otherwise.

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

\[
F^l(x^k) + \frac{\gamma_k}{T^2_k} \sum_{t \in \mathcal{I}} \frac{\theta(g(t, x^k) \delta_k)}{\delta_k} \leq F(u) + \frac{\gamma_k}{T^2_k} \sum_{t \in \mathcal{I}} \frac{\theta(g(t, u) \delta_k)}{\delta_k} + \nu_k(u),
\]

with

\[
\nu_k(u) := \frac{1}{\log(r_1 + k)} + \epsilon_k \| u \|_2^2 + \epsilon_k, \quad \lim_{k \to \infty} \nu_k(u) = 0.
\]
Because \( \theta \) is nondecreasing and nonnegative it follows that
\[
F^i(x^i) + \frac{\gamma_k}{|T^k_1|} \sum_{t \in T^k_1} \frac{\theta(g(t, x^i) \delta_k)}{\delta_k} \leq F(u) + \frac{\gamma_k}{|T^k_1|} \sum_{t \in T^k_1} \frac{\theta(G(u) \delta_k)}{\delta_k} + \nu_k(u)
\]
\[
\leq F(u) + \frac{\gamma_k}{\delta_k} \theta(G(u) \delta_k) + \nu_k(u).
\]
\[(31)\]

(2) Let us prove now that the sequence \( \{x^i\} \) is bounded. Suppose the contrary. Then there exists a subsequence \( \{x^i\}_{k \in K} \) such that
\[
\lim_{k \to \infty} \|x^i\| = +\infty, \quad \lim_{k \to \infty} \|x^i\| = d \neq 0, \quad d \in Q_\infty.
\]
Let \( l \) be arbitrary. Let \( \alpha^i_1 < g(t, \cdot)_\infty(d) \quad \forall t \in T^k_1 \) and \( \beta^l < (F^i)_\infty(d) \). Then, as pointed out in Remark 2.1, there exists \( k_1 \) such that
\[
F^i(x^i) \geq \beta^l \|x^i\|, \quad g(t, x^i) \geq \alpha^i_1 \|x^i\| \quad \forall t \in T^k_1, \quad \forall k \in K \quad \text{such that} \quad k \geq \max\{l, k_1\}.
\]
Because \( \theta \) is nonnegative and nondecreasing, dividing both members of inequality (31) by \( \|x^i\| \) we deduce
\[
\beta^l + \frac{\gamma_k}{|T^k_1|} \frac{\theta(\alpha^i_1 \|x^i\| \delta_k)}{\|x^i\| \delta_k} \leq \frac{F(u)}{\|x^i\|} + \frac{\gamma_k \theta(G(u) \delta_k)}{\|x^i\| \delta_k} + \frac{\nu_k(u)}{\|x^i\|} \quad \forall t \in T^k_1.
\]
\[(33)\]

Because \( \theta \) is nondecreasing and \( \delta_k G(u) \leq 0 \), then \( \theta(G(u) \delta_k) \leq \theta(0) \quad \forall k \) so that the right-hand side of (33) converges to zero as \( k \to \infty \). As a consequence, \( \forall \tau^l > 0 \) we have for \( k \) large enough,
\[
\beta^l + \frac{\gamma_k}{|T^k_1|} \frac{\theta(\alpha^i_1 \|x^i\| \delta_k)}{\|x^i\| \delta_k} \leq \tau^l \quad \forall t \in T^k_1.
\]
\[(34)\]

Let us show now that
\[
\alpha^i_1 \leq 0 \quad \forall t \in T^k_1.
\]
\[(35)\]

