

REGINA S. BURACHIK

ALFREDO N. IUSEM

**A generalized proximal point algorithm for the  
nonlinear complementarity problem**

*RAIRO. Recherche opérationnelle*, tome 33, n° 4 (1999),  
p. 447-479

[http://www.numdam.org/item?id=RO\\_1999\\_\\_33\\_4\\_447\\_0](http://www.numdam.org/item?id=RO_1999__33_4_447_0)

© AFCET, 1999, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Recherche opérationnelle » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

## A GENERALIZED PROXIMAL POINT ALGORITHM FOR THE NONLINEAR COMPLEMENTARITY PROBLEM (\*)

by Regina S. BURACHIK (<sup>1,\*\*</sup>) and Alfredo N. IUSEM (<sup>2,\*\*\*</sup>)

Communicated by Jean-Pierre CROUZEIX

---

Abstract. – We consider a generalized proximal point method (GPPA) for solving the nonlinear complementarity problem with monotone operators in  $\mathbf{R}^n$ . It differs from the classical proximal point method discussed by Rockafellar for the problem of finding zeroes of monotone operators in the use of generalized distances, called  $\varphi$ -divergences, instead of the Euclidean one. These distances play not only a regularization role but also a penalization one, forcing the sequence generated by the method to remain in the interior of the feasible set, so that the method behaves like an interior point one. Under appropriate assumptions on the  $\varphi$ -divergence and the monotone operator we prove that the sequence converges if and only if the problem has solutions, in which case the limit is a solution. If the problem does not have solutions, then the sequence is unbounded. We extend previous results for the proximal point method concerning convex optimization problems.

Keywords: Nonlinear complementarity problem, proximal point methods, monotone operators.

### 1. INTRODUCTION

In this paper we are concerned with proximal algorithms for solving the nonlinear complementarity problem in  $\mathbf{R}^n$ . We start with some preliminaries. The operator  $T : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$  is *monotone on a subset*  $C \subset \mathbf{R}^n$  if

$$\langle u - v, x - y \rangle \geq 0,$$

for all  $x, y \in C$  and all  $u \in T(x), v \in T(y)$ . A monotone operator is called *maximal* if for any other monotone operator  $\tilde{T}$  with  $\tilde{T}(x) \supseteq T(x)$  for all  $x \in \mathbf{R}^n$ , it holds that  $\tilde{T} = T$ .

---

(\*) Received August 1996.

(\*\*) Research of this author was supported by CNPq grant No. 301280/94-0.

(\*\*\*) Research of this author was partially supported by CNPq grant No. 301280/86.

(<sup>1</sup>) Universidade Federal de Rio de Janeiro, COPPE/UFRJ, Programa de Engenharia de Sistemas e Computação, Caixa Postal 68511, CEP 21945-970, Rio de Janeiro - RJ - Brazil.

(<sup>2</sup>) Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina, 110, CEP 22460-320, Rio de Janeiro, RJ, Brazil.

1991 AMS Classification numbers: 90C25, 90C30.

Given  $T : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$ , the nonlinear complementarity problem,  $\text{NCP}(T)$ , consists of finding  $z \in \mathbf{R}^n$  such that there exists  $u \in T(z)$  satisfying

$$z \geq 0, \quad u \geq 0, \quad \langle z, u \rangle = 0. \quad (1)$$

An in-depth study of  $\text{NCP}(T)$  can be found in [9]. This problem arises in many situations; the most obvious one is obtained by taking  $T(x) = \nabla f(x)$  with  $f: \mathbf{R}^n \rightarrow \mathbf{R}$  differentiable, in which case (1) are the first order Karush-Kuhn-Tucker conditions of  $\min f(x)$  s.t.  $x \geq 0$ .

Let  $C$  be a closed and convex set. The  $\text{NCP}(T)$  is a particular case of the well-known *Variational Inequality Problem for  $T$  and  $C$* , which we will denote by  $\text{VIP}(T, C)$ . This problem is defined as:

find  $x^* \in C$  such that there exists  $y^* \in T(x^*)$  with

$$\langle y^*, x - x^* \rangle \geq 0 \text{ for all } x \in C.$$

The most important aspects of this problem are studied in [14]. It is easy to see that when  $C = \mathbf{R}_+^n$ ,  $\text{VIP}(T, C)$  becomes  $\text{NCP}(T)$ . The “unconstrained” version of  $\text{VIP}(T, C)$ , *i.e.*, when  $C = \mathbf{R}^n$ , is the problem of finding the zeroes of  $T$ . In fact,  $\text{VIP}(T, C)$  for an arbitrary  $C$  can be seen as the problem of finding the zeroes of an operator, namely  $T + N_C$ , where  $N_C$  is the *normality operator associated to  $C$*  (see Sect. 2). The proximal algorithm, or more exactly, “the proximal point algorithm”, according to Rockafellar’s terminology (PPA, for short, from now on), is basically the successive approximation method for finding zeroes of monotone operators in Hilbert spaces. This method, which is therefore not new [15, 19], seems to have been applied the first time to convex minimization by Martinet (see [17, 18]). The first important results (like approximate versions, linear and finite convergence) in the more general framework of maximal monotone operators are due to Rockafellar [22]. This algorithm is still the object of intensive investigation (see [16] for a modern survey on the method).

The PPA can be seen as a regularization method in which the regularization parameter need not to approach 0, thus avoiding a possible ill behavior of the regularized problems. The PPA for the problem of finding zeroes of  $T$  generates a sequence  $\{x^k\} \subset \mathbf{R}^n$  in the following way: it starts with any  $x^0 \in \mathbf{R}^n$  and, given  $x^k$ ,  $x^{k+1}$  is taken so that

$$0 \in \bar{T}_k(x^{k+1}),$$

where

$$\bar{T}_k(x) = T(x) + \lambda_k(x - x^k), \quad (2)$$

and  $\lambda_k > 0$ .

It is shown in [22], Theorems 1 and 2, that  $\{x^k\}$  converges to a zero of  $T$ , provided that  $\lambda_k$  is bounded away from 0 and the set of zeroes is nonempty. It is also proved that the sequence converges strongly at a linear rate if  $T^{-1}$  is Lipschitz continuous at 0,  $\lambda_k$  is nondecreasing.

As we remarked before, the PPA is used to solve the problem of finding zeroes of monotone operators, *i.e.*, the “unconstrained” variational inequality problem associated with a maximal monotone operator. Now we will exhibit a similar kind of algorithms with all the advantages of the PPA, but suitable to  $NCP(T)$ .

In the classical proximal point method, each subproblem involves a quadratic regularization. Indeed, the second term that appears in iteration (2) is precisely the gradient of the quadratic norm. We consider below generalized proximal algorithms in which this quadratic distance is replaced by a distance-like function adapted to the set  $\mathbf{R}_+^n$ .

Though the operators  $\bar{T}_k$  in (2) are better conditioned in principle than  $T$  (*e.g.*,  $\bar{T}_k$  has a unique zero when  $T$  could have several or none), the subproblems are structurally as hard to solve as the original problem. In this paper, we consider a generalized proximal point algorithm, for  $NCP(T)$ , which generates subproblems which are structurally simpler than the original problem, as we show below.

In the algorithm considered in this paper, the subproblems are of the form  $0 \in T_k(x^{k+1})$ , but in this case  $T_k(x) = T(x) + \lambda_k \nabla_x d_\varphi(x, x^k)$ , where  $d_\varphi$  is a  $\varphi$ -divergence. This means, basically, that  $d_\varphi(\cdot, x^k)$  is a strictly convex function defined on  $\mathbf{R}_+^n$  whose gradient diverges at the boundary of  $\mathbf{R}_+^n$ . As a consequence,  $T_k$  has always a unique zero, and it lies in the interior of  $\mathbf{R}_+^n$ . So that the subproblems are genuinely unconstrained. We make this point clearer with the following example.

Take  $T$  point-to-point and  $d_\varphi$  as in Example 1 of Section 2. We will prove that under suitable assumptions, our algorithm solves the  $NCP(T)$ . The PPA applied to this problem generates a subproblem of the form

$$x^t(T(x) + \lambda_k(x - x^k)) = 0, \quad x \geq 0, \quad T(x) + \lambda_k x \geq \lambda_k x^k,$$

while our scheme will reduce to the following system of nonlinear equations

$$T(x)_j + \lambda_k \log x_j = \lambda_k \log x_j^k, \quad j = 1, \dots, n.$$

The difference between PPA and the algorithm we propose in this example is clear. The PPA subproblems are  $NCP$ 's of the same nature as the original

problem, considerably harder to solve, from a computational point of view, than the system above. For instance, if  $T$  is continuously differentiable, the mentioned system can be easily solved with Newton method, while the PPA algorithm has the additional combinatorial complication of determining the set of zero components of  $x$ .

The distance-like functions we will consider are called  $\varphi$ -divergences. These  $\varphi$ -divergences are “distances” adequate to the positive orthant. They are defined using a strictly convex function  $\varphi$  that satisfies the conditions

$$\varphi(1) = \varphi'(1) = 0, \text{ and } \varphi''(1) > 0.$$

The distance-like function obtained by this  $\varphi$  is given by the formula

$$d_\varphi(x, y) = \sum_{j=1}^n y_j \varphi\left(\frac{x_j}{y_j}\right). \quad (3)$$

These distance-like functions, called  $\varphi$ -divergences, provide a regularization term that penalizes the proximity to the boundary of  $\mathbf{R}_+^n$ , forcing the sequence  $\{x^k\}$  to be in the interior of  $\mathbf{R}_+^n$  and making the subproblems unconstrained. Another kind of regularization, suitable for the  $\text{VIP}(T, C)$ , is the one obtained using the so called *Bregman distances* (introduced in [2]) instead of the quadratic distance in (2). In [6] it is proved welldefinedness and convergence of the sequence for the  $\text{VIP}(T, C)$  with this special kind of regularization, adapted to any convex and closed set  $C \subset H$ , for  $H$  an arbitrary Hilbert space. In summary, the proximal algorithms we will study replace the quadratic distance by other distance-like function whose properties are chosen so that they behave with respect to the feasible set in an analogous way as the norm behaves in  $\mathbf{R}^n$ .

Generalizing the scheme (2) of the classical proximal point method, we define the proximal point method with  $\varphi$ -divergences as

**Initialization:** Take  $x^0$  such that

$$x^0 > 0. \quad (4)$$

**Iterative step:** Given  $x^k \in \mathbf{R}_{++}^n$ , if  $x^k$  solves  $\text{NCP}(T)$ , stop. Otherwise, find  $x^{k+1} \in \mathbf{R}_{++}^n$  such that

$$0 \in Tx^{k+1} + \lambda_k \nabla_x d_\varphi(x^{k+1}, x^k), \quad (5)$$

where  $\lambda_k$  is a sequence of positive numbers and  $\nabla_x$  denotes the gradient with respect to the first argument. In Section 5 we will prove that under adequate assumptions (5) uniquely determines  $x^{k+1}$ .

Observe that we have replaced the regularization term in (2) (which is, as we mentioned before, the gradient of the quadratic norm), by  $\nabla_x d_\varphi(x, x^k)$ . The properties of this distance imply that all iterates are in the interior of  $\mathbf{R}_+^n$ .

