RAIRO. RECHERCHE OPÉRATIONNELLE

CARLOS HUMES JR.
PAULO JOSÉ DA SILVA E SILVA

Strict convex regularizations, proximal points and augmented lagrangians

RAIRO. Recherche opérationnelle, tome 34, nº 3 (2000), p. 283-303

<http://www.numdam.org/item?id=RO_2000__34_3_283_0>

RAIRO Operations Research
RAIRO Oper. Res. **34** (2000) 283-303

STRICT CONVEX REGULARIZATIONS, PROXIMAL POINTS AND AUGMENTED LAGRANGIANS (*)

by Carlos HUMES Jr. ([1, 2]) and Paulo José DA SILVA E SILVA ([3, 4])

Communicated by Pierre TOLLA

Abstract. *Proximal Point Methods (PPM) can be traced to the pioneer works of Moreau [16], Martinet [14, 15] and Rockafellar [19, 20] who used as regularization function the square of the Euclidean norm. In this work, we study PPM in the context of optimization and we derive a class of such methods which contains Rockafellar's result. We also present a less stringent criterion to the acceptance of an approximate solution to the subproblems that arise in the inner loops of PPM. Moreover, we introduce a new family of augmented Lagrangian methods for convex constrained optimization, that generalizes the P_E^+ class presented in [2].*

Keywords: Proximal points methods, augmented Lagrangians, convex programming.

1. INTRODUCTION

In recent years, Proximal Point Methods (PPM) have been receiving plenty of attention in the literature. These methods can be characterized roughly along two main lines: Proximal Point Methods with embedded penalties (typically Bregman distances and φ-divergences [4, 10–12, 24]) and "pure regularization" methods. The classical method in the last approach is presented in [20]: at each iteration, we calculate

$$x^{i+1} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2} \|x - x^i\|_2^2 \right\},$$

where $f : \mathbb{R}^n \to (-\infty, \infty]$ is the function we want to minimize.

In this article, we prove the convergence of a generalization of the classical PPM where the square of the Euclidean norm is substituted by a strict convex

(*) Received February, 1998.

([1]) Departemento de Ciência da Computação, Instituto de Matemática e Estatística – USP, rua de Matao 1010, Cidade Universitaria, CEP 05315–970 Sao Paulo, SP, Brazil.

([2]) Research of this autor is suported by CNPq and PRONEX, Convênio 76.97.1008.00.

([3]) Departamento de Matemática Aplicada, Instituto de Matemática e Estatística – USP, rua de Matao 1010, Cidade Universitaria, CEP 05315–970 Sao Paulo, SP, Brazil.

([4]) Research of this autor is suported by FAPESP – Processo 96/09939-0.

Recherche opérationnelle/Operations Research, 0399-0559/00/03/$ 7.00

function with null gradient at the origin. We also show that the subproblems may be solved with a fixed relative precision.

The remainder of this paper is organized as follows: in Section 2, we introduce a generalization of the PPM, that we call ϕ-PPM, and we prove a convergence theorem. In Section 3, we weaken the assumption of exact solution to the subproblems arising in the ϕ-PPM. In Section 4, we derive a new family of augmented Lagrangians based on the exact version of the ϕ-PPM. Some final comments are presented in Section 5, including a discussion of the limitations of the approach used here.

2. PROXIMAL POINT METHODS AND STRICTLY CONVEX FUNCTIONS

In this section, we consider the problem of minimizing a convex function, $f(\cdot)$, in $\mathbb{R}^n$. We assume that:

ASSUMPTION 1: $f : \mathbb{R}^n \to (-\infty, \infty]$ is convex, proper, lower semi-continuous (lsc) and bounded from below.

To solve this minimization problem, we introduce a generalized proximal method based on a regularization function $\phi : \mathbb{R}^n \to (-\infty, \infty]$:

ϕ-Proximal Point Method (ϕ-PPM)

Let $\{\alpha_i\}_{i \in \mathbb{N}}$ be a sequence of positive real numbers bounded from above by $\widetilde{\alpha} > 0$.

1. *Start with some $x^1 \in \operatorname{dom} f$.*
2. *Calculate*

$$x^{i+1} \doteq \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ f_i(x) \doteq f(x) + \alpha_i \phi(x - x^i) \right\},$$

If $x^{i+1} = x^i$, stop.

Where $\phi(\cdot)$ obey the following assumption:

ASSUMPTION 2: The function $\phi(\cdot)$ is strictly convex and lsc, $\phi(0) = 0$ and $\phi(\cdot)$ is differentiable at the origin with $\nabla \phi(0) = 0$ ([1]).

Note that this assumption implies that $\phi(\cdot)$ is inf-compact, *i.e.*, the level set

$$L_\lambda \doteq \{x \mid \phi(x) \leqq \lambda\}$$

([1]) The assumption that $\phi(0) = 0$ is not necessary, but it simplifies almost all expressions. In particular, it implies that $\phi(\cdot)$ is non-negative.

is compact for all λ, since $L_0 = \{x \mid \phi(x) \leqq 0\} = \{0\}$ (see Sect. 8.7.1 in [18]).

Note that the sequence $\{x^i\}_{i\in\mathbb{N}}$ is well defined. In fact, Assumptions 1 and 2 guarantee that $f_i(\cdot)$ is inf-compact, hence it attains its minimum. Furthermore, $f_i(\cdot)$ is strictly convex, due to the strict convexity of $\phi(\cdot)$, hence the minimizer is unique.

Another important remark is that the ϕ-PPM has as fixed points the minimizers of $f(\cdot)$. This is a trivial consequence of the assumption that $\nabla\phi(0) = 0$.

We will start the convergence analysis of the ϕ-PPM by presenting an auxiliary lemma that characterizes some important properties of $\phi(\cdot)$ near the origin.

LEMMA 1: *Let $\phi(\cdot)$ be a function under Assumption 2, then for any sequence $\{z^i\}_{i\in\mathbb{N}}$ the following statements are equivalent:*

(i) $\phi(z^i)\to 0$;

(ii) $z^i\to 0$;

(iii) *There is a $N > 0$ such that for all $i > N$, $\phi(\cdot)$ is subdifferentiable at z^i. Moreover, $\forall i > N$, $\forall\gamma^i \in \partial\phi(z^i)$, $\gamma^i\to 0$.*

(iv) *There is a $N > 0$ such that for all $i > N$, $\phi(\cdot)$ is subdifferentiable at z^i. Moreover, $\forall i > N$, $\forall\gamma^i \in \partial\phi(z^i)$, $\langle\gamma^i, z^i\rangle\to 0$.*

Proof:

1. $(i) \Rightarrow (ii)$. If $\phi(z^i)\to 0$ then $\{z^i\}_{i\in\mathbb{N}}$ is bounded, as $\phi(\cdot)$ is inf-compact. Let $\{z^i\}_{i\in\mathcal{K}}$ be any convergent subsequence of $\{z^i\}_{i\in\mathbb{N}}$. Since $\phi(\cdot)$ is lsc it follows that

$$\phi(\bar{z}) \leqq \lim_{i\in\mathcal{K}} \phi(z^i) = 0 = \min_{z\in\mathbb{R}^n}\{\phi(z)\}.$$

Then $\bar{z}$ is the argument that minimize $\phi(\cdot)$, *i.e.* the origin.