Suppose the contrary; i.e., there exists some \( t \in T^k_1 \) with \( \alpha^i_1 > 0 \). Consider
\[
\lim_{k \to \infty} \frac{\gamma_k}{|T^k_1|} \frac{\theta(\alpha^i_1 \|x^i\| \delta_k)}{\|x^i\| \delta_k} = \lim_{k \to \infty} \left( \frac{\alpha^i_1 \gamma_k}{|T^k_1|} \frac{\theta(\alpha^i_1 \|x^i\| \delta_k)}{\|x^i\| \delta_k} \right) \geq \lim_{k \to \infty} \left( \frac{\alpha^i_1 \gamma_k}{m_k} \right) \left( \frac{\theta(\alpha^i_1 \|x^i\| \delta_k)}{\|x^i\| \delta_k} \right),
\]
where \( \lim_{k \to \infty} (\theta(\alpha^i_1 \|x^i\| \delta_k))/\alpha^i_1 \|x^i\| \delta_k = \theta_\infty(1) \). 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, because \( \theta \) is nonnegative we deduce from (34) that \( \beta^l \leq \tau^l \) for all positive \( \tau^l \); i.e., \( \beta^l \leq 0 \). From (35), letting \( \beta^l \to (F^i)_\infty(d), \alpha^i_1 \to g(t, \cdot)_\infty(d) \quad \forall t \in T^k_1 \), it follows that
\[
(F^i)_\infty(d) \leq 0, \quad g(t, \cdot)_\infty(d) \leq 0 \quad \forall t \in T^k_1.
\]
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 \( x_\infty \) be a limit point of the sequence \( \{x^i\} \). Because \( Q \) is closed, \( x_\infty \in Q \). Furthermore, because \( T_1 \) and \( T_2 \) are compact, there exist \( t_i \in T_1, i = 1, 2 \), and subsequences \( \{x^{i+1}\}_{k \in K}, \{t^{i+1}\}_{k \in K}, i = 1, 2 \), such that
\[
\lim_{k \to \infty} x^k = x_\infty, \quad \lim_{k \to \infty} t^{i+1} = t_i, \quad i = 1, 2.
\]
\[(37)\]

Let \( l \) be arbitrary. Let \( \tilde{\alpha}^i_1 < g(t, x_\infty) \quad \forall t \in T^{i+1}, \tilde{\beta}^l < F^i(x_\infty) \). Then by continuity there exists \( k \) such that
\[
F^i(x^i) \geq \tilde{\beta}^l, \quad g(t, x^i) \geq \tilde{\alpha}^i_1 \quad \forall t \in T^{i+1}, \quad \forall k \in K \quad \text{such that} \quad k \geq \max\{l, k_1\}.
\]
As a consequence, because \( \theta \) is nondecreasing, we deduce from inequality (31) that
\[
\tilde{\beta}^l + \sum_{t \in T^k} \frac{\gamma_k}{|T^k_1|} \frac{\theta(\tilde{\alpha}^i_1 \delta_k)}{\delta_k} \leq F(u) + \frac{\gamma_k}{|T^k_1|} \frac{\theta(G(u) \delta_k)}{\delta_k} + \nu_k(u), \quad \text{with} \quad \lim_{k \to \infty} \nu_k(u) = 0.
\]
\[(38)\]

Because \( 0 \leq \theta(G(u) \delta_k) \leq \theta(0) \) and \( \lim_{u \to -\infty} \theta(u) = 0 \), then
\[
\lim_{k \to \infty} \frac{\gamma_k}{\delta_k} \frac{\theta(G(u) \delta_k)}{\delta_k} = 0
\]
under one of the conditions (a), (b), and (c).
It follows that the right-hand side of (38) converges to \( F(u) \) as \( k \to \infty \). As a consequence, because \( \theta \) is nonnegative, we get that

\[
\tilde{\beta}^l \leq F(u), \quad \tilde{\beta} + \frac{\gamma_k}{|T^l_k|} \frac{\theta(\tilde{\alpha}_k | \tilde{\delta}_k)}{\tilde{\delta}_k} \leq F(u) + \kappa_k(u) \quad \forall t \in T^l_k, \quad \text{where } \lim_{k \to \infty} \kappa_k(u) = 0. \tag{40}
\]

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

\[
\tilde{\alpha}_k^l \leq 0 \quad \forall t \in T^l_k. \tag{41}
\]

Letting \( \tilde{\beta}^l \to F^l(x^\infty), \tilde{\alpha}_k^l \to g(t, x^\infty) \quad \forall t \in T^l_k \), it follows from (40) and (41) that

\[
F^l(x^\infty) \leq F(u), \tag{42}
\]

and

\[
g(t, x^\infty) \leq 0 \quad \forall t \in T^l_k \quad \forall l. \tag{43}
\]