Convergence results are available for this algorithm applied to the convex optimization problem, *i.e.*, when  $T = \partial f$ , with  $f$  convex and closed. This analysis is rather involved and can be found in [10] and [11]. In these works it is proved that, under reasonable hypothesis on  $\varphi$ , the sequence generated by the algorithm converges to a minimizer of  $f$  on  $\mathbf{R}_+^n$ , as long as the set of minimizers is not empty. Further results on proximal-like methods for convex optimization can be found in [25].

For the general  $\text{NCP}(T)$ , Auslender and Haddou proved in [1] convergence of the sequence for a specific  $\varphi$ -divergence, namely, the Kullback-Liebler divergence, defined as

$$d_\varphi(x, y) = \sum_{j=1}^n x_j \log\left(\frac{x_j}{y_j}\right) + y_j - x_j,$$

corresponding to  $\varphi(t) = t \log t - t + 1$  in (3). Later on, we will describe more carefully these results.

For an operator  $T$ , we define

$$\begin{aligned} S^* &:= \{x \mid x \text{ solves } \text{NCP}(T)\}, \\ D(T) &:= \{x \in \mathbf{R}^n \mid T(x) \neq \emptyset\}, \\ R(T) &:= \cup_{x \in D(T)} T(x). \end{aligned}$$

We will show here that under appropriate assumptions this method generates a sequence that converges to  $x^* \in S^*$  if and only if  $S^* \neq \emptyset$ .

## 2. $\varphi$ -DIVERGENCES

In this section we discuss a special class of distance-like functions, adequate to the positive orthant. They are denoted by  $d_\varphi(\cdot, \cdot)$ , defined on  $\mathbf{R}_{++}^n \times \mathbf{R}_{++}^n$ . Take  $\varphi : \mathbf{R}_{++} \rightarrow \mathbf{R}_+$ , strictly convex, closed and twice continuously differentiable, satisfying

$$\varphi(1) = \varphi'(1) = 0, \text{ and } \varphi''(1) > 0.$$

The set of  $\varphi$  satisfying these conditions will be called  $\Phi$ .

DEFINITION 1: If  $\varphi \in \Phi$ , then  $d_\varphi(\cdot, \cdot) : \mathbf{R}_{++}^n \times \mathbf{R}_{++}^n \rightarrow \mathbf{R}_+$  defined by

$$d_\varphi(x, y) = \sum_{j=1}^n y_j \varphi\left(\frac{x_j}{y_j}\right),$$

is said to be a  $\varphi$ -divergence.

The notion of  $\varphi$ -divergence was first introduced in [8] and furtherly developed in [24].

DEFINITION 2:  $d_\varphi$  is said to be boundary coercive iff

$$\lim_{t \rightarrow 0^+} \varphi'(t) = -\infty.$$

DEFINITION 3:  $d_\varphi$  is zone coercive if and only if it is boundary coercive and

$$\lim_{t \rightarrow +\infty} \varphi'(t) = +\infty.$$

Clearly zone coerciveness implies boundary coerciveness. We consider some specific subsets of  $\Phi$ .

DEFINITION 4: Define  $\Phi_3$ ,  $\Phi_4$ , and  $\Phi_5$  as

$$\Phi_3 := \{\varphi \in \Phi \mid \varphi'(t) \leq \varphi''(1) \log t \text{ for all } t > 0\}.$$

$$\Phi_4 := \left\{ \varphi \in \Phi \mid \left(1 - \frac{1}{t}\right) \varphi''(1) \leq \varphi'(t) \leq \varphi''(1) \log t \text{ for all } t > 0 \right\}.$$

$$\Phi_5 := \left\{ \varphi \in \Phi_4 \mid \left(1 - \frac{1}{t}\right) \varphi''(1) < \varphi'(t) \text{ for } t > 0, t \neq 1, \right.$$

$$\left. \lim_{t \rightarrow 0^+} t\varphi'(t) > -\varphi''(1) \text{ and } \lim_{t \rightarrow +\infty} \varphi'(t) > \varphi''(1) \right\}.$$

(We skip subindices 1, 2 for the sake of compatibility with [10].)

The following proposition establishes the basic properties of  $d_\varphi(\cdot, \cdot)$  and its proof can be found in [24].

PROPOSITION 1: Let  $\varphi \in \Phi$ . Then

- (i)  $d_\varphi(x, y) \geq 0$  for all  $x, y \in \mathbf{R}_{++}^n$ ;
- (ii)  $d_\varphi(x, y) = 0$  iff  $x = y$ ;
- (iii) the level sets of  $d_\varphi(\cdot, y)$  and  $d_\varphi(x, \cdot)$  are bounded for all  $x, y \in \mathbf{R}_{++}^n$ , respectively;

(iv)  $d_\varphi(\cdot, \cdot)$  is jointly continuous and convex on  $x, y$  and strictly convex in  $x$ ;

(v)  $\lim_{k \rightarrow \infty} d_\varphi(y, y^k) = 0$  iff  $\lim_{k \rightarrow \infty} y^k = y$ .

In order to describe the subdifferential properties of  $d_\varphi(\cdot, y)$  (with  $y \in \mathbf{R}_{++}^n$ ) we will assume from now that  $d_\varphi(\cdot, y)$  has been extended in the usual way, i.e., taking  $d_\varphi(x, y) = +\infty$  for all  $x \notin \mathbf{R}_{++}^n$ .

For a convex  $f : \mathbf{R}^n \rightarrow \mathbf{R}$ , we denote by  $dom f$  the effective domain of  $f$ . By  $\partial f$  we denote the subdifferential of  $f$  (i.e.  $\partial f(x)$  is the set of subgradients of  $f$  at  $x$ ), and by  $\partial d_\varphi(\cdot, y)$  the subdifferential of  $d_\varphi$  as a function of its first argument, in the same way as  $\nabla d_\varphi(\cdot, y)$  denotes the gradient of  $d_\varphi$  as a function of its first argument. For  $X \subset \mathbf{R}^n$ ,  $\partial X$  will denote the boundary of  $X$ .

The following lemma is a straightforward consequence of [23], Theorem 25.6, and its proof will be omitted here.

LEMMA 1: *If  $\varphi \in \Phi$  is boundary coercive, then the domain of  $\partial d_\varphi(\cdot, y)$  is  $\mathbf{R}_{++}^n$  for all  $y \in \mathbf{R}_{++}^n$ , i.e.,*

$$\partial d_\varphi(x, y) = \begin{cases} \sum_{j=1}^n \varphi' \left( \frac{x_j}{y_j} \right) e^j & \text{if } x \in \mathbf{R}_{++}^n, \\ \emptyset & \text{otherwise,} \end{cases}$$

where  $\{e^j\}$  is the canonical basis of  $\mathbf{R}^n$ .

Let us present now some examples of  $\varphi$ -divergences.

Example 1: Let  $\varphi_1 = t \log t - t + 1$

$$d_{\varphi_1}(x, y) = \sum_{j=1}^n \left( x_j \log \frac{x_j}{y_j} + y_j - x_j \right),$$

i.e.,  $d_{\varphi_1}$  is the so called Kullback-Liebler divergence, which can be extended to  $\mathbf{R}_+^n \times \mathbf{R}_{++}^n$ . It is easy to see that  $\varphi_1 \in \Phi_5$ . The function  $d_{\varphi_1}(x, \cdot)$  is a strictly convex function, well defined for  $x$  in the boundary of  $\mathbf{R}_{++}^n$  and with bounded level sets.

Example 2: Let  $\varphi_2 = t - \log t - 1$ . Then

$$d_{\varphi_2}(x, y) = d_{\varphi_1}(y, x).$$

It is easy to check that  $\varphi_2 \in \Phi_4$  but  $\varphi_2 \notin \Phi_5$ .



*Example 3:* Let  $\varphi_3(t) = (\sqrt{t} - 1)^2$ . Then

$$d_{\varphi_3}(x, y) = \sum_{j=1}^n (\sqrt{x_j} - \sqrt{y_j})^2.$$

It can be easily verified that  $\varphi_3 \in \Phi_5$ .

All these  $\varphi$ -divergences are boundary coercive, but only  $\varphi_1$  is zone coercive.

*Remark 1:* Consider now the two families of  $\varphi$ -divergences  $\gamma_\lambda(\cdot)$  and  $\rho_\lambda(\cdot)$  given by:

$$\gamma_\lambda(t) := \lambda\varphi_1(t) + (1 - \lambda)\varphi_2(t),$$

and

$$\rho_\lambda(t) := \lambda\varphi_1(t) + (1 - \lambda)\varphi_3(t),$$

where  $\lambda \in (0, 1]$  and  $\varphi_1$ ,  $\varphi_2$  and  $\varphi_3$  as above. Elementary calculus shows that all the elements of these families are zone coercive and belong to  $\Phi_5$ .

### 3. PARAMONOTONICITY AND PSEUDOMONOTONICITY

Our convergence theorems require two conditions on the operator  $T$ , namely para- and pseudomonotonicity, which we discuss next. The notion of paramonotonicity was introduced in [7] and further studied in [12]. It is defined as follows.

**DEFINITION 5:**  $T$  is paramonotone on a convex set  $C$  if it is monotone on  $C$  and  $\langle z - z', w - w' \rangle = 0$  with  $z, z' \in C$ ,  $w \in T(z)$ ,  $w' \in T(z')$ , implies  $w \in T(z')$ ,  $w' \in T(z)$ .

The next proposition presents the main properties of paramonotone operators.

**PROPOSITION 2:**

- i) If  $T$  is the subdifferential  $\partial f$  of a convex function  $f : H \rightarrow \mathbf{R}$ , then  $T$  is paramonotone on  $H$ .
- ii) If  $T$  is paramonotone on  $C$ ,  $x^*$  solves  $VIP(T, C)$  and  $x \in C$  satisfies that there exists an element  $\bar{u} \in T(x)$  such that  $\langle \bar{u}, x^* - x \rangle \geq 0$  then  $x$  also solves  $VIP(T, C)$ .
- iii) If  $T_1$  and  $T_2$  are paramonotone on  $C$  then  $T_1 + T_2$  is paramonotone in  $C$ .

*Proof:* (i) and (ii) are Propositions 1 and 2 of [12]. (iii) follows easily from Definition 5.

Other conditions guaranteeing paramonotonicity can be found in [12].

The notion of pseudomonotonicity, as introduced in [5], Definition (7.5), is formulated in a much more general framework than ours. For simplicity of the exposition, however, we restrict ourselves to  $\mathbf{R}^n$ .

DEFINITION 6: *Let  $G$  be a closed and convex subset of  $D(T)$ . An operator  $T : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$  is said to be pseudomonotone from  $G$  to  $\mathcal{P}(\mathbf{R}^n)$  if and only if it satisfies the following condition:*

*Take any sequence  $\{x^k\} \subset G$ , converging to an element  $x \in G$ , and any sequence  $\{w^k\} \subset \mathbf{R}^n$ , with  $w^k \in T(x^k)$  for all  $k$ , such that*

$$\limsup_{k \rightarrow \infty} \langle w^k, x^k - x \rangle \leq 0.$$

*Then for each  $y \in G$  there exists an element  $w \in T(x)$ , such that*

$$\langle w, x - y \rangle \leq \liminf_{k \rightarrow \infty} \langle w^k, x^k - y \rangle.$$

The term pseudomonotonicity is also used for operators such that

$$\forall x, y \in C \quad \langle T(x), y - x \rangle \geq 0 \Rightarrow \langle T(y), y - x \rangle \geq 0$$

(see e.g. [13]). In our framework, the concept of pseudomonotonicity is the one presented in Definition 6.