2. $(ii) \Rightarrow (iii)$. Since $z^i\to 0$ there is an $N > 0$ such that for all $i > N$, z^i is in the interior of $\operatorname{dom}\phi$ where $\phi(\cdot)$ is subdifferentiable. For $i > N$, let $\gamma^i \in \partial\phi(z^i)$, this sequence is bounded (see Th. 24.7 in [18]) and, due to the outer semi-continuity (osc) of $\partial\phi(\cdot)$, all accumulations points of γ^i must be in $\partial\phi(0) = \{0\}$.

3. $(iii) \Rightarrow (iv)$. It suffices to show that $\{z^i\}_{i\in\mathbb{N}}$ is bounded. Since $\gamma^i\to 0$ and

$$\frac{\langle\gamma^i, z^i\rangle}{\|z^i\|} \geqq \frac{\phi(z^i) - \phi(0)}{\|z^i\|} = \frac{\phi(z^i)}{\|z^i\|} \geqq 0,$$

it follows that

$$\frac{\phi(z^i)}{\|z^i\|} \to 0.$$

Then, $\{z^i\}_{i\in\mathbb{N}}$ must be bounded, otherwise we would have a contradiction with the inf-compactness of $\phi(\cdot)$, which is equivalent to level coercitivity (see Sect. 3.27 in [21]).

4. $(iv) \Rightarrow (i)$. $\langle \gamma^i, z^i \rangle \geqq \phi(z^i) - \phi(0) = \phi(z^i) \geqq 0$, the result follows. □

Now we can prove the convergence theorem for the ϕ-PPM.

THEOREM 1: *Assume that Assumptions 1 and 2 hold and that the sequence $\{x^i\}_{i\in\mathbb{N}}$ is generated by the ϕ-PPM. If a subsequence $\{x^i\}_{i\in\mathcal{K}}$ converges to a point $\bar{x}$, then $\bar{x}$ is a minimizer of $f(\cdot)$ and the subsequence $\{x^{i+1}\}_{i\in\mathcal{K}}$ converges to $\bar{x}$ as well.*

Proof: Let $\widetilde{\alpha}$ be an upper bound of $\{\alpha_i\}_{i\in\mathbb{N}}$. In order to avoid technical difficulties associated with a vanishing subsequence of α_i we define:

$$\widetilde{x}^i \doteq \operatorname*{argmin}_{x\in\mathbb{R}^n}\{f(x) + \widetilde{\alpha}\phi(x - x^i)\}.$$

Clearly,

$$f(\widetilde{x}^i) + \widetilde{\alpha}\phi(\widetilde{x}^i - x^i) \leqq f(x^{i+1}) + \widetilde{\alpha}\phi(x^{i+1} - x^i)$$
$$f(x^{i+1}) + \alpha_i\phi(x^{i+1} - x^i) \leqq f(\widetilde{x}^i) + \alpha_i\phi(\widetilde{x}^i - x^i).$$

The above inequalities and $\widetilde{\alpha} \geqq \alpha_i$ imply that

$$f(x^{i+1}) \leqq f(\widetilde{x}^i).$$

Hence

$$f(x^i) \geqq f(\widetilde{x}^i) + \widetilde{\alpha}\phi(\widetilde{x}^i - x^i) \geqq f(x^{i+1}) + \widetilde{\alpha}\phi(\widetilde{x}^i - x^i).$$

Since, by Assumption 1, $f(\cdot)$ is bounded from below we have:

$$0 = \lim_{i\to\infty} f(x^i) - f(x^{i+1}) \geqq \widetilde{\alpha}\phi(\widetilde{x}^i - x^i) \geqq 0.$$

Then

$$\phi(\widetilde{x}^i - x^i) \to 0. \tag{1}$$

The definition of $\widetilde{x}^i$ ensures that there are $\widetilde{\gamma}_f^i \in \partial f(\widetilde{x}^i)$ and $\widetilde{\gamma}_\phi^i \in \partial\phi(\widetilde{x}^i - x^i)$ such that

$$\widetilde{\gamma}_f^i + \widetilde{\alpha}\widetilde{\gamma}_\phi^i = 0.$$

From equation (1) and Lemma 1, it follows that:

$$\widetilde{x}^i - x^i \to 0 \text{ and } \widetilde{\gamma}_f^i = -\widetilde{\alpha}\widetilde{\gamma}_\phi^i \to 0.$$

Now, let $\{x^i\}_{i\in\mathcal{K}}$ be any convergent subsequence of $\{x^i\}_{i\in\mathbb{N}}$, $x^i \to_{\mathcal{K}} \bar{x}$. The above result implies that $\widetilde{x}^i \to_{\mathcal{K}} \bar{x}$ and, due to the outer semi-continuity of $\partial f(\cdot)$,

$$0 \in \partial f(\bar{x}).$$

Now, let us prove that $\{x^{i+1}\}_{i\in\mathcal{K}}$ converges to $\bar{x}$. Indeed

$$f(x^{i+1}) + \alpha_i\phi(x^{i+1} - x^i) \leqq f(\bar{x}) + \alpha_i\phi(\bar{x} - x^i) \leqq f(x^{i+1}) + \alpha_i\phi(\bar{x} - x^i).$$

Therefore

$$0 \leqq \phi(x^{i+1} - x^i) \leqq \phi(\bar{x} - x^i).$$

As $\phi(\bar{x} - x^i) \to_{\mathcal{K}} 0$, then $\phi(x^{i+1} - x^i) \to_{\mathcal{K}} 0$ and finally, $x^{i+1} \to_{\mathcal{K}} \bar{x}$. □

COROLLARY 1: *Under the same assumptions of Theorem 1, if $\{x^i\}_{i\in\mathbb{N}}$ has a limit point, then*

$$f(x^i) \to \inf_{x\in\mathbb{R}^n} \{f(x)\}.$$

Proof: Let $\{x^i\}_{i\in\mathcal{K}}$ be any convergent subsequence of $\{x^i\}_{i\in\mathbb{N}}$, $x^i \to_{\mathcal{K}} \bar{x}$. Theorem 1 asserts that $\bar{x}$ is a minimizer of $f(\cdot)$ and that $x^{i+1} \to_{\mathcal{K}} \bar{x}$. Let $\gamma_\phi^{i+1} \in \partial\phi(x^{i+1} - x^i)$ such that $\exists \gamma_f^{i+1} \in \partial f(x^{i+1})$, $\alpha_i\, \gamma_\phi^{i+1} = -\gamma_f^{i+1}$.

