A proximal method with logarithmic barrier for nonlinear complementarity problems

Rolando Gárciga Otero* Alfredo Iusem †

November 6, 2014

Abstract

We study the proximal method with the regularized logarithmic barrier, originally stated by Attouch and Teboulle for positively constrained optimization problems, in the more general context of nonlinear complementarity problems with monotone operators. We consider two sequence generated by the method. We prove that one of them, called the ergodic sequence, is globally convergent to the solution set of the problem, assuming just monotonicity of the operator and existence of solutions; for convergence of the other one, called the proximal sequence, we demand some stronger property, like paramonotonicity of the operator or the so called "cut property" of the problem.

1 Introduction

Given a point-to-set operator $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ and a closed and convex set $C \subset \mathbb{R}^n$, the *variational inequality problem* $\mathrm{VIP}(T, C)$ consists of finding $x^* \in C$ such that $\langle u^*, x - x^*\rangle \geq 0$ for some $u^* \in T(x^*)$ and all $x \in C$.

When T is the subdifferential ∂f of a convex function $f : \mathbb{R}^n \to \mathbb{R}$, we are in the so-called optimization case, and $\mathrm{VIP}(T, C)$ reduces to the problem of minimizing f on C. When $C = \mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x_j \geq 0 \ (1 \leq j \leq n)\}$, i.e., C is the nonnegative orthant, $\mathrm{VIP}(T, C)$ reduces to the *nonlinear complementarity problem* $\mathrm{NCP}(T)$, which consists of finding $x^* \in \mathbb{R}^n, u^* \in T(x^*)$ such that $x^* \geq 0,\ \ u^* \geq 0,\ \ \langle u^*, x^*\rangle = 0$.

The proximal point algorithm, introduced in [25], is one of the main tools for solving variational inequality problems. It generates a sequence $\{x^k\} \subseteq C$, starting at any $x^0 \in C$, where x^{k+1} is a solution of $\mathrm{VIP}(T_k, C)$, with $T_k(x) := \lambda_k T(x) + x - x^k$ for some $\lambda_k > 0$. When T is maximal monotone, and $\mathrm{dom}(T) \cap \mathrm{int}(C) \neq \emptyset$, $\mathrm{VIP}(T_k, C)$ has a unique solution and the sequence $\{x^k\}$ converges

*Instituto de Economia da Universidade Federal de Rio de Janeiro, Avenida Pasteur 250, Rio de Janeiro, RJ, Brazil (**rgarciga@ie.ufrj.br**). The work of this author was partially supported by PRONEX-Optimization.

†Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ, 22460-320, Brazil (**iusp@impa.br**). Partially supported by CNPq grant no. 301280/86

to a solution of VIP(T, C) under the only assumption of existence of solutions of VIP(T, C) (see e.g. [27]).

Formulated in this way, the proximal point method is a regularization procedure, which at each step replaces T by the better conditioned operator T_k. Later on, the proximal method was modified so as to achieve also a penalization effect, generating sequences which stay in the interior of C (assumed to be nonempty). This effect have been attained through the replacement of the linear regularization term $x - x^k$ in the definition of T_k by another one involving e.g. Bregman distances with zone C (see [9], [10], [12], [13], [16], [23]), φ-divergences (see [19], [20], [21] for the optimization case, [5] for complementarity problems and [17] for saddle-point computations), or the logarithmic quadratic barrier (see [1] and [2]). The general idea consists of defining T_k so that it has a unique zero which belongs to the interior of C, in which case VIP(T_k, C) reduces to the unconstrained problem of finding the zero of T_k, and the iterative step take the form $x^{k+1} = T_k^{-1}(0)$. In all these cases, one starts with a "distance-like" differentiable function δ defined at least on int$(C)\times$ int(C), whose partial derivative with respect to the first argument, say δ', diverges on the boundary of C, and defines $T_k(x) = \lambda_k T(x) + \delta'(x, x^k)$, where $\{\lambda_k\}$ is a sequence of positive real numbers, called *regularization coefficients*, bounded away from 0.

We discuss next the convergence results for the above mentioned scheme, which depend on the properties of T, C and the choice of δ. The expected result is the convergence of the sequence given by

$$x^{k+1} = T_k^{-1}(0) \tag{1}$$

to a solution of VIP(T, C), whenever this problem has solutions. A blanket assumption is the maximal monotonicity of T. This hypothesis turns out to be sufficient in the unconstrained case (i.e. $C = \mathbb{R}^n$), when δ is taken as the square of the Euclidean distance, i.e. $\delta(x, y) = \|x - y\|^2$. This is the classical proximal point method analyzed e.g. in [27].