We recall now the classical definition of upper semicontinuity.

DEFINITION 7: *Let  $S$  and  $S_1$  be two topological spaces and  $T$  a mapping of  $S$  in  $\mathcal{P}(S_1)$ . Then  $T$  is said to be an upper semicontinuous set valued mapping of  $S$  if for each point  $s_0 \in S$  and each open neighborhood  $V$  of  $T(s_0)$  in  $S_1$ , there exists a neighborhood  $U$  of  $s_0$  in  $S$  (with  $U$  depending on  $V$ ), such that  $T(U) \subset V$ .*

We denote by  $\text{int } X$  the topological interior of a subset  $X \subset \mathbf{R}^n$ .

For a closed and convex set  $V \subset \mathbf{R}^n$ , let  $\delta_V$  be the indicator function of  $V$ , i.e.

$$\delta_V(x) = \begin{cases} 0 & \text{if } x \in V \\ +\infty & \text{otherwise.} \end{cases}$$

We define the normality operator  $N_V$  of  $V$  as  $N_V(x) = \partial\delta_V(x)$  (i.e. the subdifferential of  $\delta_V$  at  $x$ ). Indeed,

$$N_V(x) = \begin{cases} \emptyset & \text{if } x \notin V \\ \{w \mid \langle w, y - x \rangle \leq 0 \text{ for any } y \in V\} & \text{otherwise.} \end{cases}$$

As a direct consequence of the definition above, it holds that  $N_V(x) = 0$  for any  $x \in \text{int} V$ .

**PROPOSITION 3:** *With the notation above, it holds that*

- i)  $N_V(x)$  is paramonotone;
- ii)  $N_V(x)$  is pseudomonotone from  $V$  to  $\mathcal{P}(\mathbf{R}^n)$ .

*Proof:* Item (i) is an application of Proposition 2(i) to the convex function  $\delta_V(\cdot)$ . Let us prove (ii). Take  $\{x^k\}$ ,  $\{w^k\}$ ,  $x$  and  $y$  as in Definition 6. Since  $w^k \in N_V(x^k)$  and  $y \in D(N_V) = V$ , it follows from the definition of the normal cone  $N_V(x^k)$  that  $\langle w^k, x^k - y \rangle \geq 0$ . On the other hand, it is obvious from the definition of  $N_V(x)$ , that  $0 \in N_V(x)$ . Take  $w = 0$  and conclude that  $\langle w, x - y \rangle = 0 \leq \liminf_{k \rightarrow \infty} \langle w^k, x^k - y \rangle$ .  $\square$

**PROPOSITION 4:** *If  $T : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$  is maximal monotone and  $\text{int}(D(T))$  contains a nonempty closed and convex set  $G$ , then  $T$  is pseudomonotone from  $G$  to  $\mathcal{P}(\mathbf{R}^n)$ .*

*Proof:* (See [20], p. 106.)

The next proposition lists several conditions which ensure pseudomonotonicity.

**PROPOSITION 5:** *Consider a maximal monotone operator  $T : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$ . If any of the following conditions holds:*

- i)  $D(T)$  is closed (henceforth convex), set valued function from each line segment in  $D(T)$  to  $\mathbf{R}^n$ ;
  - ii)  $T$  is point-to-point and hemicontinuous, i.e., for all  $x, y \in \mathbf{R}^n$  the map  $\varphi(t)$  defined as  $\varphi(t) = \langle T((1-t)x + ty), x - y \rangle$  is continuous;
  - iii)  $T = \nabla f$  with  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  convex and differentiable;
- then  $T$  is pseudomonotone from  $D(T)$  to  $\mathcal{P}(\mathbf{R}^n)$ .

*Proof:*

- i) This result can be found in [5], Proposition (7.4).
- ii) (See [20], p. 107.)

iii) Follows from (ii) and the fact that the gradient of a convex and differentiable function is hemicontinuous, proved in [20], p. 94.  $\square$

It is easy to verify that every point-to-point and continuous operator is pseudomonotone. In particular, if we take  $T(x) = Ax$  with  $A \in \mathbf{R}^{n \times n}$ ,  $A$  positive semidefinite and nonsymmetric is an example of a monotone and pseudomonotone operator which is not the subdifferential of any convex function.

**4. STATEMENT OF THE ALGORITHM**

Let  $T : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$  be a maximal monotone operator and  $\{\lambda_k\}$  a sequence of positive real numbers bounded above by some  $\tilde{\lambda} > 0$  and  $\varphi \in \Phi_5$ . We define algorithm GPPA as follows.

i) **Initialization:**

$$x^0 \in \mathbf{R}_{++}^n. \tag{6}$$

ii) **Iterative step:** Given  $x^k \in \mathbf{R}_{++}^n$ , if  $x^k$  is a solution of  $\text{NCP}(T)$ , stop. Otherwise define  $T_k : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$  as  $T_k(\cdot) = T(\cdot) + \lambda_k \nabla_x d_\varphi(\cdot, x^k)$  and let  $x^{k+1} \in \mathbf{R}_{++}^n$  be such that

$$0 \in T_k(x^{k+1}). \tag{7}$$

As we mentioned in Section 1, for the case in which  $T = \partial f$ , it has been shown in [10] and [11] that when the set of minimizers of  $f$  is not empty, the sequence given by (6-7) converges to one of these minimizers, as long as any of several technical hypotheses hold (e.g. when the set of minimizers is bounded, or when  $x^0$  is close enough to the set of solutions, or when  $\varphi \in \Phi_4$ ). For a general  $\text{NCP}(T)$ , Auslender and Haddou considered only the case of  $\varphi$  as in Example 1 and proved in [1] convergence to a solution when the following conditions hold:

- (H1)  $S^* \neq \emptyset$ .
- (H2)  $\mathbf{R}_{++}^n \cap D(T) \neq \emptyset$ .
- (H3) Each subproblem has a solution in  $\mathbf{R}_+^n$ .
- (H4) There exists  $u^* \in S^*$  such that for all  $\epsilon > 0$ , and for all  $K \subset \mathbf{R}_{++}^n \cap \{y \in \mathbf{R}_+^n \mid \forall x^* \in S^*, \|y - x^*\| \geq \epsilon\}$ , where  $K$  is bounded, it holds that

$$\inf_{u \in K, c \in Tu} \langle c, u - u^* \rangle > 0.$$

We will prove that at each step  $k$  of the algorithm,  $x^{k+1}$  exists (Hypothesis (H3)). It is unique by requiring some additional conditions on  $T$ . Our proof will hold for any  $\varphi \in \Phi_5$  (see Def. 4). Instead of condition (H4), we will make some hypotheses on  $T$ , not involving the solution set.

For proving in Section 5 that the sequence given by (6-7) is well defined and contained in  $\mathbf{R}_{++}^n$  we will need some preliminary material. For any convex set  $X$  in  $\mathbf{R}^n$ ,  $\text{ri}(X)$  will denote the relative interior of  $X$ .

**PROPOSITION 6:** *Let  $T_1, T_2 : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$  be maximal monotone operators. Suppose that  $D(T_1) \cap \text{int}(D(T_2)) \neq \emptyset$ . Then  $T_1 + T_2$  is a maximal monotone operator.*

*Proof:* (See [21], Th. 1.)

*Remark 2:* Suppose that  $T$  is of the form  $T = \tilde{T} + N_V$ , where

- a)  $\tilde{T}$  is maximal monotone, with  $D(\tilde{T})$  closed;
- b)  $\text{int}(D(\tilde{T})) \supset G$ , for some nonempty closed and convex set  $G$ ;
- c)  $V$  is a closed and convex set with  $V \cap \text{int}(D(\tilde{T})) \neq \emptyset$ .

Then  $T$  is maximal monotone and pseudomonotone from  $D(T) = D(\tilde{T}) \cap V$  to  $\mathcal{P}(\mathbf{R}^n)$ . Indeed, as a consequence of Proposition 6 and condition (c),  $T$  is maximal monotone. Let us prove now the pseudomonotonicity. By Proposition 4, conditions (a) and (b) imply that  $\tilde{T}$  is pseudomonotone. Since  $N_V$  is also pseudomonotone and the sum of pseudomonotone operators is also pseudomonotone (see [20], p. 97), we deduce that  $T$  is pseudomonotone.

*Remark 3:* It is easy to check that for closed and convex sets  $C$  and  $V$ , the solution set of  $\text{VIP}(\tilde{T} + N_V, C)$  coincides with the solution set of  $\text{VIP}(\tilde{T}, V \cap C)$  for any monotone operator  $\tilde{T}$ . Our algorithm is devised for  $\text{NCP}(T) = \text{VIP}(T, \mathbf{R}_+^n)$ , but, by taking  $T := \tilde{T} + N_V$ , it can be used for  $\text{VIP}(\tilde{T}, \mathbf{R}_+^n \cap V)$  for any closed and convex set  $V \subset \mathbf{R}^n$ . In this case the constraints in  $V$  are transferred to the subproblems. For instance in the linear programming case we would have  $\tilde{T}(x) = c$  (constant) and  $V = \{x \in \mathbf{R}^n : Ax = b\}$ . In this case, the subproblems become

$$\min \langle c, x \rangle + \lambda_k d_\varphi(x, x^k) \text{ s.t. } Ax = b,$$

thus yielding an interior point algorithm for the linear programming problem.

We could have a simpler convergence proof assuming, instead of pseudomonotonicity, continuity or even local boundedness of  $T$  ( $T$  is said to be *locally bounded* at  $x$  if there exists a neighborhood  $U$  of  $x$  such that  $T(U)$  is bounded), but that would not cover cases like  $T = \tilde{T} + N_V$ , since  $N_V$  is

unbounded on the boundary of  $V$ . Our pseudomonotonicity assumption, on the other hand, covers this situation when  $\tilde{T}$  is well behaved, e.g. when it satisfies conditions (a), (b) and (c) of this remark.

Note that existence of  $x^{k+1}$  satisfying (7) is not immediate at all, and will be ensured only under some extra assumptions on  $T$  and  $\varphi$ . This issue is the matter of the next section.

**5. WELLDEFINEDNESS OF THE SEQUENCE**

Now we introduce a function that will be useful in the sequel.

Let  $T : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$  be a maximal monotone operator and  $C$  a nonempty, closed and convex subset of  $\mathbf{R}^n$  such that  $D(T) \cap C \neq \emptyset$ . We define the function

$$h_{T,C} : \mathbf{R}^n \rightarrow \mathbf{R} \cup \{\infty\}$$

as

$$h_{T,C}(x) := \sup_{(y,v) \in G_C(T)} \langle v, x - y \rangle, \tag{8}$$

where  $G_C(T) := \{(y, v) | v \in T(y), y \in D(T) \cap C\}$ .

This function, which is called the gap function associated with  $T$  and  $C$ , has interesting properties, some of which are described in the following lemma.

LEMMA 2: Let  $T, C$  and  $h_{T,C}$  be as given by the definition above. Then

- i)  $h_{T,C}$  is convex.
- ii)  $h_{T,C} \geq 0$  in  $D(T) \cap C$ .

*Proof:*

- i)  $h_{T,C}(x)$  is the supremum of a family of affine transformations, which are, in particular, convex functions, and the supremum of a family of convex functions is always convex.