Since $x^{i+1} - x^i \to_{\mathcal{K}} 0$ it follows that $\gamma_\phi^{i+1} \to_{\mathcal{K}} 0$. Therefore,

$$f(\bar{x}) \leqq \lim_{i \xrightarrow{\mathcal{K}} \infty} f\left(x^{i+1}\right) \leqq \lim_{i \xrightarrow{\mathcal{K}} \infty} f(\bar{x}) + \alpha_i \left\langle \gamma_\phi^{i+1}, \bar{x} - x^{i+1} \right\rangle = f(\bar{x}).$$

As $\{f(x^i)\}_{i\in\mathbb{N}}$ is non-increasing and bounded, it is convergent, and we get the desired result. □

3. AN INEXACT VERSION OF ϕ-PPM

In [20], Rockafellar showed that instead of solving exactly the subproblem

$$\text{Find } x^{i+1} = \operatorname{argmin}\Big\{ f(x) + \frac{\alpha_i}{2}\|x - x^i\|^2 \Big\},$$

the PPM converges to a minimizer of $f(\cdot)$ if we find a x^{i+1} such that for some $\gamma_f^i \in \partial f(x^{i+1})$

$$\|\gamma_f^i + \alpha_i(x^{i+1} - x^i)\| \leqq \sigma_i \alpha_i, \quad \sum_{i=1}^{\infty} \sigma_i < \infty, \tag{A'}$$

$$\|\gamma_f^i + \alpha_i(x^{i+1} - x^i)\| \leqq \sigma_i \alpha_i \|x^{i+1} - x^i\|, \quad \sum_{i=1}^{\infty} \sigma_i < \infty. \tag{B'}$$

He used (A′) to prove convergence and (B′) to guarantee linear convergence rate.

This result was improved by Solodov and Svaiter in [22] ([2]). Adding a projection after each proximal step, they proved that it is enough to find x^{i+1} such that

$$\|\gamma_f^i + \alpha_i(x^{i+1} - x^i)\| \leqq \sigma \max\big\{\|\gamma_f^i\|, \ \alpha_i\|x^{i+1} - x^i\|\big\},$$

for some fixed σ is in $[0, 1)$. A similar result can be proved for more general Bregman distances, substituting the projection by another extra-gradient step [23]. They also proved linear convergence rate.

Inspired by the last results, we introduce a similar acceptance criterion. Under this criterion we prove convergence without the need for an extra-gradient step, but we do not analyze convergence rate. It must be noted that the results in [20, 22, 23] deal with the problem of finding zeroes for a maximal monotone operator and we constrain ourselves to the minimization of a convex function, where it can be shown that under the criterion of [22] no extra-gradient step is needed [8].

THEOREM 2: *Suppose that Assumptions 1 and 2 hold. Let $\sigma \in [0, 1)$ and $\{\alpha_i\}_{i\in\mathbb{N}}$ be a positive real sequence bounded from above by $\tilde{\alpha}$. Let $\{x^i\}_{i\in\mathbb{N}}$*

([2]) The reader should keep in mind that the articles [20] and [22], deal with the case $\phi(\cdot) = \frac{1}{2}\|\cdot\|^2$.

be any sequence generated iteratively such that for some $\gamma_f^i \in \partial f(x^{i+1})$ *and* $\gamma_\phi^i \in \partial\phi(x^{i+1} - x^i)$,

$$\|\gamma_f^i + \alpha_i \gamma_\phi^i\| \leqq \sigma \frac{\alpha_i \langle \gamma_\phi^i, x^{i+1} - x^i \rangle}{\|x^{i+1} - x^i\|}. \tag{AC}$$

Suppose that one of the following conditions hold:

(i) $\{\alpha_i\}_{i\in\mathbb{N}}$ *is bounded from below by some* $\widehat{\alpha} > 0$;

(ii) *The sequence* $\Delta x^i \doteq x^{i+1} - x^i$ *is contained in some compact,* K, *in the interior of* $\operatorname{dom}\phi$;

(iii) *There is an* $\epsilon > 0$ *such that for all* $x \in \mathbb{R}^n$ *and* $\gamma_\phi \in \partial\phi(x)$

$$\frac{\langle \gamma_\phi, x \rangle}{\|\gamma_\phi\|\|x\|} \geqq \epsilon,$$

then $\gamma_f^i \to 0$. *In particular, any accumulation point of* $\{x^i\}_{i\in\mathbb{N}}$ *is a minimizer of* $f(\cdot)$.

Proof: In order to prove this result, we shall consider, by contradiction, a subsequence $\{\gamma_f^i\}_{i\in\mathcal{K}}$ such that $\gamma_f^i \not\to_{\mathcal{K}} 0$.

As $0 \leqq \alpha_i \leqq \widetilde{\alpha}$, we can assume without the loss of generality that $\exists \bar{\alpha} \geqq 0,\ \alpha_i \to_{\mathcal{K}} \bar{\alpha}$. The basis for the proof are relations (2) and (3) below:

$$\begin{aligned} f(x^i) &\geqq f(x^{i+1}) + \langle \gamma_f^i, x^i - x^{i+1} \rangle \\ &= f(x^{i+1}) + \langle \alpha_i \gamma_\phi^i, x^{i+1} - x^i \rangle + \langle \gamma_f^i + \alpha_i \gamma_\phi^i, x^i - x^{i+1} \rangle \\ &\geqq f(x^{i+1}) + \alpha_i \langle \gamma_\phi^i, x^{i+1} - x^i \rangle - \|\gamma_f^i + \alpha_i \gamma_\phi^i\| \|x^{i+1} - x^i\| \\ &\geqq f(x^{i+1}) + (1 - \sigma)\alpha_i \langle \gamma_\phi^i, x^{i+1} - x^i \rangle, \end{aligned}$$

where the last inequality is a consequence of the acceptance criterion (AC). Therefore

$$\alpha_i \langle \gamma_\phi^i, x^{i+1} - x^i \rangle \to 0. \tag{2}$$

On the other, using (AC) once more:

$$\|\gamma_f^i\| \leqq 2\alpha_i \|\gamma_\phi^i\|. \tag{3}$$

The first case we consider is associated with $\bar{\alpha} > 0$ (automatically guaranteed if (i) holds). In this case, equation (2) implies that

$$\langle \gamma_\phi^i, x^{i+1} - x^i \rangle \to_{\mathcal{K}} 0.$$

Applying Lemma 1 and the above relation (3), it follows that $\gamma_f^i \to_{\mathcal{K}} 0$.

The next cases are associated with $\bar{\alpha} = 0$. If condition (ii) holds, then $\forall i \in \mathcal{K}$, $\gamma_\phi^i \in \partial\phi(K)$ and therefore $\{\gamma_\phi^i\}_{i\in\mathcal{K}}$ is bounded. Using equation (3) it follows that $\gamma_f^i \rightarrow_{\mathcal{K}} 0$.