In the constrained case, i.e., when $C \neq \mathbb{R}^n$, the only known case for which convergence has been proved assuming just monotonicity of T and existence of solutions corresponds to the logarithmic-quadratic method of Auslender, Teboulle and Ben Tiba (see [2]), where $C = \mathbb{R}^n_+$, i.e., the nonnegative orthant of $\mathbb{R}^n$, and $\delta : \mathbb{R}^n_{++} \times \mathbb{R}^n_{++} \to \mathbb{R}_+$ is chosen as

$$\delta(x, y) = \frac{\nu}{2} \|x - y\|^2 + \mu \sum_{j=1}^{n} y_j^2 \varphi(y_j^{-1} x_j), \tag{2}$$

with $\varphi(r) = r - ln(r) - 1$ and $\mu, \nu > 0$ (here $\mathbb{R}^n_{++}, \mathbb{R}_+$ denote the interior of the nonnegative orthant and the nonnegative halfline respectively).

In all other constrained cases (i.e., with $C \neq \mathbb{R}^n$), it has not been possible to establish convergence of the sequence given by (1) without further assumptions (though no counterexample has been exhibited). In order to obtain the desired convergence results, three alternative strategies have been developed:

a) Imposing on T a condition stronger than monotonicity or pseudomonotonicity, e.g. *paramonotonicity* or the so-called *cut property* (which involves also the set C), whose formal definitions can be found in Section 2.

b) Perturbing the generated sequence; for instance, the step given by (1) becomes an auxiliary one ($y^k = T_k^{-1}(0)$), and then the next iterate x^{k+1} is taken as a certain convex combination of x^k and y^k ($\{x^k\}$ is the so-called *ergodic sequence*).

c) Imposing some regularity condition on C (or perhaps jointly on C and T).

The study of all the above mentioned methods follows one of the three strategies above. The convergence analysis of the proximal point method with Bregman distances or φ-divergences follow strategy (a), assuming hypotheses which imply the cut property. Strategy (b) has been used in [14] and [18], replacing the original proximal sequence by the ergodic one, and in [4] and [28], where a re-scalarization is performed. Strategy (c) has been applied in [24], where convergence is ensured assuming that the boundary of C contains no line segments, and in [18], where the operator T is assumed to be a saddle-point one (which is not paramonotone) but the problem VIP(T, C) is assumed to enjoy strict complementarity.

A modified version of the distance-like δ given by (2), was introduced in [1], where $C = \mathbb{R}^n_+$ and δ is taken as $d : \mathbb{R}^n_+ \times \mathbb{R}^n_{++} \to \mathbb{R}_+$ defined as

$$d(x, y) = \frac{\nu}{2}||x - y||^2 + \mu \sum_{j=1}^{n} y_j \varphi(y_j^{-1} x_j). \tag{3}$$

Note that the difference between (2) and (3) is the exponent of the first factor in each term of the summation (2 in (2), 1 in (3)). The method induced by d was studied in [1] for the optimization case, i.e. when T is the subdifferential of a convex function. Since subdifferentials of convex functions are paramonotone and enjoy the cut property, no further assumption was required in the convergence analysis of [1]. In this paper, we consider the regularization given by (3) with a general maximal monotone T, i.e. for the nonlinear complementarity problem.

We will apply strategies (a) and (b), establishing convergence of the proximal sequence under the cut property, and convergence of the ergodic sequence in the absence of this property.

We emphasize that this case has not been covered in any of the above mentioned papers (note e.g. that for $C = \mathbb{R}^n_+$ the boundary of C contains segments, so that we cannot consider an approach like the one in [24]). We also point out that the convergence analysis for the complementarity case differs substantially from the optimization case studied in [1]. When T is not the subdifferential of a convex function f, one cannot exploit the monotonicity properties of the sequence of functional values $\{f(x^k)\}$. Instead, we will invoke properties of the enlargements of monotone operators introduced in [6], like the *transportation formula* established in [7] (see Section 2).

2 Preliminaries

We denote as $\langle\cdot,\cdot\rangle$ the Euclidean inner product. We recall next several monotonicity-like properties of operators, needed for the convergence analysis.

Definition 2.1. A point-to-set operator $T:\mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ is said to be *monotone* if

$$\langle w-w', x-x'\rangle \geq 0, \text{ for all } x,\ x' \in \mathbb{R}^n \text{ and all } w\in T(x),\ w'\in T(x'),$$

where $\langle\cdot,\cdot\rangle$ denotes the Euclidean inner product. If the inequality above is strict whenever $x \neq x'$ then T is called *strictly monotone*.