Now passing to the limit as \( l \to +\infty \), we get

\[
g(\tilde{t}_2, x^\infty) \leq 0. \tag{44}
\]

Finally,

\[
g(t^{k+1}_2, x^\infty) = g(t^{k+1}_2, x^k) + [g(t^{k+1}_2, x^\infty) - g(t^{k+1}_2, x^k)],
\]

so that, according to (20),

\[
g(t^{k+1}_2, x^\infty) \geq G(x^k) - \mu_x + [g(t^{k+1}_2, x^\infty) - g(t^{k+1}_2, x^k)]. \tag{45}
\]

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 \( u \in C \ (u \in Q \text{ and } G(u) < 0 \text{ if } (A_2) \), by continuity, we get

\[
F(x^\infty) \leq v(P) \quad \forall l. \tag{46}
\]

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

\[
F(x^\infty) \geq f(t^{(l-1)}_1, x^\infty) = f(t^{(l-1)}_1, x^{(l-1)}) + [f(t^{(l)}_1, x^\infty) - f(t^{(l-1)}_1, x^{(l-1)})].
\]

According to (20), \( F(x^{(l)}) \leq f(t^{(l)}_1, x^{(l)}) + \mu_{(l)} \). Passing to the limit in these inequalities, and using (46), we get \( F(x^\infty) \leq v(P) \), which proves that \( x^\infty \in S_P \). □

Remark 3.4. The functions \( \theta_1 \) and \( \theta_2 \) satisfy the assumption \( \theta(0) = 0 \), but not \( \theta_3 \).

Remark 3.5. When \( D \) is defined with a finite number of inequalities \( q \), then we can take \( T^0 = \{1, \ldots, q\} \), and we do not need to compute in RPSALG the element \( t^{k+1}_k \). Obviously, the convergence proof remains valid. Furthermore, in that case we can choose \( m_k = q \) which leads to parameters \( \gamma_k \) smaller than for \( m_k = r_2 + k \).

Remark 3.6. A unified framework for penalty and barrier methods was developed in Auslender [3] 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 §2 are only valid for convex functions, e.g., (12), (13), or Proposition 2.1).

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

It should also be noted that both schemes require summing up over \( T^l_k \) 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 Martinet [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

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

Because $T_2$ is a metric space, $\mathcal{C}(T_2)$ is separable and every finite signed Borel measure on $T_2$ is regular (see, for instance, Bonnans and Shapiro [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.,

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

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

(d) We consider the following function, associated with the Lagrangian function

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

(e) We define the usual Lagrangian dual problem below, associated with our primal problem $(P)$

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

(f) The so-called weak duality inequality $v(P) \leq v(D)$ 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 $\tilde{\sigma} \in M_+ (T_2)$ and $\bar{x} \in \arg \min \psi(\tilde{\sigma})$ are such that

$$g(t, \bar{x}) \leq 0 \quad \forall t \in T_2,$$

and

$$\langle g_\cdot, \tilde{\sigma} \rangle = 0,$$

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

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

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

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

$$\left\| \nabla F(x^k) + \frac{\gamma_k}{|T_2^2|} \sum_{i \in T_2^2} \theta^i g(t, x^k) \delta_i \nabla_x g(t, x^k) + 2 \epsilon_k x^k \right\| \leq \sqrt{2} \epsilon_k.$$  

(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 §3, and

(c') $\theta \in \mathcal{T}_2$, $\lim_{k \to \infty} \delta_k = +\infty$, $\gamma_k/m_k > \varepsilon$ for a certain $\varepsilon > 0$, $\{\gamma_k/\delta_k\}$ 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

\[
\sigma^k := \frac{\gamma_k}{|T^k_2|} \sum_{t \in T^k_2} \theta'(g(t, x^k) \delta_t) \alpha_t,
\]  

where \( \alpha_t \) 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.,

\[
\lim_{k \to \infty} \langle h, \sigma^k \rangle = \langle h, \sigma \rangle, \quad \forall h \in \mathfrak{E}(T_2).
\]