If condition (iii) holds, by relation (2),

$$\alpha_i \|\gamma_\phi^i\| \|x^{i+1} - x^i\| \epsilon \rightarrow_{\mathcal{K}} 0.$$

Suppose, by contradiction, that there is a subsequence $\mathcal{K}' \subset \mathcal{K}$ and a $\delta > 0$ such that $\forall i \in \mathcal{K}'$, $\alpha_i\|\gamma_\phi^i\| > \delta$. Then we should have $\|x^{i+1} - x^i\| \rightarrow_{\mathcal{K}'} 0$. But, from Lemma 1, this would imply that $\alpha_i \gamma_\phi^i \rightarrow_{\mathcal{K}'} 0$, a contradiction. Then we have $\alpha_i \gamma_\phi^i \rightarrow_{\mathcal{K}} 0$, and from equation (3), $\gamma_f^i \rightarrow_{\mathcal{K}} 0$.

Hence, in any case we contradict the existence of a subsequence such that $\gamma_f^i \not\rightarrow_{\mathcal{K}} 0$. Therefore $\gamma_f^i \rightarrow 0$ and the outer semi-continuity of $\partial f(\cdot)$ implies that any accumulation point of this sequence is a minimizer of $f(\cdot)$. □

4. AUGMENTED LAGRANGIANS

In this section, we use the exact version of the ϕ-PPM, presented in Section 2, to derive a class of augmented Lagrangians for convex programming. This class is a generalization of the P_E^+ class presented in [2].

Through this section we are concerned with the following convex programming problem:

$$\begin{cases} \min & f(x) \\ s.t. & g(x) \leqq 0 \\ & x \in X^0 \end{cases} \tag{P}$$

where $f(\cdot)$ is a convex function from $\mathbb{R}^n$ to $\mathbb{R}$, $g(\cdot)$ is a function from $\mathbb{R}^n$ to $\mathbb{R}^m$ such that each component function, $g_i(\cdot)$, is convex and X^0 is a convex set. We also assume that its optimal value is finite. This problem will be sometimes referred as the *primal* problem.

There is a concave maximization problem closely related to the primal problem, it is called the *minimax dual* problem:

$$\begin{cases} \max & F(\lambda) = \inf_{x\in X^0} \{L(x,\lambda)\} \\ s.t. & \lambda \geqq 0 \end{cases} \tag{D}$$

where $L : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is the Lagrangian function associated with (P):

$$L(x, \lambda) = f(x) + \langle \lambda, g(x) \rangle.$$

A revision of duality is important as we shall treat augmented Lagrangian methods as the solution to the dual problem using proximal points methods (like in [3, 19, 24]).

4.1. Preliminary results

The dual problem and its connection to the primal have been extensively studied [3, 6, 9, 18, 25] and, for the sake of completeness, we review the main results used in this section.

A key concept to duality theory is the convex (concave) conjugate of a function:

DEFINITION 1: Let $f : \mathbb{R}^n \to [-\infty, \infty]$, the convex conjugate of $f(\cdot)$, denoted $f^*(\cdot)$, is the convex function, $f^* : \mathbb{R}^n \mapsto [-\infty, \infty]$, defined by:

$$f^*(x^*) = \sup_z \{\langle z, x^* \rangle - f(z)\}.$$

Analogously, the concave conjugate of $f(\cdot)$, denoted $f_*(\cdot)$, is the concave function, $f_* : \mathbb{R}^n \mapsto [-\infty, \infty]$, defined by:

$$f_*(x^*) = \inf_z \{\langle z, x^* \rangle - f(z)\}.$$

One of the most important results in the theory of conjugate functions is the Fenchel's duality theorem [18]. This theorem presents sufficient conditions to assert that the Fenchel's inequality,

$$\inf_x \{f(x) - g(x)\} \geqq \sup_{x^*} \{g_*(x^*) - f^*(x^*)\}, \tag{4}$$

holds as an equality (strong duality theorem). In [17], Rockafellar characterizes the points where $f(\cdot) - g(\cdot)$ achieve its minimum if strong duality holds.

THEOREM 3: *(Fenchel's Duality Theorem* [18]*). Let $f(\cdot)$ be a proper convex function on $\mathbb{R}^n$, and let $g(\cdot)$ be a proper concave function on $\mathbb{R}^n$. One has*

$$\inf_x \{f(x) - g(x)\} = \sup_{x^*} \{g_*(x^*) - f^*(x^*)\},$$

if either of the following conditions is satisfied:

1. ri(dom f) $\cap$ ri(dom g) $\neq \emptyset$;

2. *$f(\cdot)$ is lsc, $g(\cdot)$ upper semi-continuous (usc), and* $\mathrm{ri}(\mathrm{dom}\ g_*) \cap \mathrm{ri}(\mathrm{dom}\ f^*) \neq \emptyset$.

Moreover, Under (1) the supremum is attained at some x^*, while under (2) the infimum is attained at some x; if (1) and (2) both hold, the infimum and supremum are necessarily finite.

THEOREM 4: *(Rockafellar* [17]*). Let $f(\cdot)$ ($g(\cdot)$) be a proper convex (concave) function on $\mathbb{R}^n$. Suppose Fenchel's inequality holds as an equality. Then, $\bar{x}$ is a point where $f(\cdot) - g(\cdot)$ achieves its minimum, if and only if $\partial f(\bar{x})$ and $\partial g(\bar{x})$ have some $\bar{x}^*$ in common. Moreover, such vectors $\bar{x}^*$ are then precisely the points where $g_*(\cdot) - f^*(\cdot)$ achieves its maximum.*

The conjugacy leads to an important relationship between the primal optimal value and the dual objective function. In order to study this, it is convenient to introduce:

PROPOSITION 1: (Laurent, Sect. 7.2 in [13]). *Let $v(\cdot)$ denote the perturbation function associated with the optimization problem (P)*

$$v(y) \doteq \inf_{\left\{x \in X^0 | g(x) \leq y\right\}} \left\{f(x)\right\}, \qquad \forall y \in \mathbb{R}^m,$$

and let $F(\cdot)$ be objective function of the dual problem (D). Then, for all $\lambda \in \mathbb{R}^m_+$ we have:

$$F(\lambda) = -v^*(-\lambda).$$

Another important concept to augmented Lagrangians methods is:

DEFINITION 2: We call a function $P : \mathbb{R}^m \to (-\infty, \infty]$ a penalty function if it is convex, proper and non-decreasing.

And again, conjugacy leads to interesting bridges between primal and dual problems.

PROPOSITION 2: (Bertsekas, Sect. 5.4.5 in [3]). *Let $P(\cdot)$ be a penalty function, $v(\cdot)$ be the perturbation function and $F(\cdot)$ be the dual objective function. If the relative interiors of the effective domains of $v(\cdot)$ and $P(\cdot)$ intersect or the relative interiors of the effective domains of $F(\cdot)$ and $P^*(\cdot)$ intersect, then the following equalities are valid:*

$$\inf_{x \in X^0} \left\{f(x) + P(g(x))\right\} = \inf_{y \in \mathbb{R}^m} \left\{v(y) + P(y)\right\} = \sup_{\lambda \geq 0} \left\{F(\lambda) - P^*(\lambda)\right\}.$$

Since the "monotonicity" property is essential to penalty functions, we shall characterize this property in a convenient way to our presentation.