Definition 2.2. A monotone operator is *maximal* if its graph,

$$G(T) = \{(x,v)\in \mathbb{R}^n\times\mathbb{R}^n : v\in T(x)\},$$

is not properly contained in the graph of any other monotone operator.

Definition 2.3. A monotone operator $T:\mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ is *paramonotone* (on its domain) if it satisfies:

$$\langle w-w', z-z'\rangle = 0, \quad w\in T(z), \quad w'\in T(z') \implies \quad w\in T(z'), \quad w'\in T(z).$$

We remark that strictly monotone operators are always paramonotone.

Definition 2.4. An operator $T:\mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ is *pseudomonotone* if it satisfies:

$$\langle w', z-z'\rangle \geq 0,\ w\in T(z),\ w'\in T(z') \implies \quad \langle w, z-z'\rangle \geq 0.$$

Moreover, a pseudomonotone operator is *pseudomonotone*$_*$ (see [11], [15], [24]), if

$$\langle w, z-z'\rangle = 0 = \langle w', z-z'\rangle,\ w\in T(z),\ w'\in T(z') \implies \exists k>0 \text{ such that } kw'\in T(z).$$

Note that any monotone operator is pseudomonotone and, by definition, paramonotone operators are monotone.

We recall next the *cut property* introduced by [11] for single-valued operators and in [15], [24] for set-valued ones, which is strongly connected to paramotonocity.

Definition 2.5. The variational inequality problem VIP(T,C) satisfies the *cut property* when for any solution x^* the condition

$$\hat{x}\in C, \quad \hat{v}\in T(\hat{x}), \quad \langle \hat{v}, x^*-\hat{x}\rangle \geq 0$$

implies that $\hat{x}$ is a solution too.

If the set-valued operator T is paramonotone on C, with C closed and convex, then $VIP(T,C)$ satisfies the cut property (see [17]); the same is true when T is pseudomonotone$_*$, as shown in [11] for single-valued operators and in Theorem 4.1(i) of [15] for set-valued ones. It was observed in [11] that pseudomonotonicity$_*$ is in a certain way a minimal condition ensuring that the cut property holds. This result was improved upon in Theorem 4.1(ii) of [15], where it was proved that among all pseudomonotone mappings on C, with C convex, those having convex and compact values, and satisfying the cut property on every compact subset of C, are pseudomonotone$_*$ on the interior of C.

We mention that the cut property cannot be expected to hold in variational inequalities associated to saddle point operators (see Remark 1.2 in [22]).

We close the section with some elements of the theory of enlargements of monotone operators, introduced in [6], also needed in our convergence analysis.

Definition 2.6. Given a monotone operator $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ and a nonnegative parameter ε, the $\varepsilon-$enlargement of T is the operator $\widetilde{T} : \mathbb{R}_+ \times \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ defined by

$$\widetilde{T}(\varepsilon, x) = \{u \in \mathbb{R}^n : \langle v - u, y - x\rangle \geq -\varepsilon,\ \forall (y, v) \in G(T)\}.$$

We will occasionally use the notation $T^\varepsilon(x)$ as an alternative to $\widetilde{T}(\varepsilon, x)$. We recall that when T is maximal monotone then the graph of $\widetilde{T}$ is closed (see Proposition 3.4 in [7]).

Next we state the so called *transportation formula.*

Proposition 2.7. *Given any set of m triplets $\{(\beta_i, x^i, v^i)\}_{i=1}^m \subseteq G(\widetilde{T})$ and any $\alpha \in \Delta_m := \{\alpha \in \mathbb{R}^m_+ : \sum_{i=1}^m \alpha_i = 1\}$, it holds that $\hat{\beta} \geq 0$ and that $(\hat{\beta}, \hat{x}, \hat{v})$ belongs to $G(\widetilde{T})$, where*

$$\hat{x} = \sum_{i=1}^m \alpha_i x^i,\ \ \hat{v} = \sum_{i=1}^m \alpha_i v^i \quad \text{and} \quad \hat{\beta} = \sum_{i=1}^m \alpha_i \left[\beta_i + \langle v^i - \hat{v}, x^i - \hat{x}\rangle\right].$$

Proof. See Theorem 3.11 in [7]. □

3 A proximal-like algorithm

We consider the following entropy proximal scheme for solving the nonlinear complementarity problem NCP(T), to be denoted as Algorithm EPNLC. It requires positive parameters μ and ν, a sequence of regularizing parameters $\{\lambda_k\} \subset [\bar{\lambda}, \infty)$ for some $\bar{\lambda} > 0$, and the auxiliary function $\varphi : \mathbb{R}_{++} \to \mathbb{R}_+$ defined by $\varphi(r) = r - ln(r) - 1$.