- ii) Take  $y = x$  in (8). □

We need now the notion of regularity, a property of maximal monotone operators introduced in [4]. Let  $G(T)$  denote the graph of  $T$ , i.e.  $G(T) = \{(y, v) | y \in \mathbf{R}^n, v \in T(y)\}$ .

DEFINITION 8: A maximal monotone operator  $T : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$  is said to be regular iff for all  $u \in R(T)$  and for all  $x \in D(T)$ , it holds that  $\sup_{(y,v) \in G(T)} \langle v - u, x - y \rangle < \infty$ .

PROPOSITION 7: If  $T = \partial f$ , with  $f$  closed, proper and lower semicontinuous, then  $T$  is regular.

*Proof:* (See [4], p. 167.)

We want to analyze the range of the sum of two maximal monotone operators  $T_1, T_2 : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$ , i.e., we will study the sets

$$R(T_1 + T_2) = \bigcup_{u \in D(T_1) \cap D(T_2)} T_1 u + T_2 u$$

and

$$R(T_1) + R(T_2) = \bigcup_{u \in D(T_1), v \in D(T_2)} T_1 u + T_2 v.$$

For any set  $X$  in  $\mathbf{R}^n$ ,  $\text{conv}(X)$  will denote the convex hull of  $X$ , and  $\overline{X}$  the closure of  $X$ .

The following lemma was proved in [4], in the more general case of a Hilbert space.

LEMMA 3: Let  $T_0 : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$  a maximal monotone operator and  $F \subset \mathbf{R}^n$ . If for all  $u \in F$  there exists  $x \in \mathbf{R}^n$  such that

$$\sup_{(y,v) \in G(T_0)} \langle v - u, x - y \rangle < \infty, \tag{9}$$

then,

- i)  $\text{conv}(F) \subseteq \overline{R(T_0)}$ ;
- ii)  $\text{conv}(\text{int}(F)) \subseteq R(T_0)$ .

The following proposition is new and can be useful for other applications.

PROPOSITION 8: Let  $T_1, T_2$  be monotone operators defined in  $\mathbf{R}^n$ . Suppose that they satisfy the following conditions:

- a)  $T_1$  is regular;
- b)  $D(T_1) \cap D(T_2) \neq \emptyset$  and  $R(T_1) = \mathbf{R}^n$ ;
- c)  $T_1 + T_2$  is maximal monotone.

Then  $R(T_1 + T_2) = \mathbf{R}^n$ .

*Proof:* We apply Lemma 3 to  $F := R(T_1) + R(T_2)$  and  $T_0 := T_1 + T_2$ . The result will follow from Lemma 3 (i) or (ii) and assumption (b), if we prove that  $F$  satisfies (9). We proceed to verify (9) for  $F = R(T_1) + R(T_2)$ . Let  $u \in R(T_2) + R(T_1)$ , and take  $x \in D(T_1) \cap D(T_2)$  and  $w \in T_2(x)$ . Then  $u = w + (u - w)$ .

Since  $R(T_1) = \mathbf{R}^n$ , there exists  $y \in \mathbf{R}^n$  such that  $u - w \in T_1(y)$ . Then, since  $T_1$  is regular, given  $u - w \in R(T_1)$  and  $x \in D(T_1)$ , there exists some  $\gamma \in \mathbf{R}$  such that

$$\sup_{(z,s) \in G(T_1)} \langle s - (u - w), x - z \rangle = \gamma.$$

Then, for any  $(z, s) \in G(T_1)$ , we have

$$\langle s - (u - w), x - z \rangle \leq \gamma. \tag{10}$$

Take  $v \in T_2(z)$ , with  $z \in D(T_2) \cap D(T_1)$ . By monotonicity of  $T_2$ , we have

$$\langle v - w, z - x \rangle \geq 0, \tag{11}$$

because  $w \in T_2(x)$ . Adding (10) and (11), we obtain

$$\langle (s + v) - u, x - z \rangle \leq \gamma,$$

for any  $z \in D(T_1 + T_2)$ ,  $s \in T_1(z)$ ,  $v \in T_2(z)$ , i.e., for any  $s + v \in (T_1 + T_2)(z)$ . Therefore,

$$\sup_{(z,t) \in G(T_1+T_2)} \langle t - u, x - z \rangle < +\infty$$

and (9) is established for  $F = R(T_1) + R(T_2)$  and  $T_0 = T_1 + T_2$ . □

Now we can state and prove the following lemma.

LEMMA 4: *Let  $T : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$  be a maximal monotone operator such that  $D(T) \cap \mathbf{R}^{n}_{++} \neq \emptyset$ . Take  $y \in \mathbf{R}^{n}_{++}$  and  $\lambda > 0$ , assume that  $\varphi \in \Phi$  is zone coercive and define  $\hat{T}(\cdot) := T(\cdot) + \lambda \nabla_x d_\varphi(\cdot, y)$ . Then*

- i)  $\hat{T}$  is maximal monotone;
- ii)  $R(\hat{T}) = \mathbf{R}^n$ .

*Proof:*

- i) The operator  $\hat{T}$  is obviously monotone. Let  $T_1 := T$ , and  $T_2 := \lambda \nabla_x d_\varphi(\cdot, y)$ . By Lemma 1,  $D(T_1) \cap \text{int}(D(T_2)) = D(T) \cap \mathbf{R}^{n}_{++} \neq \emptyset$  by assumption. By Proposition 6,  $\hat{T} = T_1 + T_2$  is maximal monotone.
- ii) It is a consequence of Proposition 8 which can be applied because conditions (a), (b) and (c) hold for  $T_2 = T$  and  $T_1 = \lambda \nabla_x d_\varphi(\cdot, y)$ . More precisely,



- a)  $\lambda \nabla_x d_\varphi(\cdot, y)$  is regular, by Proposition 7;
- b)  $D(T) \cap \mathbf{R}_{++}^n \neq \emptyset$  and  $R(\nabla_x d_\varphi(\cdot, y)) = \mathbf{R}^n$  by zone coerciveness;
- c)  $\hat{T}$  is maximal monotone by (i).

Then  $R(\hat{T}) = \mathbf{R}^n$  by Proposition 8.

**COROLLARY 1:** *If  $T_1$  and  $T_2$  are maximal monotone operators such that  $D(T_2) \cap \text{int}(D(T_1)) \neq \emptyset$  then*

- i)  $T_1 + T_2$  is maximal monotone.
- ii) *If furthermore  $T_2$  is the subdifferential of a proper and lower semicontinuous convex function, and  $T_2$  is onto, then  $T_1 + T_2$  is onto.*

*Proof:* (i) follows from Proposition 6. Conclusion (ii) follows from Propositions 7 and 8. □

**PROPOSITION 9:** *Let  $T : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$  be maximal monotone. If  $D(T)$  is bounded, then  $T$  is onto.*

*Proof:* (See [3], Cor. 2.2.)

From now on, we work with a maximal monotone operator  $T : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$ .

Now we present our existence results.

**THEOREM 1:** *Suppose that the following conditions hold:*

- i)  $D(T) \cap \mathbf{R}_{++}^n \neq \emptyset$ ;
- ii)  $\varphi$  is zone coercive.

*Then, for each  $x^0$  satisfying (6), (7) defines a unique sequence  $\{x^k\}$  contained in  $\mathbf{R}_{++}^n$ .*

*Proof:* The proof will be performed by induction on  $k$ .  $x^0 \in \mathbf{R}_{++}^n$  by (6). Suppose that  $x^k \in \mathbf{R}_{++}^n$ . Let  $B_k(\cdot) := \lambda_k \nabla d_\varphi(\cdot, x^k)$ . Then  $B_k$  is strictly monotone because  $\varphi$  is strictly convex. This implies strict monotonicity of  $T_k := T + B_k$ . By Lemma 4(ii),  $T_k$  has a zero in  $D(T_k)$ , which is unique by strict monotonicity. We call this zero  $x^{k+1}$ . We have to show that this zero belongs to  $\mathbf{R}_{++}^n$ . By Lemma 1,  $D(T_k) = D(T) \cap \mathbf{R}_{++}^n$ , and since  $x^{k+1} \in D(T_k)$  we obtain that  $x^{k+1} \in \mathbf{R}_{++}^n$ . We conclude that  $\{x^k\} \subset \mathbf{R}_{++}^n$ . □

The theorem below introduces a different kind of hypothesis on the operator  $T$ , namely, the finiteness of the function  $h_{T, \mathbf{R}_+^n}$ . The following

results explain the meaning of this assumption and give some conditions under which it holds for a maximal monotone operator  $T$ .

We will consider  $h_{T, \mathbf{R}_+^n}$  as in (8), for  $C = \mathbf{R}_+^n$ .

DEFINITION 9: Suppose that  $D(T) \cap C \neq \emptyset$ . We will say that  $T$  is  $C$ -stable at  $x \in D(T) \cap C$  iff  $h_{T,C}(x) < \infty$ . For the case in which  $C = \mathbf{R}_+^n$ , we are considering the  $NCP(T)$ . We will say that  $T$  is stable iff  $T$  is  $\mathbf{R}_+^n$ -stable.

The following lemma establishes a relation between stability and existence of solutions of the  $NCP(T)$ .

LEMMA 5: Suppose that  $\text{int}(D(T)) \cap \mathbf{R}_+^n \neq \emptyset$ . If  $T$  is stable at some  $a \in D(T) \cap \mathbf{R}_+^n$ , then  $0 \in \overline{R(T + N_{\mathbf{R}_+^n})}$ .

Proof: For proving our claim, we will use Lemma 3 for  $F = \{0\}$  and  $T_0 = T + N_{\mathbf{R}_+^n}$ . Take  $y \in D(T) \cap \mathbf{R}_+^n$  and  $v \in (T + N_{\mathbf{R}_+^n})(y)$ . Then there exist  $u \in N_{\mathbf{R}_+^n}(y)$  and  $w \in T(y)$  such that  $v = u + w$ . We have that

$$\langle v, a - y \rangle = \langle u + w, a - y \rangle = \langle u, a - y \rangle + \langle w, a - y \rangle \leq \langle w, a - y \rangle \leq h_{T, \mathbf{R}_+^n}(a),$$

where we used in the expression above the definition of  $N_{\mathbf{R}_+^n}(y)$  and the fact that  $a \in \mathbf{R}_+^n$ . Taking in the last chain of inequalities supremum over the graph of  $T + N_{\mathbf{R}_+^n}$ , we obtain:

$$\sup_{(y,v) \in G(T+N_{\mathbf{R}_+^n})} \langle v, a - y \rangle \leq h_{T, \mathbf{R}_+^n}(a).$$

On the other hand, it is easy to check that the supremum above coincides with the one in (9) for the mentioned choices of  $T_0$  and  $F$ . Then our hypothesis on stability at  $a$  implies that condition (9) holds. Under this condition, Lemma 3 (i) yields that  $0 \in \overline{R(T + N_{\mathbf{R}_+^n})}$ . □

Remark 4: It is not true in general that  $0 \in R(T + N_{\mathbf{R}_+^n})$  under the assumptions of Lemma 5 as the following example shows.