PROPOSITION 3: *Let $h(\cdot)$ be a lower semi-continuous proper convex function. $h(\cdot)$ is non-decreasing if and only if its conjugate $h^*(\cdot)$ is such that:*

$$h^*(\cdot) = h^*(\cdot) + \delta(\cdot \mid \mathbb{R}^n_+),$$

i.e., dom $h^* \subset \mathbb{R}^n_+$.

Proof: This is a very simple result whose proof we present only because we could not find any reference to it.

1. Suppose that $h(\cdot)$ is non-decreasing.

Take any $y \in \mathbb{R}^n$ such that $y_j < 0$. Let $x \in \text{dom } h$ and let e^j denote the j-th vector of the canonical basis, then

$$\begin{aligned} \forall \alpha \geqq 0, \ h^*(y) &\geqq \langle x - \alpha e^j, y \rangle - h(x - \alpha e^j) \\ &\geqq \langle x, y \rangle + \alpha |y_j| - h(x) \Rightarrow \\ h^*(y) &= \infty. \end{aligned}$$

2. Suppose that $h^*(\cdot) \equiv +\infty$ out of the positive orthant, then

$$\begin{aligned} h(x) = h^{**}(x) &= \sup_y \{\langle y, x \rangle - h^*(y) - \delta(y \mid \mathbb{R}^n_+)\} \\ &= \sup_{y \geqq 0} \{\langle y, x \rangle - h^*(y)\}. \end{aligned}$$

Now, given $a \geqq b$, for all $y \geqq 0$ we have

$$\begin{gathered} \langle y, a \rangle \geqq \langle y, b \rangle \Rightarrow \\ \langle y, a \rangle - h^*(y) \geqq \langle y, b \rangle - h^*(y) \Rightarrow \\ h(a) \geqq h(b). \end{gathered}$$

□

4.2. Relationship between augmented Lagrangians and the ϕ-PPM.

If we apply the ϕ-PPM to the dual, we get the following iteration step:

$$\begin{aligned} \lambda^{i+1} &= \underset{\lambda \geqq 0}{\text{argmax}} \{F(\lambda) - \alpha_i \phi(\lambda - \lambda^i)\} \\ &= \underset{\lambda (\geqq 0)}{\text{argmax}} \{F(\lambda) - \alpha_i \phi(\lambda - \lambda^i) - \delta(\lambda \mid \mathbb{R}^m_+)\}. \end{aligned}$$

In view of the above expression, it is reasonable to use Proposition 2 to derive a primal step which would help to compute the iteration of the proximal method. The penalty function that is natural to consider is:

$$P(y, \lambda, \alpha) \doteq \left(\alpha\phi(\cdot - \lambda) + \delta(\cdot \mid \mathbb{R}^m_+)\right)^*(y).$$

Note that $P(\cdot, \lambda, \alpha)$ is the convex conjugate of a function that is proper, convex and inf-compact, and hence $0 \in \operatorname{int dom} P(\cdot, \lambda, \alpha)$ (see [7], Sect. 1.3.10), therefore we may apply Proposition 2. Moreover, this penalty is differentiable, since it is the convex conjugate of a *strictly* convex function (see [7], Th. 4.1.1).

We conclude that:

$$\begin{aligned}\inf_{x \in X^0} \{f(x) + P(g(x), \lambda^i, \alpha_i)\} &= \inf_{y \in \mathbb{R}^m} \{v(y) + P(y, \lambda^i, \alpha_i)\} \\ &= \sup_{\lambda \geqq 0} \{F(\lambda) - P^*(\lambda, \lambda^i, \alpha_i)\}.\end{aligned}$$

Let $\bar{x}$ be a point where the infimum on the penalized primal problem attained. It is clear that $f(\bar{x}) = v(g(\bar{x}))$. Then, if we call $\bar{u} \doteq g(\bar{x})$, we get the equality:

$$\begin{aligned}v(\bar{u}) + P(\bar{u}, \lambda^i, \alpha_i) &= \inf_{y \in \mathbb{R}^m} \{v(y) + P(y, \lambda^i, \alpha_i)\} \\ &= \sup_{\lambda \geqq 0} \{F(\lambda) - P^*(\lambda, \lambda^i, \alpha_i)\}.\end{aligned}$$

From Theorem 4 we know that

$$\lambda^{i+1} = \operatorname*{argmax}_{\lambda \geqq 0} \{F(\lambda) - P^*(\lambda, \lambda^i, \alpha_i)\} \subset \partial_y P(\bar{u}, \lambda^i, \alpha_i) = \{\nabla_y P(\bar{u}, \lambda^i, \alpha_i)\}.$$

Based on these ideas we can formulate the following primal-dual method:

Generalized augmented Lagrangian

1. *Initialization: Take any* λ^1 *dual feasible and* $\widetilde{\alpha}$ *a positive real number.*
2. *Iteration:*
   - *Chose* $0 < \alpha_i \leq \widetilde{\alpha}$ *and calculate an optimal solution* x^i *of the problem:*

$$\begin{cases} \min & f(x) + P(g(x), \lambda^i, \alpha_i) \\ s.t. & x \in X^0 \end{cases}$$

where

$$P(y, \lambda^i, \alpha_i) := \left(\alpha_i\phi(\cdot - \lambda^i) + \delta(\cdot \mid \mathbb{R}^m_+)\right)^*(y).$$

- *Let* $\lambda^{i+1} \doteq \nabla_y P(g(x^i), \lambda^i, \alpha_i)$.
- *Stop if* $\lambda^{i+1} = \lambda^i$ ([3]).

Now, we can formulate the main result of this section:

THEOREM 5: *If the penalized primal problems have optimal solutions, then the proximal point method applied to the dual problem is equivalent to the generalized augmented Lagrangian method. This implies that every limit point of the generated sequence,* $\{\lambda^i\}_{i\in\mathbb{N}}$, *is a solution to the dual problem.*

Proof: It is a direct consequence of the arguments presented before the algorithm and the convergence of the ϕ-PPM (Th. 1). □

The above results lead us to ask if the generated sequence, $\{\lambda^i\}_{i\in\mathbb{N}}$, has any accumulation point. The next theorem is well known and guarantees the existence of accumulation points.

THEOREM 6: *The level sets of the dual objective function are compact if, and only if Slater's constraint qualification holds.*

Proof: For a proof see [5]. □

The above results correspond to convergence theorems for the dual problem, but our main objective is to solve the primal. Hence, we need to show that any accumulation point of the unconstrained minima calculated in the augmented Lagrangian method, $\{x^i\}_{i\in\mathbb{N}}$, is a solution to the original problem (P).