**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 \( S_p \).

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

\[
\forall t \in T_2 \cap I_k: \frac{\gamma_k}{|T^k_2|} \theta'(g(t, x^k) \delta_t)(0 - \delta_t g(t, x^k)) \leq \frac{\gamma_k}{|T^k_2|} (\theta(0) - \theta(g(t, x^k) \delta_t)) \leq \frac{\gamma_k}{|T^k_2|} \theta(0),
\]

and because \( \theta' \) is nonnegative, it follows, from the definition of \( \sigma^k \), that

\[
\langle g_{x^k}, \sigma^k \rangle \geq \sum_{t \in T_2 \cap I_k} \frac{\gamma_k}{|T^k_2|} \theta'(g(t, x^k) \delta_t) g(t, x^k) \geq -\frac{\gamma_k}{\delta_k} \theta(0).
\]  

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

\[
\hat{\sigma}^k := \frac{\sigma^k}{\| \sigma^k \|}, \quad k \in K.
\]

Then recall that the separability of \( \mathfrak{E}(T_2) \) entails that the ball \( B^* := \{ \sigma \in M(T_2) : \| \sigma \| \leq 1 \} \) is weak\(^*\)-sequentially compact. As a consequence of that, and because 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 \( \{ \hat{\sigma}^{k_i} \}_{i \in K'} \), with \( K' \subset K \), such that

\[
\lim_{k \to \infty, k \in K'} x^k = x_\infty \in S_p, \quad w^* - \lim_{k \to \infty, k \in K'} \hat{\sigma}^{k_i} = \hat{\sigma} \in M_1(T_2), \quad \| \hat{\sigma} \| = 1.
\]  

Now from (49) we obtain

\[
\frac{\| \nabla F(x^k) \|}{\| \sigma^k \|} + \langle \nabla_s g_{x^k}, \hat{\sigma}^{k_i} \rangle + \frac{2\epsilon_k x^k}{\| \sigma^k \|} \leq \sqrt{2\epsilon_k}, \quad k \in K'.
\]  

Before taking limits for \( k \to \infty, k \in K' \), we write

\[
\langle \nabla_s g_{x^k}, \hat{\sigma}^{k_i} \rangle = \langle \nabla_s g_{x^k} - \nabla_s g_{x_\infty}, \hat{\sigma}^{k_i} \rangle + \langle \nabla_s g_{x_\infty}, \hat{\sigma}^{k_i} \rangle.
\]  

Using the uniform convergence over the compact set \( T_2 \), we get

\[
\lim_{k \to \infty, k \in K'} \| \nabla_s g_{x^k} - \nabla_s g_{x_\infty} \| = 0.
\]

So, because 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 \( \hat{\sigma}^{k_i} \) to \( \hat{\sigma} \), taking limits in (53) we conclude

\[
\langle \nabla_s g_{x_\infty}, \hat{\sigma} \rangle = 0.
\]

Let us write now

\[
\langle g_{x^k}, \hat{\sigma}^{k_i} \rangle = \langle g_{x^k} - g_{x_\infty}, \hat{\sigma}^{k_i} \rangle + \langle g_{x_\infty}, \hat{\sigma}^{k_i} \rangle.
\]  

\[
\| g_{x^k} - g_{x_\infty} \| + \| g_{x_\infty} \| = \| g_{x^k} \| + \| g_{x_\infty} \| = \| g_{x_\infty} \|,
\]

and because \( \| g_{x_\infty} \| = \| g_{x_{\infty}} \| \leq M \), we have

\[
\langle g_{x^k}, \hat{\sigma}^{k_i} \rangle = \langle g_{x_{\infty}}, \hat{\sigma}^{k_i} \rangle + \langle g_{x^k} - g_{x_{\infty}}, \hat{\sigma}^{k_i} \rangle \leq M - \sqrt{2\epsilon_k}.
\]  

From (50) and (54) we conclude

\[
\langle g_{x^k}, \hat{\sigma}^{k_i} \rangle \leq \sqrt{2\epsilon_k}, \quad k \in K'.
\]