Example 4: Take  $n = 1$  and  $f(t) = e^{-t}$ . Then  $h_{\nabla f, \mathbf{R}_+}$  has no minimizers. Indeed,

$$h_{\nabla f, \mathbf{R}_+}(x) = \sup_{y \in \mathbf{R}_+} (-e^{-y})(x - y) = e^{-x-1}.$$

Then  $\nabla f : \mathbf{R} \rightarrow \mathbf{R}$  is stable. However, we know that  $\nabla f$  has no zeroes. Take again  $C$  as a closed and convex set, we summarize below some conditions under which  $h_{T,C}$  is finite for any  $x \in D(T) \cap C$ . We denote by  $S(T, C)$  the solution set of  $VIP(T, C)$ .

PROPOSITION 10: Let  $T, C$  and  $h_{T,C}$  be as in (8).

a)  $T$  is  $C$ -stable (i.e.,  $C$ -stable at any  $x \in D(T) \cap C$ ) if any one of the following conditions holds:

1)  $T = \partial f$  where  $f$  is a convex function such that

$$\inf_C f > -\infty. \quad (12)$$

2)  $T + N_C$  is regular and  $S(T, C) \neq \emptyset$ .

3)  $T$  is strongly monotone (with modulus  $\alpha > 0$ ), i.e.,

$$\langle v - w, x - y \rangle \geq \alpha \|x - y\|^2 \quad \text{whenever } v \in T(x), w \in T(y).$$

4)  $\overline{D(T)}$  is bounded.

b)  $T$  is  $C$ -stable at a point  $x^0 \in D(T) \cap C$  if there exist  $\alpha > 0$  and  $c \in \mathbf{R}$  with

$$\langle v, y - x^0 \rangle \geq -c, \quad (13)$$

for all  $(y, v) \in G_C(T)$  such that  $\|y\| > \alpha$ .

*Proof:*

a)

1)

$$\begin{aligned} h_{\partial f, C}(x) &= \sup_{G_C(\partial f)} \langle v, x - y \rangle \leq \sup_{y \in C} \langle u, x - y \rangle \\ &\leq \sup_{y \in C} (f(x) - f(y)) = f(x) - \inf_C f < \infty, \end{aligned}$$

where we used (8), monotonicity, the subgradient property and (12) respectively.

2) If  $S(T, C) \neq \emptyset$  then there exists  $z \in D(T) \cap C$  with  $0 \in (T + N_C)z$ . Then taking  $u = 0 \in R(T + N_C)$  in the definition of regularity we have that for any  $x \in D(T + N_C) = D(T) \cap C$ ,

$$h_{T, C}(x) = \sup_{(y, v) \in G_C(T)} \langle v, x - y \rangle < \infty.$$

3) As  $x \in D(T) \cap C$ , then there exists some  $u \in T(x)$ . Using this fact and the definition of strong monotonicity, we can write:

$$\begin{aligned} h_{T, C}(x) &= \sup_{(y, v) \in G_C(T)} \langle v, x - y \rangle \leq \sup_{y \in D(T) \cap C} \langle u, x - y \rangle - \alpha \|x - y\|^2 \\ &= \sup_{y \in D(T) \cap C} \|x - y\| \left( \frac{\langle u, x - y \rangle}{\|x - y\|} - \alpha \|x - y\| \right). \end{aligned}$$

As  $u \in T(x)$  is fixed, we can take  $R > 0$  large enough, so that whenever  $\|x - y\| > R$ , the expression between parentheses is nonpositive. This implies that the supremum can be taken for  $y$  such that  $\|x - y\| \leq R$ , and hence it is finite. This establishes the finiteness of  $h_{T,C}(x)$ .

- 4) Suppose that  $\overline{D(T)}$  is bounded. By (8) and monotonicity, we have for any  $u \in T(x)$ ,

$$h_{T,C}(x) = \sup_{(y,v) \in G_C(T)} \langle v, x - y \rangle \leq \sup_{y \in C} \langle u, x - y \rangle < \infty,$$

where the last inequality holds by boundedness of  $\overline{D(T)}$ .

- b) Suppose that there exist  $\alpha$  and  $c$  as in (13). We claim that it is enough to evaluate the supremum in  $h_{T,C}$  in a bounded subset of  $C$ . Indeed, take  $u \in T(x^0)$ , (which is possible because  $x^0 \in D(T)$ ), and then

$$\begin{aligned} 0 \leq h_{T,C}(x^0) &= \sup_{y \in D(T) \cap C, v \in T(y)} \langle v, x^0 - y \rangle \\ &= \max \left\{ c, \sup_{y \in C, \|y\| \leq \alpha, v \in T(y)} \langle v, x^0 - y \rangle \right\} \\ &\leq \max \left\{ c, \sup_{y \in C, \|y\| \leq \alpha} \langle u, x^0 - y \rangle \right\}, \end{aligned}$$

where the second equality is a direct consequence of (13) and the rightmost inequality holds by monotonicity. But the rightmost expression is finite because  $y$  is taken in a bounded set. So  $h_{T,C}(x^0) < \infty$  and the  $C$ -stability at  $x^0$  has been established.  $\square$

From now on, we will write  $h := h_{T, \mathbf{R}_+^n}$

LEMMA 6: For  $h$  as above, suppose that  $D(T) \cap \mathbf{R}^{n}_{++} \neq \emptyset$ . Then  $h$  has zeroes in  $\mathbf{R}_+^n$  if and only if  $S^*$  is nonempty, and in such a case  $S^*$  is precisely the set of zeroes of  $h$ .

Proof: We will see first that if  $x \in \mathbf{R}_+^n$  is a zero of  $h$ , then it is a solution of  $NCP(T)$ . Define  $A := T + N_{\mathbf{R}_+^n}$ , so that  $D(A) = D(T) \cap \mathbf{R}_+^n$ . If  $h(x) = 0$ , then, by (8), we have that for any  $y \in D(T) \cap \mathbf{R}_+^n, v \in Ty$ ,

$$\langle v, x - y \rangle \leq 0.$$

On the other hand, if we take any  $w \in N_{\mathbf{R}_+^n}(y)$ , then, as  $x \in \mathbf{R}_+^n$ , by definition of  $N_{\mathbf{R}_+^n}$ ,

$$\langle w, x - y \rangle \leq 0.$$

Adding both inequalities above, we obtain

$$\langle 0 - (v + w), x - y \rangle \geq 0,$$

for any  $y \in D(T) \cap \mathbf{R}_+^n = D(A)$  and any  $v + w \in Ay = Ty + N_{\mathbf{R}_+^n}(y)$ .

By Proposition 6,  $A$  is maximal monotone. Then the latter fact together with the last inequality, yield

$$0 \in Ax = Tx + N_{\mathbf{R}_+^n}(x),$$

and this implies that  $x \in S^*$ .

Conversely, suppose now that we have an element  $x \in S^*$ . Then for any  $y \in \mathbf{R}_+^n$ , there exists an element  $v \in Tx$  with

$$\langle v, y - x \rangle \geq 0.$$

Using the last inequality and monotonicity, we have that for any  $w \in Ty$ ,  $\langle w, y - x \rangle \geq 0$ . Then it follows from (8) that

$$h(x) = \sup_{(y,w) \in G_{\mathbf{R}_+^n}(T)} \langle w, x - y \rangle \leq 0.$$

As  $h(x) \geq 0$  we conclude that  $h(x) = 0$ , as we wanted to prove. □

**THEOREM 2:** *Suppose that the following conditions hold:*

- i)  $S^* \neq \emptyset$ ;
- ii)  $D(T) \cap \mathbf{R}_{++}^n \neq \emptyset$ ;
- iii)  $h(x) < \infty$  for all  $x \in \mathbf{R}_+^n \cap D(T)$ ;
- iv)  $\varphi$  is boundary coercive.

If  $x^k \notin S^*$ , then (7) uniquely defines a vector  $x^{k+1} \in \mathbf{R}_{++}^n$ .

*Proof:* Now we do not have the property  $R(T_k) = \mathbf{R}^n$ , which is essential in Theorem 1 and is implied by zone coerciveness of  $\varphi$ . This hypothesis is replaced here by boundary coerciveness of  $\varphi$  and the finiteness of  $h$  in  $\mathbf{R}_+^n \cap D(T)$ . We proceed by induction.  $x^0 \in \mathbf{R}_{++}^n$  by (6). By inductive hypothesis and (7), there exists  $x^k \in D(T) \cap \mathbf{R}_{++}^n$  with  $0 \in T_{k-1}(x^k)$ . So we have that  $0 \leq h(x^k) < \infty$ . By assumption,  $x^k \notin S^*$ . By (i) and Lemma 6  $h(x^k) > 0$ . In this case, define the set

$$S_k := \left\{ x \in \mathbf{R}_{++}^n \mid d_\varphi(x, x^k) < \frac{h(x^k)}{\lambda_k} \right\}.$$

By Proposition 1(iii), (iv),  $S_k$  is open, convex and bounded. Then its closure is also convex. Observe that  $x^k \in S_k$  because  $d_\varphi(\cdot, x^k)$  is continuous and  $d_\varphi(x^k, x^k) = 0 < h(x^k)/\lambda_k$  and that also  $x^k \in \mathbf{R}_{++}^n$  by inductive hypothesis. Let  $N_k := N_{\overline{S_k}}$ , the normality operator of  $\overline{S_k}$ . Being the subdifferential of a closed and convex function,  $N_k$  is a maximal monotone operator with domain  $D(N_k) = \overline{S_k}$ , which is also a bounded set. This implies, using Proposition 9, that  $N_k$  is onto. Now we define  $B_k(\cdot) := N_k(\cdot) + \lambda_k \nabla_x d_\varphi(\cdot, x^k)$ . In order to prove that  $B_k$  is maximal monotone, we check the condition of Proposition 6. By the induction hypothesis and Lemma 1,  $x^k \in D(N_k) \cap \mathbf{R}_{++}^n = \overline{S_k} \cap D(\nabla_x d_\varphi(\cdot, x^k))$ . Then  $D(N_k) \cap \text{int}(D(\nabla_x d_\varphi(\cdot, x^k))) \neq \emptyset$  and so Proposition 6 applies. Then  $B_k$  is maximal. Also,  $D(B_k)$  is bounded, being a subset of  $S_k$ , and using Proposition 9 we conclude that  $R(B_k) = \mathbf{R}^n$ . Let us consider now the operator  $A_k =: T + B_k$ . We check now that this operator is also maximal monotone. By Proposition 6, it will be enough to show that  $D(T) \cap \text{int}(D(B_k)) \neq \emptyset$ . Actually, because  $D(B_k) = \mathbf{R}_{++}^n \cap \overline{S_k}$ , we can write

$$x^k \in D(T) \cap \mathbf{R}_{++}^n \cap S_k = D(T) \cap \text{int}(D(B_k)).$$

Then  $A_k$  is maximal monotone. Now using Proposition 9 again we get that  $A_k$  is onto, because its domain is a subset of  $D(B_k)$ , which is bounded. Then there exists  $y \in D(A_k) = D(T) \cap D(B_k) \subset D(T) \cap \mathbf{R}_{++}^n$ , such that

$$0 \in T(y) + N_k(y) + \lambda_k \nabla_x d_\varphi(y, x^k). \tag{14}$$

Let us call  $u^k, w^k$  and  $v^k$  the elements in  $\mathbf{R}^n$  such that

$$u^k \in T(y), w^k \in N_k(y), \text{ and } v^k \in \lambda_k \nabla_x d_\varphi(y, x^k),$$

and

$$0 = u^k + w^k + v^k. \tag{15}$$

We claim that  $y \in S_k$ . Since  $y \in D(A_k) \subset D(B_k) \subset D(\nabla_x d_\varphi(\cdot, x^k)) = \mathbf{R}_{++}^n$  by Lemma 1, it will be enough to show, in view of the definition of  $S_k$ , that

$$d_\varphi(y, x^k) < \frac{h(x^k)}{\lambda_k}. \tag{16}$$

We proceed to establish (16). Applying the gradient inequality to  $d_\varphi(\cdot, x^k)$ , which is a strictly convex function, by Proposition 1(iv) we obtain

$$0 = d_\varphi(x^k, x^k) > d_\varphi(y, x^k) + \left\langle \frac{v^k}{\lambda_k}, x^k - y \right\rangle. \tag{17}$$