4.3. Primal sequence

We first present a condition that guarantee the existence of the *primal sequence*:

PROPOSITION 4: *Suppose that* $\operatorname{dom}\phi = \mathbb{R}^m$. *If the solution set of (P) is non-empty and bounded, then the level sets of the functions* $f(\cdot) + P(g(\cdot), \lambda, \alpha)$ *are compact for all* $\lambda \geqq 0$ *and* $\alpha > 0$. *This implies that the primal step in the generalized augmented Lagrangian is well defined.*

Proof: This proof is a generalization of the proof of Proposition 5.7 in [2], considering here non-separable functions. It is presented in the appendix. □

([3]) In this case λ^i is a dual solution, as it would be a fixed point of the ϕ-PPM.

Finally, we show the main convergence theorem for the primal sequence computed by the generalized augmented Lagrangian:

THEOREM 7: *Let* $\{(x^i, \lambda^i)\}_{i\in\mathbb{N}}$ *be a sequence generated by the generalized augmented Lagrangian method. If the primal problem satisfies Slater's constraint qualification, then*

$$\limsup_{i\to\infty} g(x^i) \leqq 0,$$
$$\lim_{i\to\infty} \langle \lambda^{i+1}, g(x^i)\rangle = 0,$$

and $f(x^i)$ *converges to the optimal value of the primal problem,* $v(0)$. *Therefore, any accumulation point of the primal sequence is a solution to the primal problem.*

Proof: The boundedness of the dual sequence and Theorem 1 imply that $\lambda^{i+1} - \lambda^i \to 0$. Hence, $\forall \gamma_\phi^i \in \partial\phi(\lambda^{i+1} - \lambda^i)$, $\gamma_\phi^i \to 0$.

Using the algorithm, we know that:

$$\lambda^{i+1} = \nabla_y P(g(x^i), \lambda^i, \alpha_i) \Leftrightarrow g(x^i) \in \partial\big(\alpha_i \phi(\lambda^{i+1} - \lambda^i) + \delta(\lambda^{i+1} \mid \mathbb{R}_+^m)\big).$$

This is equivalent to the existence of subgradients $\gamma_\phi^i \in \partial\phi(\lambda^{i+1} - \lambda^i)$ and $\gamma_\delta^i \in \partial\delta(\lambda^{i+1} \mid \mathbb{R}_+^m)$, such that

$$g(x^i) = \alpha_i \gamma_\phi^i + \gamma_\delta^i. \tag{5}$$

Since $\gamma_\delta^i \leqq 0$, we have:

$$\limsup_{i\to\infty} g(x^i) \leqq \lim_{i\to\infty} \alpha_i \gamma_\phi^i = 0.$$

Moreover, the definition of γ_δ^i implies that $\langle \gamma_\delta^i, \lambda^{i+1}\rangle = 0$, and then

$$\lim_{i\to\infty} \langle \lambda^{i+1}, g(x^i)\rangle = \lim_{i\to\infty} \langle \lambda^{i+1}, \alpha_i \gamma_\phi^i\rangle = 0,$$

where the last equality follows from the boundedness of $\{\lambda^i\}_{i\in\mathbb{N}}$ and $\{\alpha_i^i\}_{i\in\mathbb{N}}$, and $\gamma_\phi^i \to 0$.

Finally, we prove that $f(x^i) \to v(0)$. The definition of λ^{i+1}, implies that x^i is the minimum of the ordinary Lagrangian with λ^{i+1} as multiplier, hence

$$F(\lambda^{i+1}) = f(x^i) + \langle \lambda^{i+1}, g(x^i)\rangle. \tag{6}$$

Using Corollary 1 and the strong duality theorem, we conclude that:

$$v(0) = \lim_{i\to\infty} F(\lambda^{i+1}) = \lim_{i\to\infty} f(x^i) + 0.$$

□

We end this section by presenting a condition to the boundedness of $\{x^i\}_{i\in\mathbb{N}}$.

PROPOSITION 5: *Suppose that the primal problem satisfies Slater's constraint qualification and let $\{x^i\}_{i\in\mathbb{N}}$ be a sequence computed by the generalized augmented Lagrangian method. If the solution set of (P) is non-empty and bounded, then $\{x^i\}_{i\in\mathbb{N}}$ is bounded.*

Proof: This proof is a generalization of the proof of Proposition 5.10 in [2], considering here non-separable functions. It is presented in the appendix. □

5. FINAL REMARKS

The results presented here can be viewed as twofold: the presentation of a new class of proximal point methods (including a strong result on imprecise inner loops) and the corresponding introduction of a new class of augmented Lagrangians.

5.1. The ϕ-PPM

It should be noted that the classical proximal point method [20] is equivalent to ϕ-PPM with the choice $\phi(x) = \|x\|_2^2$. In this case, Féjer monotonicity of the sequence generated by the algorithm with respect to minimizers set has been shown in [10, 12]. A trivial exercise is to show that Féjer monotonicity also holds for $\phi(x) = \theta(\|x\|_2)$ where $\theta : \mathbb{R} \to \mathbb{R}$ is differentiable convex function, strictly increasing on $\mathbb{R}_+$. This is a direct consequence of a chain rule for subdifferentials in [7] (Th. 3.6.1). For these functions, the convergence of the sequence generated by the algorithm is guaranteed if the solution set is non-empty.

For more general regularization functions, Theorem 1 states that if there are accumulation points, all of them are solutions to the minimization problem, as usual in Nonlinear Programming algorithms.

An interesting point is that the result of Theorem 1 holds even if we compute the proximal step inexactly, with a non-vanishing precision, as stated in Theorem 2.

5.2. Augmented Lagrangians

The class of augmented Lagrangian methods presented here can be seen as a generalization of the P_E^+ class described in Bertsekas [2]. The main difference is that we do not impose coerciveness, *i.e.* we do not consider that the derivative of the penalty, $\nabla_y P(y, \bar{\lambda}, \bar{\alpha})$, must go to infinity as $\|y\| \to \infty$. This is true as we can use a regularization whose effective domain is not $\mathbb{R}^n$. Typically, this allows for penalties that are asymptotically affine or even affine away from the origin, allowing for a potential bridge between the penalty functions here presented and exact penalties (4). Examples of the penalties are presented in Figure 1.