This implies that

\[
\hat{\sigma}^{k_i} = \frac{\sigma^{k_i}}{\| \sigma^{k_i} \|} \to \hat{\sigma} \in \mathfrak{E}(T_2), \quad \| \hat{\sigma}^{k_i} \| = 1,
\]

because the sequence \( \{ \sigma^{k_i} \} \) is strongly bounded and \( \mathfrak{E}(T_2) \) is weak\(^*\)-sequentially compact. Since \( \mathfrak{E}(T_2) \) is also sequentially weak\(^*\)-compact, the sequence \( \{ \sigma^{k_i} \} \) converges weak\(^*\)-almost everywhere to some limit point \( \hat{\sigma} \).
With the same arguments as above, taking limits for \( k \to \infty, \ k \in K' \), we get
\[
\lim_{k \to \infty, \ k \in K'} \langle g_{x_k}, \tilde{\sigma}^k \rangle = \langle g_{x_{\infty}}, \tilde{\sigma} \rangle \leq 0,
\]
(58)
because \( \tilde{\sigma} \in M_+(T_2) \) and \( g(t, x_{\infty}) \leq 0 \ \forall \ t \in T_2 \).

Now, dividing both members of (51) by \( \|\sigma^k\|\), we get
\[
\langle g_{x_k}, \sigma^k \rangle \geq -\frac{\gamma_k}{\delta_k \|\sigma^k\|} \theta(0).\]
(59)
Because \( \lim_{k \to \infty} (\gamma_k / \delta_k \|\sigma^k\|) = 0 \), passing to the limit in this inequality, we get, with (58),
\[
\langle g_{x_{\infty}}, \tilde{\sigma} \rangle = 0.
\]
(60)
Let us now consider \( u \) satisfying Slater’s condition. Because \( \|\tilde{\sigma}\| = 1 \) and \( \tilde{\sigma} \in M_+(T_2) \), it follows that \( \langle g_u, \tilde{\sigma} \rangle < 0 \). Because \( g(t, \cdot) \) 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 that
\[
0 = \langle h, \tilde{\sigma} \rangle = \langle \nabla_x g_{x_{\infty}}, \tilde{\sigma} \rangle, u - x_{\infty} \rangle \leq \langle g_u - g_{x_{\infty}}, \tilde{\sigma} \rangle < 0,
\]
a contradiction.

(ii) Because the sequence \( \{\sigma^k\} \) is bounded, and again applying that \( B^* \) is weak*—sequentially compact, there will exist at least a \( w^* \)—limit point. Let \( \sigma_{\infty} \) be an arbitrary \( w^* \)—limit point of this sequence. Because \( x^k \) is bounded with limit points in \( S_P \), there exists a subsequence \( \{\sigma^k\}_{k \in K} \) such that
\[
\lim_{k \to \infty, \ k \in K} x^k = x_{\infty} \in S_P, \quad w^* - \lim_{k \to \infty, \ k \in K} \sigma^k = \sigma_{\infty} \in M_+(T_2).
\]
(61)
Using the same arguments as for (58) we get \( \langle g_{x_{\infty}}, \sigma_{\infty} \rangle \leq 0 \). Now, because either \( \theta(0) = 0 \) or \( \lim_{k \to \infty} (\gamma_k / \delta_k) = 0 \), and passing to the limit in (51), we get
\[
\langle g_{x_{\infty}}, \sigma_{\infty} \rangle = 0.
\]
(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, like all Remez-type methods, requires solving nonconvex optimization problems in Step 2. From a computational point of view, 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 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 \( h: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) is convex, lsc, and \( C^1 \) on \( Q \). For more general cases we refer to Remark 5.2 below.