Using (15) in (17) and rearranging terms

$$\frac{1}{\lambda_k} (\langle u^k, x^k - y \rangle + \langle w^k, x^k - y \rangle) > d_\varphi(y, x^k). \tag{18}$$

Since  $w^k \in N_k(y)$ , and  $x^k \in S_k$ , we have by definition of  $N_k$  that

$$\langle w^k, x^k - y \rangle \leq 0. \tag{19}$$

Using (19) in (18) we get

$$0 \leq \lambda_k d_\varphi(y, x^k) < \langle u^k, x^k - y \rangle. \tag{20}$$

Since  $\langle u^k, x^k - y \rangle \leq \sup_{z \in \mathbf{R}_+^n \cap D(T), v \in Tz} \langle v, x^k - z \rangle = h(x^k)$ , we obtain from (20) that  $d_\varphi(y, x^k) < \frac{h(x^k)}{\lambda_k}$ , *i.e.* (16) holds. We have proved that  $y \in S_k = \text{int}(\bar{S}_k)$  and therefore  $N_k(y) = \{0\}$ . Where we used the well-known fact that any convex set  $X$  for which  $\text{int}(X) \neq \emptyset$  satisfies  $\text{int}(X) = \text{int}(\bar{X})$ . Then  $w^k = 0$  and by (15)  $0 = u^k + v^k$ . In view of (14),  $0 \in T_k(y)$ . By strict monotonicity of  $T_k$ ,  $y$  is the only zero of  $T_k$ . By (7),  $y = x^{k+1}$ . As  $y \in \mathbf{R}_{++}^n$ , then  $x^{k+1} \in \mathbf{R}_{++}^n$  and the induction step is complete. □

**6. CONVERGENCE RESULTS**

We have not been able to obtain the convergence result for any  $\varphi \in \Phi_4$ , as it is proved for the case  $T = \partial f$  in [11]. We must restrict our set of functions  $\varphi$  to  $\Phi_5$  (see Def. 4). We remark that this hypothesis on  $\varphi$  is weaker than those imposed by Auslender and Haddou in [1], and the assumptions of our existence and convergence results are satisfied, *e.g.*, by the families of  $\varphi$ -divergences mentioned in Remark 1.

In this section we will establish the convergence properties of the sequence generated by (6-7). We will use in our analysis the Kullback-Liebler divergence introduced in Example 1, which will be denoted by  $\psi$ , *i.e.*,

$$\psi(t) = t \log t - t + 1.$$

We recall here that this  $\varphi$ -divergence can be extended to  $\mathbf{R}_+^n \times \mathbf{R}_{++}^n$ . Before presenting the convergence results, we need some previous tools. Let us call  $\alpha := \varphi''(1)$ .

Consider  $\varphi \in \Phi_5$ , i.e., such that

$$\left(1 - \frac{1}{t}\right) < \frac{\varphi'(t)}{\alpha} \leq \log t, \text{ for all } t > 0, t \neq 1, \tag{21}$$

and

$$\lim_{t \rightarrow \infty} \varphi'(t) > \alpha, \quad \lim_{t \rightarrow 0^+} t\varphi'(t) > -\alpha.$$

Define  $\bar{\varphi}(t) := t\varphi'(t) - \alpha(t - 1)$ , and consider  $d_{\bar{\varphi}}(\cdot, \cdot)$  with the same structure as in Definition 1, i.e.,

$$d_{\bar{\varphi}}(\cdot, \cdot) : \mathbf{R}_{++}^n \times \mathbf{R}_{++}^n \rightarrow \mathbf{R}_+$$

and

$$d_{\bar{\varphi}}(x, y) = \sum_{j=1}^n y_j \bar{\varphi}\left(\frac{x_j}{y_j}\right). \tag{22}$$

Observe that  $d_{\bar{\varphi}}(\cdot, \cdot)$  is not necessarily a  $\varphi$ -divergence because  $\bar{\varphi}$  may fail to satisfy some of the conditions of Section 2. Nevertheless  $d_{\bar{\varphi}}(\cdot, \cdot)$  shares some properties of  $\varphi$ -divergences, as the next lemma shows.

LEMMA 7: Let  $d_{\bar{\varphi}}(\cdot, \cdot)$  be defined as above with  $\varphi \in \Phi_5$ . Then it satisfies:

- i)  $d_{\bar{\varphi}}(x, y) \geq 0$  for all  $x, y \in \mathbf{R}_{++}^n$ ;
- ii) If  $d_{\bar{\varphi}}(x, y) = 0$  then  $x = y$ ;
- iii) If  $\{x^k\}, \{y^k\} \subset \mathbf{R}_{++}^n, x^k \rightarrow x, y^k \rightarrow y$  and  $\lim_{k \rightarrow \infty} d_{\bar{\varphi}}(x^k, y^k) = 0$ , then  $x = y$ .

Proof: i) By definition of  $d_{\bar{\varphi}}(\cdot, \cdot)$ , we can write

$$\begin{aligned} d_{\bar{\varphi}}(x, y) &= \sum_{j=1}^n y_j \bar{\varphi}\left(\frac{x_j}{y_j}\right) = \sum_{j=1}^n y_j \left(\frac{x_j}{y_j} \varphi'\left(\frac{x_j}{y_j}\right) - \alpha\left(\frac{x_j}{y_j} - 1\right)\right) \\ &= \sum_{j=1}^n x_j \varphi'\left(\frac{x_j}{y_j}\right) - \alpha(x_j - y_j) \\ &= \sum_{j=1}^n x_j \left(\varphi'\left(\frac{x_j}{y_j}\right) - \alpha\left(1 - \frac{y_j}{x_j}\right)\right). \end{aligned}$$

Defining  $t_j := \frac{x_j}{y_j}$  we get from the last formula:

$$d_{\bar{\varphi}}(x, y) = \sum_{j=1}^n x_j \left(\varphi'(t_j) - \alpha\left(1 - \frac{1}{t_j}\right)\right),$$



and this expression is strictly positive by (21) and the fact that  $x_j > 0$  for all  $j$ .

ii) Suppose that  $d_{\bar{\varphi}}(x, y) = 0$ . Then all terms in the expression above must be zero. Since  $x \in \mathbf{R}_{++}^n$ , we conclude that  $\varphi'(t_j) = \alpha(1 - \frac{1}{t_j})$  for all  $j$ . From (21), we get  $t_j = 1$  for all  $j$ . Then  $x_j = y_j$  for all  $j$ . Then we can conclude that  $x = y$ .

iii) Use (22) with  $x = x^k$  and  $y = y^k$ , and  $t_j^k := \frac{x_j^k}{y_j^k}$  to get

$$d_{\bar{\varphi}}(x^k, y^k) = \sum_{j=1}^n y_j^k \bar{\varphi}(t_j^k). \tag{23}$$

Since  $d_{\bar{\varphi}}(x^k, y^k) \geq 0$  and each term in (23) is nonnegative, we have that all of them must go to zero. That is,

$$\lim_{k \rightarrow \infty} y_j^k \bar{\varphi}(t_j^k) = 0 \text{ for all } j. \tag{24}$$

We will consider two cases.

- a) Suppose that  $j$  is such that  $y_j = 0$ . We will prove that  $\lim_{k \rightarrow \infty} x_j^k = 0$ . Suppose that  $\lim_{k \rightarrow \infty} x_j^k > 0$ . We will show that this assumption leads to a contradiction. In this situation, we have that  $\lim_{k \rightarrow \infty} t_j^k = +\infty$ . Hence, using the inequality coming after inequality (21),

$$\lim_{k \rightarrow \infty} \varphi'(t_j^k) > \alpha. \tag{25}$$

We know that

$$y_j^k \bar{\varphi}(t_j^k) = x_j^k \left( \varphi'(t_j^k) - \alpha \left( 1 - \frac{1}{t_j^k} \right) \right). \tag{26}$$

By (25) and our assumption on the  $\lim_{k \rightarrow \infty} x_j^k$ , the rightmost expression in (26) has a positive limit, which contradicts (24). This means that  $\lim_{k \rightarrow \infty} x_j^k$  must be 0. So we proved that if  $y_j = 0$ , then  $x_j = 0$ .

- b) Suppose that  $j$  is such that  $y_j > 0$ . Then by (24) we must have

$$\lim_{k \rightarrow \infty} \bar{\varphi}(t_j^k) = 0.$$

Since  $\varphi \in \Phi_{\xi}$ , it is easy to see that this can only happen if

$$\lim_{k \rightarrow \infty} t_j^k = 1, \tag{27}$$

and (27) implies directly that  $x_j = y_j$  also in this case.

So we conclude that in both cases  $x = y$ . □

Finally we present our first convergence result for the Proximal Point Method with  $\varphi$ -divergences.

**THEOREM 3:** *Let  $T : \mathbf{R}^n \rightarrow \mathcal{P}(\mathbf{R}^n)$  be a maximal monotone operator with closed domain. Consider  $NCP(T)$ , and suppose that the following conditions hold:*

- i)  $D(T) \cap \mathbf{R}_{++}^n \neq \emptyset$ ;
- ii)  $NCP(T)$  has solutions, i.e.,  $S^* \neq \emptyset$ ;
- iii)  $T$  is pseudomonotone from  $D(T)$  to  $\mathcal{P}(\mathbf{R}^n)$ .
- iv)  $\lambda_k \in (0, \tilde{\lambda}]$  for some  $\tilde{\lambda} > 0$ ;
- v)  $\varphi \in \Phi_5$ ;
- vi) either
  - vi<sub>1</sub>)  $\varphi$  is zone coercive, or
  - vi<sub>2</sub>)  $\varphi$  is boundary coercive and  $h(x) < \infty$  for all  $x \in \mathbf{R}_+^n \cap D(T)$ .

Then, either the algorithm stops at some iteration  $k$  and  $x^k$  solves  $NCP(T)$ , or it generates an infinite sequence  $\{x^k\}$  such that

- a)  $\{x^k\}$  is bounded and so it has cluster points;
- b) If  $\bar{x}$  is a cluster point of  $\{x^k\}$ , then there exists  $\bar{u} \in T(\bar{x})$  such that  $\langle \bar{u}, x^* - \bar{x} \rangle = 0$  for any  $x^* \in S^*$ .