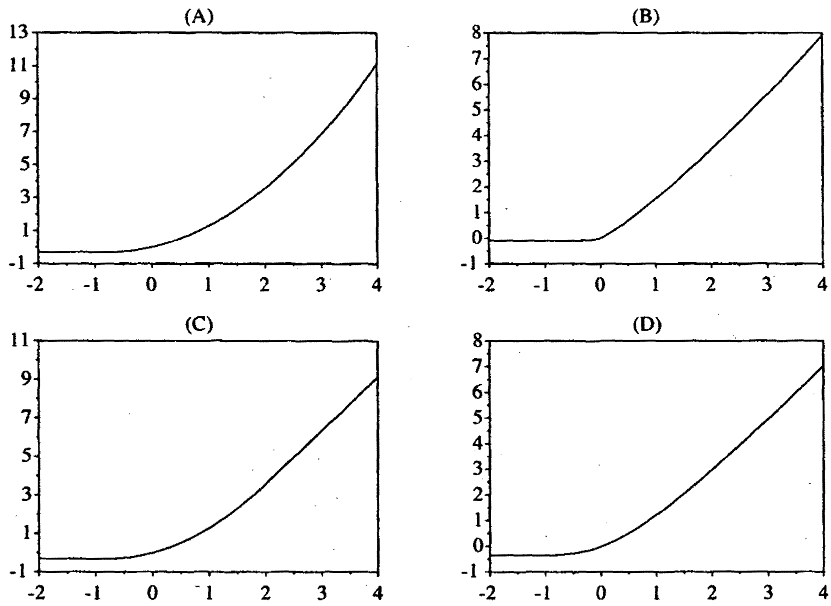

Figure 1. – Type of penalties: (A) the classical penalty, used in the standard augmented Lagrangian. (B) The penalty $P(x, \lambda) = x\lambda + \frac{3}{4}x^{\frac{4}{3}}$, derived from $\phi(x) = \frac{1}{4}x^4$. (C) The regularization function used is the classical $\phi(x) = \frac{1}{2}x^2$ only in a interval containing the origin. Outside this interval the function is considered equal to $+\infty$. This makes the penalty affine above a threshold point (2, in the example). (D) This penalty was generated using the regularization $\phi(x) = -\ln(\cos(x))$. Since $\phi'(x) = \tan(x)$, the penalty, will have bounded slope and asymptotically converges to an affine function.

Another strong point in the above results is that the optimality of the primal accumulation points was easy to obtain, which is not the case when we consider augmented Lagrangians associated with proximal point methods based on Bregman distances or φ-divergences. This positive aspect can

(4) See Bertsekas [2], Chapter 4.

be seen as a consequence of the fact that the ϕ-PPM keeps the shape of the regularization instead of changing it to impose coerciveness near the boundary of $\mathbb{R}^m_+$.

Another difference between the P_E^+ class and the penalties presented here is that separability is not imposed, although it is natural to use separable penalties due to computational considerations.

A drawback of the augmented Lagrangian methods we introduce, is that the penalty functions are not twice differentiable on $\mathbb{R}^n$, due to the addition of $\delta(\cdot \mid \mathbb{R}^n_+)$ to the regularization term that generates such functions.

APPENDIX A. EXISTENCE AND BOUNDEDNESS OF THE PRIMAL ITERATES

We present the proof of Propositions 4 and 5, based on the concepts of recession directions and recession functions as presented in [18], Chapters 8 and 9, whose notation we preserve. Initially we introduce the notation:

DEFINITION 3: For any $\alpha > 0$ and $\lambda \geqq 0$, we will use the following notation:

$$\forall x \in \mathbb{R}^n, \quad p_{\lambda,\alpha}(x) \doteq P\big(g(x), \lambda, \alpha\big),$$

where $P(\cdot, \lambda, \alpha)$ is defined in Section 4.2.

LEMMA 2: *Suppose that* $\operatorname{dom}\phi = \mathbb{R}^m$ *and that* $\exists \bar{x},\ g(\bar{x}) \leqq 0$. *Let* $\alpha > 0$ *and* $\lambda \geqq 0$. *Let* $\mathcal{R} \subset \mathbb{R}^n$ *denote the intersection of the sets of all directions of recession of all the components of the constraint function* $g(\cdot)$. *Then,*

$$p_{\lambda,\alpha}0^+(h) = \begin{cases} 0, & \text{if } h \in \mathcal{R}; \\ \infty, & \text{if } h \notin \mathcal{R}. \end{cases}$$

Proof: Given any proper closed convex function, $f(\cdot)$, we have by definition:

$$f0^+(h) = \lim_{t\to\infty} \frac{f(x+th) - f(x)}{t} \tag{7}$$

for any $x \in \operatorname{dom} f$ and any $h \in \mathbb{R}^n$.

Note that, from the definition of the penalty, it is easy to show that $p_{\lambda,\alpha}(y) \leqq 0$ for all $y \leqq 0, \lambda \geqq 0, \alpha > 0$. It follows that $p_{\lambda,\alpha}(\bar{x}) \leqq 0$.

1. $h \in \mathcal{R}$

From the definition of the penalty we have:

$$p_{\lambda,\alpha}(\cdot) \geqq -\alpha\phi(-\lambda). \tag{8}$$

And, since $h \in \mathcal{R}$:

$$g(\bar{x} + th) \leqq g(\bar{x}), \ \forall t \geqq 0.$$

Using the last two inequalities and the fact that the penalty is non-decreasing we conclude that:

$$-\alpha\phi(-\lambda) - p_{\lambda,\alpha}(\bar{x}) \leqq p_{\lambda,\alpha}(\bar{x} + th) - p_{\lambda,\alpha}(\bar{x}) \leqq 0, \ \forall t \geqq 0.$$

It follows (from 7) that:

$$p_{\lambda,\alpha}0^+(h) = 0.$$

2. $h \notin \mathcal{R}$.
   Without loss of generality we will suppose that h is not a direction of recession of $g_1(\cdot)$.

$$g_1 0^+(h) > \delta > 0.$$

Let $e^1 = (1, 0, \ldots, 0)^t \in \mathbb{R}^m$. Then,

$$\begin{aligned} & p_{\lambda,\alpha}(\bar{x} + th) - p_{\lambda,\alpha}(\bar{x}) \\ & \quad = \sup_{z \geqq 0} \{\langle g(\bar{x} + th), z\rangle - \alpha\phi(z - \lambda)\} - p_{\lambda,\alpha}(\bar{x}) \\ & \quad \geqq \sup_{\gamma \geqq 0} \{\langle g(\bar{x} + th), \lambda + \gamma e^1\rangle - \alpha\phi(\gamma e^1)\} - p_{\lambda,\alpha}(\bar{x}) \\ & \quad = \langle g(\bar{x} + th), \lambda\rangle + \sup_{\gamma \geqq 0} \{\gamma g_1(\bar{x} + th) - \alpha\phi(\gamma e^1)\} - p_{\lambda,\alpha}(\bar{x}). \end{aligned}$$

Dividing both sides by t and taking limits for $t \to \infty$ we have:

$$p_{\lambda,\alpha}0^+(h) \geqq \langle g0^+(h), \lambda\rangle + \lim_{t\to\infty} \sup_{\gamma \geqq 0} \left\{ \frac{\gamma g_1(\bar{x} + th) - \alpha\phi(\gamma e^1)}{t} \right\} - 0.$$

As $g_1 0^+(h) > \delta$, this can be further simplified to:

$$p_{\lambda,\alpha}0^+(h) \geqq \langle g0^+(h), \lambda\rangle + \lim_{t\to\infty} \sup_{\gamma \geqq 0} \left\{ \gamma\delta - \frac{\alpha\phi(\gamma e^1)}{t} \right\}.$$

As the recession function of a proper convex function is proper, we only need to prove that the limit above is ∞.