For \( \delta_k > 0, \ p_k > 0, \ \gamma_k > 0, \) and \( \theta \in \mathcal{F} \), we set
\[
I_k(x) = \frac{1}{p_k} \log \left( \int_{T_1} \exp \left( \langle (a(t), x) - b(t) \rangle \right) dt \right), \quad E_k(x) = \gamma_k \int_{T_2} \frac{\theta(g(t, x) \delta_k)}{\delta_k} dt,
\]
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
\[
(P^k_{\epsilon_p}) \inf \{R_k(x) + \epsilon_k \|x\|^2 \ | \ x \in Q \}.
\]
(63)
Observe that \( I_k \) is convex (Fang and Wu [12, Lemma 1]) and, obviously, \( E_k \) is also convex, so that \( R_k \) is convex, lsc, and \( C^1 \) on \( Q \). Consequently, the objective function of \( (P^k_{\epsilon_p}) \) is strongly convex on \( Q \) and there exists at least a point \( x^k \) satisfying
\[
x^k \in Q, \quad h(x^k) + I_k(x^k) + E_k(x^k) + \epsilon_k \|x^k\|^2 \leq h(u) + I_k(u) + E_k(u) + \epsilon_k \|u\|^2 + \epsilon_k \ \forall \ u \in Q.
\]
(64)
**Remark 5.1.** Because $h$ and $\theta$ are $C^1$, the objective function of $P_{\text{eps}}^k$ is also $C^1$, with
\begin{equation}
\nabla l_k(x) = \frac{\int_t \exp\left([a(t), x] - b(t)\right)p_k(\nabla a(t) dt}{\int_t \exp\left([a(t), x] - b(t)\right)p_k(\nabla a(t) dt}, \quad \nabla E_k(x) = \gamma_k \int_t \theta'(g(t, x)\delta_k) a(t) dt.
\end{equation}

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
\begin{equation}
\|\nabla R_k(x^k) + 2\varepsilon_k x^k\| \leq \sqrt{2}\varepsilon_k.
\end{equation}

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

(a') $\theta \in \mathcal{T}$, $\lim_{k \to \infty} (\gamma_k / \delta_k) = 0$, and $\lim_{k \to \infty} \gamma_k = +\infty$.

(b') $\theta \in \mathcal{T}$, $\lim_{k \to \infty} (\gamma_k / \delta_k) = 0$, and $\gamma_k > e$ $\forall k$, for a certain $e > 0$.

Now we describe our second algorithm as follows:

**Integral penalty smoothing algorithm—IPSA\textsc{lg}:**

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

From now on, for each $V \subset T$, we set $\nu(V) = \int_V dt$. Then we have the following lemma:

**Lemma 5.1.** Let $u \in Q$, $\tau(u) := \max\{a(t), u - b(t) : t \in T_1\}$, and let $\{u^k\}$ be a sequence in $Q$ converging to $u$. Then
\begin{equation}
l_k(u) \leq \tau(u) + \frac{\log(\nu(T_k))}{p_k} \quad \text{and} \quad \lim_{k \to \infty} l_k(u^k) = \tau(u).
\end{equation}

**Proof.**

Set
\begin{equation}
l(t) := a(t, u) - b(t), \quad l_k(t) := a(t, u^k) - b(t), \quad \text{and} \quad t^*_k \in \arg\max\{l(t) : t \in T_1\}.
\end{equation}

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 $\varepsilon_k > 0$ and $n_1$ such that
\begin{equation}
l_k(t) \geq \tau(u^k) - \beta \quad \forall t \in B(t^*_k; \varepsilon_k) := \{t \in T_1 : d(t^*_k, t) \leq \varepsilon_k\} \quad \text{and} \quad k \geq n_1.
\end{equation}

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 $d(t^*_k, t_r) \leq 1/r$ and $l_k(t_r) < \tau(u^k) - \beta$. The sequence $\{t_r\}$ converges to $t^*_k$, and because the function $\tau$ is continuous, passing to the limit, we get $\tau(u) \leq \tau(u) - \beta$, a contradiction. As a consequence, we get, denoting $B_r := B(t^*_k; \varepsilon_k)$,
\begin{equation}
l_k(u^k) \geq \frac{1}{p_k} \log(\int_{B_r} \exp[l_k(t)p_k] dt) \geq \frac{1}{p_k} \log(\int_{B_r} \exp[(\tau(u^k) - \beta)p_k] dt)
= \left(\tau(u^k) - \beta\right) + \frac{\log(\nu(B_r))}{p_k}.
\end{equation}

Passing to the limit, because $\tau$ is continuous, we obtain $\lim_{k \to \infty} l_k(u^k) \geq \tau(u) - \beta$, and then, with $\beta \to 0^+$, we obtain $\lim_{k \to \infty} l_k(u^k) = \tau(u)$.