*Proof:* The finite case follows from the stopping criterion. We consider the case of an infinite sequence. (a) We will show that  $\{x^k\}$  satisfies

$$d_\psi(z, x^{k+1}) \leq d_\psi(z, x^k) - \frac{1}{\alpha} d_{\bar{\varphi}}(x^{k+1}, x^k), \tag{28}$$

for all  $z \in S^*$ , where  $\psi(t) = t \log t - t + 1$  and  $d_{\bar{\varphi}}(\cdot, \cdot)$  is as given before in Lemma 7. By Theorem 1 or Theorem 2 we know that there exists  $x^{k+1} \in \mathbf{R}_{++}^n$  such that

$$0 \in T(x^{k+1}) + \lambda_k \nabla_x d_\varphi(x^{k+1}, x^k) \text{ for } k = 0, 1, \dots \tag{29}$$

Hence, there exists  $u^k \in T(x^{k+1})$  such that

$$\frac{u^k}{\lambda_k} = -\nabla_x d_\varphi(x^{k+1}, x^k) = -\sum_{j=1}^n \varphi' \left( \frac{x_j^{k+1}}{x_j^k} \right) e^j,$$

where we used (29) in the first equality and Lemma 1 in the second one. So, for any  $x \in \mathbf{R}_+^n$ , we have, defining  $t_j^k := \frac{x_j^{k+1}}{x_j^k}$ ,

$$\left\langle \frac{u^k}{\lambda_k}, x^{k+1} - x \right\rangle = \sum_{j=1}^n \varphi'(t_j^k)(x_j - x_j^{k+1}). \tag{30}$$

Using now the definitions of  $d_\psi$ ,  $d_{\bar{\varphi}}$  and (30) we obtain, after some algebra,

$$\begin{aligned} d_\psi(x, x^k) - d_\psi(x, x^{k+1}) - \frac{1}{\alpha} d_{\bar{\varphi}}(x^{k+1}, x^k) - \frac{1}{\alpha \lambda_k} \langle u^k, x^{k+1} - x \rangle \\ = \sum_{j=1}^n x_j (\log t_j^k - \frac{1}{\alpha} \varphi'(t_j^k)) \geq 0, \end{aligned} \tag{31}$$

where the inequality in (31) holds because  $\varphi \in \Phi_3$ .

Let's take now  $z \in S^*$ . By definition of  $S^*$ , there exists  $v^* \in T(z)$  such that  $\langle v^*, y - z \rangle \geq 0$  for all  $y \in \mathbf{R}_+^n$ . By (31) for  $x = z$  we have that

$$\begin{aligned} d_\psi(z, x^k) - d_\psi(z, x^{k+1}) - \frac{1}{\alpha} d_{\bar{\varphi}}(x^{k+1}, x^k) \\ \geq \frac{1}{\alpha} \left\langle \frac{u^k}{\lambda_k}, x^{k+1} - z \right\rangle \geq \frac{1}{\alpha} \left\langle \frac{v^*}{\lambda_k}, x^{k+1} - z \right\rangle \geq 0, \end{aligned} \tag{32}$$

where we used monotonicity in the second inequality. By (32) we get

$$d_\psi(z, x^{k+1}) \leq d_\psi(z, x^k) - \frac{1}{\alpha} d_{\bar{\varphi}}(x^{k+1}, x^k), \tag{33}$$

establishing (28). Then (a) is a consequence of the boundedness of the level sets of  $d_\psi(z, \cdot)$  (see Ex. 1). Namely, by (28) the sequence is contained in the level set

$$\{x \in \mathbf{R}_{++}^n \mid d_\psi(z, x) \leq d_\psi(z, x^0)\},$$

where  $z \in S^*$ . This implies the boundedness of the sequence and so the existence of cluster points. So (a) is proved. (b) We prove first that

$$\lim_{k \rightarrow \infty} d_{\bar{\varphi}}(x^{k+1}, x^k) = 0. \tag{34}$$

By (33) and Lemma 7(i)  $\{d_\psi(z, x^k)\}$  is nonincreasing. Since  $d_\psi(z, x^k) \geq 0$ , it follows that  $\{d_\psi(z, x^k)\}$  is convergent. Then, by (33)

$$0 \leq \frac{1}{\alpha} d_{\bar{\varphi}}(x^{k+1}, x^k) \leq d_\psi(z, x^k) - d_\psi(z, x^{k+1}). \tag{35}$$

Now (34) follows taking limits in (35). The next step will use the hypothesis that  $\varphi \in \Phi_5$ , in order to prove (b).

Let  $\bar{x}$  be a cluster point of  $\{x^k\}$  and  $\{x^{k_j}\}$  a subsequence of  $\{x^k\}$  such that  $\bar{x} = \lim_{k_j \rightarrow \infty} x^{k_j}$ . In view of (a), we may assume without loss of generality that  $\{x^{k_j+1}\}$  converges to  $\bar{y}$ . By (34)  $\lim_{k_j \rightarrow \infty} d_{\bar{\varphi}}(x^{k_j+1}, x^{k_j}) = 0$ . In this case we can apply Lemma 7(iii), which implies that  $\bar{x} = \bar{y}$ . Using now Proposition 1(v) (which holds for  $d_\psi$ ):

$$\lim_{k_j \rightarrow \infty} d_\psi(\bar{x}, x^{k_j}) = \lim_{k_j \rightarrow \infty} d_\psi(\bar{x}, x^{k_j+1}) = 0. \tag{36}$$

Now we consider (31) with  $x = \bar{x}$  and  $k = k_j$ , obtaining

$$d_\psi(\bar{x}, x^{k_j}) - d_\psi(\bar{x}, x^{k_j+1}) - \frac{1}{\alpha} d_{\bar{\varphi}}(x^{k_j+1}, x^{k_j}) \geq \frac{1}{\alpha} \left\langle \frac{u^{k_j}}{\lambda_{k_j}}, x^{k_j+1} - \bar{x} \right\rangle, \tag{37}$$

with  $u^{k_j} \in T(x^{k_j+1})$ . The left hand side of (37) goes to 0 by (34) and (36), therefore

$$\limsup_{k_j \rightarrow \infty} \left\langle \frac{u^{k_j}}{\lambda_{k_j}}, x^{k_j+1} - \bar{x} \right\rangle \leq 0. \tag{38}$$

Since  $\{\lambda_k\} \subset \mathbf{R}_{++}$  and  $\lambda_k$  is bounded above (assumption (iv)), we have by (38)

$$\limsup_{k_j \rightarrow \infty} \langle u^{k_j}, x^{k_j+1} - \bar{x} \rangle \leq 0 \text{ and } u^{k_j} \in T(x^{k_j+1}). \tag{39}$$

We use now (39) together with the pseudomonotonicity assumption. Take  $x^* \in S^*$ . By pseudomonotonicity there exists  $\bar{u} \in T\bar{x}$  such that

$$\langle \bar{u}, \bar{x} - x^* \rangle \leq \liminf_{k_j \rightarrow \infty} \langle u^{k_j}, x^{k_j+1} - x^* \rangle. \tag{40}$$

Since  $x^* \in S^*$  and  $\bar{x} \in \mathbf{R}_+^n$ , we conclude by monotonicity that the left hand side of (40) is nonnegative. On the other hand, by convergence of the sequence  $\{d_\psi(x^*, x^{k_j})\}$  and (34) for  $k = k_j$ , we get from (32) with

$k = k_j$  that the right hand side of (40) is zero. So we found  $\bar{u} \in T\bar{x}$  such that  $\langle \bar{u}, \bar{x} - x^* \rangle = 0$ .

The next result establishes that when  $T$  is also paramonotone, the sequence generated by (6-7) is convergent to a solution of  $\text{NCP}(T)$ .

**THEOREM 4:** *Under the assumptions of Theorem 3, if  $T$  is paramonotone on  $\mathbf{R}_+^n$ , then the sequence  $\{x^k\}$  given by (6)-(7) is convergent to a solution  $\bar{x}$  of  $\text{NCP}(T)$ .*

*Proof:* This theorem is just a continuation of Theorem 3. Since now  $T$  is paramonotone, combining Theorem 3(b) and Proposition 2(ii), we conclude that any cluster point of  $\{x^k\}$  is a solution of  $\text{NCP}(T)$ . It remains to prove that there is at most one cluster point. Let  $\bar{x}$  be a cluster point of  $\{x^k\}$ , i.e., there exists a subsequence  $\{x^{k_j}\}$  of  $\{x^k\}$  such that

$$\lim_{k_j \rightarrow \infty} x^{k_j} = \bar{x}. \quad (41)$$

Since  $\bar{x}$  is a solution of  $\text{NCP}(T)$ , the sequence  $\{d_\psi(\bar{x}, x^k)\}$  is nonincreasing, as established after equation (34) in the proof of Theorem 3. On the other hand, as  $\psi \in \Phi$ , by Proposition 1(v), we have that

$$\lim_{k_j \rightarrow \infty} d_\psi(\bar{x}, x^{k_j}) = 0.$$

A nonnegative and nonincreasing sequence with a subsequence converging to 0, certainly converges to 0, i.e.,

$$\lim_{k \rightarrow \infty} d_\psi(\bar{x}, x^k) = 0.$$

Again, Proposition 1(v) allows us to conclude that

$$\lim_{k \rightarrow \infty} x^k = \bar{x},$$

as we wanted to prove. □

We point out that the pseudomonotonicity assumption required in Theorems 3 and 4 is guaranteed to hold in the cases considered in Propositions 4 and 5. Conditions for existence as (vi<sub>1</sub>) and (vi<sub>2</sub>) in Theorem 3 hold for  $\varphi$  as in Remark 1 or for  $\varphi$  boundary coercive and  $T$  as in Proposition 10.

We close this section showing that the existence of solutions of  $\text{NCP}(T)$  is a necessary condition for convergence of the sequence  $\{x^k\}$  generated by (6-7). In fact, we will show that if the solution set  $S^*$  of  $\text{NCP}(T)$  is empty then  $\{x^k\}$  is unbounded.

LEMMA 8: *Suppose that all hypotheses of Theorem 3 hold, excepting for (ii). Let  $\{x^k\}$  be the sequence generated by (6)-(7). If  $S^* = \emptyset$ , then  $\{x^k\}$  is unbounded.*

*Proof:* Suppose that  $\{x^k\}$  is bounded. Hence there exists a bounded, closed and convex set  $B \subset \mathbf{R}_{++}^n$  such that  $\overline{\{x^k\}} \subset \text{int}(B)$ . Consider the operator  $\tilde{T} := T + N_B$ , where  $N_B$  is the normality operator associated with  $B$ . This operator is maximal monotone because  $\{x^k\} \subset \text{int}(B) \cap D(T)$  and Proposition 6 applies. Consider now the sequence  $\{\tilde{x}^k\}$ , generated using  $\tilde{T}$  instead of  $T$  in (7), and starting with  $x^0$  in (6) (the same initial point as  $\{x^k\}$ ). We claim that:

- 1)  $\{\tilde{x}^k\}$  is well defined and contained in  $\mathbf{R}_{++}^n$ ;
- 2)  $\{\tilde{x}^k\}$  converges to a solution of  $\text{NCP}(\tilde{T})$ ;
- 3)  $\tilde{x}^k = x^k$  for all  $k$ .

We proceed to establish these three claims.