To prove this assertion, given an $M > 0$, let t_M be big enough such that $\frac{\alpha\phi(\frac{2M}{\delta}e^1)}{t_M} < M$. It follows that for all $t \geq t_M$:

$$\sup_{\gamma \geqq 0} \left\{ \gamma\delta - \frac{\alpha\phi(\gamma e^1)}{t} \right\} \geqq \frac{2M}{\delta}\delta - M \geqq M.$$

□

PROPOSITION 4: *Suppose that* $\operatorname{dom}\phi = \mathbb{R}^m$. *If the solution set of (P) is non-empty and bounded, then the level sets of the functions* $f(\cdot) + P(\cdot, \lambda, \alpha)$ *are compact for all* $\lambda \geqq 0$ *and* $\alpha > 0$. *This implies that the primal step in the generalized augmented Lagrangian is well defined.*

Proof: From the previous lemma and Theorem 9.3 in [18] we have that:

$$\big(f(\cdot) + P(g(\cdot), \lambda, \alpha)\big)0^+(h) = \begin{cases} f0^+(h), & \text{if } h \in \mathcal{R}; \\ \infty, & \text{if } h \notin \mathcal{R}. \end{cases}$$

Remembering that the assumption that the solution set is compact implies that $f0^+(h) > 0$ whenever $h \in \mathcal{R}$, we get the desired result. □

Note that the assumption $\operatorname{dom}\phi = \mathbb{R}^m$ seems to be essential. It is already present in [2] (p. 306) in the form of items (f), (g) and (h) in the description of the P_I class. More recently, in [1], this assumption corresponds to recession properties of the penalties θ and to the choice of the parameter $\alpha(r)$.

As an example, consider the problem

$$\begin{cases} \min & x \\ s.t. & -2x \leqq 0 \end{cases}$$

and $\phi(\lambda) = -\sqrt{1 - \lambda^2}, \operatorname{dom}\phi = [-1, 1]$; it is possible to show that

$$f(x) + P(g(x), 2, 1) = -3x + (4x^2 + 1)\sqrt{\frac{1}{4x^2 + 1}},$$

which is a function that does not have bounded level sets. This function goes to $-\infty$ for $x \to \infty$, and hence the penalized problem does not have a solution. In this case, the penalty makes feasibility more important than optimality. This behavior would be reversed if the objective function was $7x$ instead of x.

Finally, we present a result that shows that the primal sequence, $\{x^i\}_{i \in \mathbb{N}}$, is bounded under standard assumptions.

PROPOSITION 5: *Suppose that the primal problem satisfies Slater's constraint qualification and let* $\{x^i\}_{i\in\mathbb{N}}$ *be a sequence computed by the generalized augmented Lagrangian method. If the solution set of (P) is non-empty and bounded, then* $\{x^i\}_{i\in\mathbb{N}}$ *is bounded.*

Proof: From Theorem 7 we know that $g(x^i)$ is bounded from above. Therefore, $\{x^i\}_{i\in\mathbb{N}}$ is contained in an intersection of level sets of the all constraints.

Suppose, by contradiction, that $\{x^i\}_{i\in\mathbb{N}}$ is unbounded. Since the solution set of (P) is non-empty and bounded, it follows that if $f(x^i)\to\infty$, since, otherwise, at least one of the accumulation points of $\frac{x^i}{\|x^i\|}$ would be a common direction of recession the objective function and the constraints. But this is a contradiction with $f(x^i)\to v(0)\in\mathbb{R}$, proved in Theorem 7. □

ACKNOWLEDGEMENTS

Before concluding, we must register our acknowledgment to the referees, whose comments and suggestions definitively improved the presentation of this article.

REFERENCES

1. 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.
2. D. BERTSEKAS, *Constrained Optimization and Lagrange Multipliers*. Academic Press, New York (1982).
3. D. BERTSEKAS, *Nonlinear Programming*. Athena Scientific (1995).
4. Y. CENSOR and J. ZENIOS, The proximal minimization algorithms with D-functions. *J. Optim. Theory Appl.* **73** (1992) 451–464.
5. J. GAUVIN, A necessary and sufficient condition to have bounded multipliers in nonconvex programming. *Math. Programming* **12** (1977) 136–138.
6. A.M. GEOFFRION, Duality in non-linear programming, a simplified applications-oriented developtment. *SIAM Rev.* **13** (1971) 65–101.
7. J.-B. HIRIART-URRUTY and C. LEMARÉCHAL, Convex analysis and minimization algorithms. II. Advanced theory and bundle methods. Springer-Verlag, Berlin (1993).
8. C. HUMES and P. SILVA, *An inexact classical proximal point algorithm viewed as a descent method in the optimization case*. Technical Report RT-MAC 99-10. Instituto de Matemática e Estatística - USP (1999).
9. C. HUMES Jr., Some comments on Lagrangian duality, optimality conditions and convexity. *Investigación Oper.* **2** (1991) 159–169.
10. A. IUSEM, Métodos de Ponto Proximal em Otimização. *Instituto de Matemática Pura e Aplicada - CNPq* (1995). Book from the 20° Colóquio Brasileiro de Matemática.
11. A. IUSEM and M. TEBOULLE, On the convergence rate of entropic proximal minimization algorithms. *Comput. Appl. Math.* **12** (1993) 153–168.

12. A. Iusem, M. Teboulle and B. Svaiter, Entropy-like proximal methods in covex programming. *Math. Oper. Res.* **19** (1994) 790–814.
13. P.-J. Laurent, *Approximation et Optimisation. Collection Enseignement des Sciences.* Hermann (1972).
14. B. Martinet, Regularisation d'inequations variationelles par approximations successives. *Rev. Française Inf. Rech. Oper.* (1970) 154–159.
15. B. Martinet, Détermination approché d'un point fixe d'une application pseudo-contractante. *C.R. Acad. Sci. Paris* **274A** (1972) 163–165.
16. J. Moreau, Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France* **93** (1965) 273–299.
17. R.T. Rockafellar, Extension of Fenchel's duality theorem for convex functions. *Duke Math. J.* **33** (1966) 81–89.
18. R.T. Rockafellar, *Convex Analysis.* Princeton University Press (1970).
19. R.T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* **1** (1976) 97–116.
20. R.T. Rockafellar, Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14** (1976) 887–898.
21. R.T. Rockafellar and R.J.-B. Wets, *Variational analysis [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, *Grundlehren Math. Wiss.* **317** (1998).
22. M. Solodov and B. Svaiter, A hybrid projection-proximal point algorithm. *J. Convex Anal.* **6** (1999).
23. M. Solodov and B. Svaiter, An inexact hybrid extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis* (to appear).
24. M. Teboulle, Entropic proximal methods with aplications to nonlinear programming. *Math. Oper. Res.* **17** (1992) 670–690.
25. J. Tind and L.A. Wolsey, An elementary survey of general duality theory in mathematical programming. *Math. Programming* **20** (1981) 241–261.