Now, using the first inequality in (67) and the continuity of $\tau$, we get $\limsup_{k \to \infty} l_k(u^k) \leq \tau(u)$ so that $\lim_{k \to \infty} l_k(u^k) = \tau(u)$. □

**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 $\{\chi^k\}$ built by IPSALG is bounded and all its limit points are in $S_p$.

**Proof.**

(1) Let $u \in C$ and set
\begin{equation}
\eta_k(u) := \gamma_k \nu(T_2) \frac{\theta(G(u)\delta_k)}{\delta_k} + \frac{\log(\nu(T_1 ))}{p_k} + e_k.
\end{equation}

Because $0 \leq \theta(G(u)\delta_k) \leq \theta(0)$, it follows from (68), and at least one of the two conditions (a') and (b') that
\begin{equation}
\lim_{k \to \infty} \eta_k(u) = 0.
\end{equation}
Now, because $\theta$ 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, because $E_k(x^k)$ is nonnegative, there exists some $\alpha \in \mathbb{R}$ such that, for $k$ sufficiently large, we have

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

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

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

Let $r_1^* \in \arg \max \{ \langle a(t), d \rangle \mid t \in T_1 \}$. Then (see §2) $F_{a_0}(d) = h_{a_0}(d) + \langle a(t_1^*), d \rangle$. Now let $\tau < \langle a(t_1^*), d \rangle$. Then, because $a(\cdot)$ 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 $u_k(t) := \langle a(t), x^k/\|x^k\| \rangle - b(t)/\|x^k\|$. Because $a(\cdot), b(\cdot)$ 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

$$u_k(t) \geq r_1 + \tau, \quad \forall t \in B_1, \quad \text{and} \quad \forall k \geq k_0, \quad k \in K.$$ \hspace{1cm} Then, $\forall k \geq k_0$ we have

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

$$= (\tau + r_1) \|x^k\| + \frac{\log(r(B_1))}{p_k},$$

and this entails

$$\liminf_{k \to +\infty, k \in K} \frac{I_k(x^k)}{\|x^k\|} \geq \tau + r_1.$$ \hspace{1cm} Then, taking $\tau \to \langle a(t_1^*), d \rangle$, which implies $r_1 \to 0^+$, we get

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

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 $h_\infty$:

$$0 = \lim_{k \to +\infty, k \in K} \frac{\alpha}{\|x^k\|} \geq \liminf_{k \to +\infty, k \in K} \frac{h(x^k)}{\|x^k\| + I_k(x^k)} + \liminf_{k \to +\infty, k \in K} \frac{I_k(x^k)}{\|x^k\|}$$

$$\geq \liminf_{k \to +\infty, k \in K} \frac{h(x^k)}{\|x^k\|} + \liminf_{k \to +\infty, k \in K} \frac{I_k(x^k)}{\|x^k\|}$$

$$\geq h_\infty(d) + \langle a(t_1^*), d \rangle = F_{a_0}(d). \quad (75)$$

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

$$0 = \lim_{k \to +\infty, k \in K} \frac{\alpha}{\|x^k\|} \geq \liminf_{k \to +\infty, k \in K} \left( \frac{h(x^k)}{\|x^k\|} + \frac{I_k(x^k)}{\|x^k\|} \right) + \liminf_{k \to +\infty, k \in K} \frac{E_k(x^k)}{\|x^k\|}$$

$$\geq h_\infty(d) + \langle a(t_1^*), d \rangle + \liminf_{k \to +\infty, k \in K} \frac{E_k(x^k)}{\|x^k\|}. \quad (76)$$

Now, we prove that

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

in which case, using (22), relations (72), (75), (77) would imply that Assumption (A1) is not satisfied, and this is a contradiction. So, suppose that (77) does not hold. Then, because $a(\cdot)$ is continuous, there exist $t^* \in T_2$, $r > 0$ 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 \( u_k(t) := \langle a(t), x^k / \| x^k \| \rangle - b(t) / \| x^k \|. \) Because \( a(\cdot), b(\cdot) \) are continuous, the sequence \( \{ u_k \}_{k \in K} \) converges uniformly on \( B_t \) to \( \langle a(\cdot), d \rangle \). Hence, there exists \( k_1 \) such that

\[
u(B_t) \geq \frac{\theta(r \delta_k)}{\delta_k} \leq \beta. \tag{82}
\]

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. □

**Remark 5.2.** When \( Q \) is bounded note that the proof of Theorem 5.1 remains valid for functions \( f_j \) and \( g_t \), not necessarily affine. Indeed, the proofs of Lemma 5.1 and parts (1) and (3) of Theorem 5.1 remain valid with almost no change. Meanwhile, part (2) becomes unnecessary because \( Q \) is bounded.