- 1) We know that  $\tilde{T}$  is maximal monotone, and its domain is  $D(\tilde{T}) = D(T) \cap B$ , which is nonempty and bounded by boundedness of  $B$ . Then by Proposition 9, the operator  $\tilde{T} + \lambda_k \nabla_x d_\varphi(\cdot, x^k)$  is onto, which asserts existence of  $\tilde{x}^k$  for any  $k$ . By strict convexity of the function  $d_\varphi(\cdot, x^k)$ , the operator  $\tilde{T} + \lambda_k \nabla_x d_\varphi(\cdot, x^k)$  is strictly monotone, which yields uniqueness of each iterate. Hence, the sequence  $\{\tilde{x}^k\}$  is well defined and contained in  $D(\tilde{T}) = D(T) \cap B \subset \mathbf{R}_{++}^n$ . In particular, the nonemptiness of  $D(\tilde{T})$ , together with the last inclusion establishes condition (i) of Theorem 3.
- 2) For proving (2), we will check first that the hypotheses (ii-vi) of Theorem 3 hold (so we assure boundedness of  $\{\tilde{x}^k\}$  and hence existence of cluster points), and then we check paramonotonicity of  $\tilde{T}$  (so we have convergence):
  - (ii) the solution set  $\text{NCP}(\tilde{T}) = \{z \in \mathbf{R}^n \mid \tilde{T}(z) + N_{\mathbf{R}_+^n}(z) = 0\}$  is nonempty because  $D(\tilde{T} + N_{\mathbf{R}_+^n}) = D(\tilde{T}) \cap \mathbf{R}_+^n$  is bounded, hence by Proposition 9, the operator  $\tilde{T} + N_{\mathbf{R}_+^n}$  has zeroes.
  - (iii) It follows from Proposition 3(ii) that  $N_B$  is pseudomonotone. By assumption,  $T$  is pseudomonotone. Using now that the sum of two

pseudomonotone operators is pseudomonotone (see [20], p. 97), we conclude that  $\tilde{T} = T + N_B$  is pseudomonotone.

Conditions (iv, v, vi.1) do not depend on  $T$ , so they automatically hold.

(vi.2) It is enough to check that

$$h_{\tilde{T}, \mathbf{R}_+^n}(x) < \infty \text{ for all } x \in \mathbf{R}_+^n \cap D(\tilde{T}).$$

Indeed, as  $B \subset \mathbf{R}_+^n$ , it is easy to check that

$$G_{\mathbf{R}_+^n}(\tilde{T}) \subset \{(y, v) \mid y \in D(T) \cap \mathbf{R}_+^n, v \in \tilde{T}(y) = T(y) + N_B(y)\}.$$

Hence, for any  $(y, v) \in G_{\mathbf{R}_+^n}(\tilde{T})$ , it holds that there exist  $u \in T(y), w \in N_B(y)$  such that

$$v = u + w. \tag{42}$$

Take now  $(y, v) \in G_{\mathbf{R}_+^n}(\tilde{T}), x \in \mathbf{R}_+^n \cap D(\tilde{T}) \subset B$ , and  $u, w$  as in (42), we can write

$$\begin{aligned} \langle v, x - y \rangle &= \langle u + w, x - y \rangle = \langle u, x - y \rangle + \langle w, x - y \rangle \\ &\leq \langle u, x - y \rangle \leq h_{T, \mathbf{R}_+^n}(x), \end{aligned}$$

where we used in the first inequality the fact that  $x \in B$ , and that  $w \in N_B(y)$ . The second inequality follows from the definition of  $h_{T, \mathbf{R}_+^n}$ . Taking in the expression above the supremum on  $(y, v) \in G_{\mathbf{R}_+^n}(\tilde{T})$ , we obtain

$$h_{\tilde{T}, \mathbf{R}_+^n}(x) = \sup_{(y, v) \in G_{\mathbf{R}_+^n}(\tilde{T})} \langle v, x - y \rangle \leq h_{T, \mathbf{R}_+^n}(x),$$

and the rightmost term is finite by assumption, as  $x \in \mathbf{R}_+^n \cap D(\tilde{T}) \subset \mathbf{R}_+^n \cap D(T)$ .

Let us check now paramonotonicity: Since  $\tilde{T} = T + \partial(\delta_B)$ ,  $\tilde{T}$  is paramonotone by Propositions 2(i) and 2(iii).

Thus, all the assumptions of Theorem 4 hold for the sequence  $\{\tilde{x}^k\}$  defined by  $\tilde{T}$  and  $\tilde{x}^0 := x^0$ , so we conclude that it converges to a solution of NCP( $T$ ). This establishes (2).

3) We will prove by induction that  $x^k = \tilde{x}^k$  for all  $k$ .

The conclusion holds for  $k = 0$  by assumption. Suppose that  $x^k = \tilde{x}^k$ . By (7)

$$0 \in T(x^{k+1}) + \lambda_k \nabla_x d_\varphi(x^{k+1}, x^k) = \tilde{T}(x^{k+1}) + \lambda_k \nabla_x d_\varphi(x^{k+1}, \tilde{x}^k), \quad (42)$$

using the fact that  $T = \tilde{T}$  in  $\text{int}(B)$ , (recall that, by definition of  $N_B$ , we have that  $N_B(x) = \{0\}$  for all  $x \in \text{int}(B)$ ) and the inductive hypothesis. Then  $x^{k+1}$  is a solution of the equation in  $y$ ,

$$0 \in \tilde{T}(y) + \lambda_k \nabla_x d_\varphi(y, x^k). \quad (43)$$

As we remarked before, the operator  $\tilde{T} + \lambda_k \nabla_x d_\varphi(\cdot, x^k)$  is onto and strictly monotone. This implies that (43) has a unique solution. By (42),  $x^{k+1}$  solves (43). Recalling now that the iterate  $\tilde{x}^{k+1}$  is defined precisely as the unique solution of (43), we conclude that  $\tilde{x}^{k+1} = x^{k+1}$ . The induction step is complete.

Applying now Theorem 4, we can conclude that  $\{\tilde{x}^k\}$ , and therefore  $\{x^k\}$ , has cluster points and all of them are solutions of  $NCP(\tilde{T})$ . Let  $x^*$  be a cluster point of  $\{x^k\}$ . It follows from our assumption on  $\{x^k\}$  that  $x^* \in \text{int}(B)$ . Since  $T(x) = \tilde{T}(x)$  for all  $x \in \text{int}(B)$ , we get  $T(x^*) = \tilde{T}(x^*)$  and then it follows from (1) that  $x^*$  is also a solution of  $NCP(T)$ , in contradiction with our hypothesis. We conclude that  $\{x^k\}$  is unbounded.  $\square$

We summarize our results in the following theorem.

**THEOREM 5:** *Assume that all hypotheses of Theorem 4 hold, excepting for (ii). Let  $\{x^k\}$  be the sequence generated by (6)-(7). Then*

- i)  $\{x^k\}$  is convergent if and only if  $S^* \neq \emptyset$ .
- ii) If  $S^* \neq \emptyset$  then the limit  $x^*$  of  $\{x^k\}$  belongs to  $S^*$ .
- iii) If  $S^* = \emptyset$  then  $\{x^k\}$  is unbounded.

*Proof:*

- i) If  $S^* \neq \emptyset$ , then the result follows from Theorem 4. If the sequence converges, then it is bounded and it follows from Lemma 8 that  $S^*$  is nonempty.
- ii) Follows from Theorem 4.
- iii) Follows from Lemma 8.

ACKNOWLEDGEMENTS

We thank two anonymous referees for their multiple suggestions, which greatly improved the paper.



## REFERENCES

1. A. AUSLENDER and M. HADDOU, An interior-proximal method for convex linearly constrained problems and its extension to variational inequalities, *Math. Programming*, 1995, 71, p. 77-100.
2. L. M. BREGMAN, The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming, *U.S.S.R. Comput. Math. and Math. Phys.*, 1967, 7, n° 3, p. 200-217.
3. H. BREZIS, *Opérateurs Monotones Maximaux et Semi-groupes de Contractions dans les Espaces de Hilbert*, North Holland, Amsterdam.
4. H. BREZIS and A. HARAUX, Image d'une somme d'opérateurs monotones et applications, *Israel J. Math.*, 1976, 23, n° 2, p. 165-186.
5. F. E. BROWDER, Nonlinear operators and nonlinear equations of evolution in Banach spaces, *Proceedings of Symposia in Pure Mathematics, American Mathematical Society*, 1976, 18, n° 2.
6. R. S. BURACHIK and A. N. IUSEM, A generalized proximal point algorithm for the variational inequality problem in a Hilbert space, *SIAM J. Optim.*, 1998, 8, p. 197-216.
7. Y. CENSOR, A. N. IUSEM and S. A. ZENIOS, An interior point method with Bregman functions for the variational inequality problem with paramonotone operators, *Math. Programming*, 1998, 81, p. 373-400.
8. I. CSISZÁR, Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.*, 1967, 2, p. 299-318.
9. P. T. HARKER and J. S. PANG, Finite dimensional variational inequalities and nonlinear complementarity problems: A survey of theory, algorithms and applications, *Math. Programming*, 1990, 48, p. 161-220.
10. A. N. IUSEM, B. F. SVAITER and M. TEBoulLE, Entropy-like proximal methods in convex programming, *Math. Oper. Res.*, 1994, 19, p. 790-814.
11. A. N. IUSEM and M. TEBoulLE, Convergence rate analysis of nonquadratic proximal and augmented Lagrangian methods for convex and linear programming, *Math. Oper. Res.*, 1995, 20, p. 657-677.
12. A. N. IUSEM, On some properties of paramonotone operators, *J. Convex Analysis*, 1998, 5, p. 269-278.
13. S. KARAMARDIAN, Complementarity problems over cones with monotone and pseudomonotone maps, *J. Optim. Theory Appl.*, 1976, 18, p. 445-455.
14. D. KINDERLEHRER and G. STAMPACCHIA, *An Introduction to Variational Inequalities and Their Applications*, Academic Press, New York, 1980.
15. M. A. KRASNOSELSKII, Two observations about the method of successive approximations, *Uspekhi Mat. Nauk*, 1955, 10, p. 123-127.
16. B. LEMAIRE, The proximal algorithm, in *International Series of Numerical Mathematics*, J. P. Penot, Ed., Birkhauser, Basel, 1989, 87, p. 73-87.
17. B. MARTINET, Régularisation d'inéquations variationnelles par approximations successives, *Revue Française d'Informatique et Recherche Opérationnelle*, 1970, 2, p. 154-159.
18. B. MARTINET, Algorithmes pour la résolution de problèmes d'optimisation et minimax, Thèse d'État, Université de Grenoble, Grenoble, 1972.
19. Z. OPIAL, Weak convergence of the sequence of successive approximations for nonexpansive mappings, *Bull. Amer. Math. Soc. (N.S.)*, 1967, 73, p. 591-597.
20. D. PASCALI and S. SBURLAN, *Nonlinear Mappings of Monotone Type*, Editura Academiei, Bucarest, 1978.

21. R. T. ROCKAFELLAR, On the maximality of sums of nonlinear monotone operators, *Trans. Amer. Math. Soc.*, 1970, *149*, p. 75-88.
22. R. T. ROCKAFELLAR, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.*, 1976, *14*, p. 877-898.
23. R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, New Jersey, 1970.
24. M. TEBoulLE, Entropic proximal mappings with applications to nonlinear programming. *Math. Oper. Res.*, 1992, *17*, p. 97-116.
25. M. TEBoulLE, Convergence of proximal-like algorithms, *SIAM J. Optim.*, 1997, *7*, p. 1069-1083.