**Remark 5.3.** When \( T_1 = \varnothing \) and \( T_2 \neq \varnothing \) we define \( F := h, I_k := 0 \), while if \( T_2 = \varnothing \) and \( T_1 \neq \varnothing \), then \( D = \mathbb{R}^n \) 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 \( \{T_i\} \) is reduced to a single element, and that \( x^i \) satisfies (66). As pointed out in Remark 5.1, \( x^i \) satisfies (64), so that Theorem 5.1 holds.

Let us introduce a linear map \( J_0: C_+ (T_2) \to M_\epsilon (T_2) \) as follows:

\[
\langle J_0 f, h \rangle := \int_{T_2} h f \, dt \quad \forall \ f \in C_+ (T_2), \ \forall \ h \in C (T_2).
\]

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

\[
\langle g_{k^*}, \sigma^k \rangle \geq \frac{\gamma_k}{\delta_k} \int_{I_k} g(t, x^k)\theta(g(t, x^k)\delta_k) dt \geq -\frac{\gamma_k}{\delta_k} \nu(T_2) \theta(0).
\]

(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 \( \lim_{k \to \infty, k \in K} \|\sigma^k\| = \infty \), and we define the measures \( \tilde{\sigma}^k := \sigma^k / \|\sigma^k\|, \ k \in K \).

Following the same arguments as in Theorem 4.2, we now conclude from (83) that

\[
0 \geq \langle g_{k^*}, \tilde{\sigma} \rangle = \lim_{k \to \infty, k \in K} \langle g_{k^*}, \tilde{\sigma}^k \rangle \geq \lim_{k \to \infty, k \in K} -\frac{\gamma_k}{\delta_k \|\sigma^k\\} \nu(T_2) \theta(0) = 0.
\]

In other words, \( \langle g_{k^*}, \tilde{\sigma} \rangle = 0 \), 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 Auslender [2], Lin et al. [16], and Teo et al. [30] deal with problems of the 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 Auslender [2], \( \theta = \theta_0 \), while in Lin et al. [16], \( \theta = \theta_0 \), so that in both cases \( \theta \in \mathcal{F}_2 \). Furthermore, in both cases \( \gamma_k = 1 \) and \( \lim_{k \to \infty} \delta_k = +\infty \), so that condition (b') holds and they coincide with IPSALG when \( \epsilon_k = 0 \ \forall \ k \). In Auslender [2], \( Q \) is supposed to be compact. In Lin et al. [16], it is assumed that (A_1) and the 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 Teo et al. [30] appears without conditions on \( \gamma_k > 0, \delta_k > 0 \) as a particular case of IPSALG with \( \epsilon_k = 0 \ \forall \ k \) and \( \theta = \theta_0 \in \mathcal{F}_1 \). However these two algorithms are different because the parameters are chosen differently. Indeed in Teo et al. [30] \( \delta_k \to \infty \) and for \( \delta_k \) fixed \( \gamma_k \) is chosen such that \( x^k \) is, in addition, feasible, while for IPSALG they must satisfy condition (a'). Moreover, in Teo et al. [30], \( Q \) is supposed to be compact, the Slater condition is also assumed, and no duality result is provided.

In Fang and Wu [12] the problem is of the 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 \( \lim_{k \to \infty} p_k = \infty \). In Fang and Wu [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 also be 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 a better understanding of their numerical behavior, future works on implementing and testing these methods should be conducted.
Acknowledgments. The second and third authors were supported by DGES of Spain and FEDER of EU, Grant MTM2005-08572-C03-01.

References