Algorithm EPNLC:

Initialization: Choose any $x^0 \in \mathbb{R}^n_{++}$ and define $y^0 = x^0$, $\sigma_0 = 0$.

Step 1: Given x^{k-1}, y^{k-1} and $\lambda_k > 0$, find $y^k \in \mathbb{R}^n$ and $v^k \in \mathbb{R}^n$ satisfying

$$v^k \in T(y^k), \tag{4}$$

$$\lambda_k v_j^k + \nu(y_j^k - y_j^{k-1}) + \mu\varphi'(y_j^k/y_j^{k-1}) = 0 \ \ (1 \le j \le n). \tag{5}$$

Step 2: Given σ_{k-1}, define $\sigma_k = \lambda_k + \sigma_{k-1}$,

$$x^k = \frac{\lambda_k}{\sigma_k} y^k + \left(1 - \frac{\lambda_k}{\sigma_k}\right) x^{k-1}. \tag{6}$$

Several comment on Algorithm EPNLC are in order. Observe first that the sequence $\{y^k\}$ generated by Algorithm EPNLC satisfies $y^k = T_k^{-1}(0)$ with

$$T_k = \lambda_k T + d'(\cdot, y^{k-1}) \tag{7}$$

and $d : \mathbb{R}^n_{++} \times \mathbb{R}^n_{++} \to \mathbb{R}_+$ given by (3), which is the regularized logarithmic barrier functional studied in [1], a variation of the log-quadratic given by (2), introduced in [2].

Note that when T is single-valued, Step 1 reduces to the solution of the system of n nonlinear equations given by (5) in the n unknowns $y_1^k, \ldots, y_n^k$, with $T(y_j^k)$ substituting for v_j^k. The solution of such system requires in general some auxiliary procedure, to be chosen according to the properties of T (e.g., one could use Newton method or a quasi-Newton one, if T is continuously differentiable). It follows that the computational burden of the iterative procedure is fully concentrated in Step 1, since Steps 2 entails just very elementary computations.

We will write now the system (4)-(5) in a more explicit way and compare it with the corresponding system for the method with the log-quadratic method of Auslender, Teboulle and Ben-Tiba presented in [2], which uses the function δ defined in (2). For the sake of simplicity, we carry out this comparison only for the case of a single-valued T. The system of equations in Step 1 of Algorithm EPNLC in the unknowns $y_1 \ldots, y_n \in \mathbb{R}$ can be written as:

$$\lambda_k T(y)_j + \left(\nu + \frac{\mu}{y_j}\right)(y^j - y_j^{k-1}) = 0 \ \ (1 \le j \le n), \tag{8}$$

while the corresponding system for the method in [2] is:

$$\lambda_k T(y)_j + \left(\nu + y_j^{k-1}\frac{\mu}{y_j}\right)(y^j - y_j^{k-1}) = 0 \ \ (1 \le j \le n). \tag{9}$$

The difference between both methods lies in the factor y_j^{k-1}, which multiplies μ in (9) and is absent in (8). Now, the expression $\nu(y_j - y_j^{k-1})$ is the regularization term ($\lambda_k T + \nu I$ is in general better conditioned than T) while the expressions $(\mu/y_j)(y^j - y_j^{k-1})$ in (8) and $y_j^{k-1}(\mu/y_j)(y^j - y_j^{k-1})$ in (9) are the penalization terms, which force the solution y of the system to be strictly positive. Consider now a component j such that $x^*_j = 0$ for all solution x^* of NCP(T) and assume that the sequence $\{y^k\}$ for any of the methods converges to a solution. Then $\lim_{k\to\infty} y^k_j = 0$ in both cases. The additional factor y^{k-1}_j multiplying μ in (9) makes the penalization term become negligible for large k as compared both to the regularization term and to the operator term $\lambda_k T(y)_j$ (remember that the λ_k's are bounded away from 0), and hence the same happens with the penalization effect, making the method in [2] numerically less stable than Algorithm EPNLC. Another attractive feature of the penalization function d given by (3) is that in its continuous version it induces a Lotka-Volterra dynamical system, enjoying several interesting properties, as discussed in [1].

The sequence $\{y^k\}$ generated by Algorithm EPNLC will be dubbed the *proximal sequence*, while the sequence $\{x_k\}$ will be called *ergodic*. The term "ergodic" is due to the following equalities, which follow easily from the definitions of σ_k and x^k in Step 2 of the algorithm, stated next for future reference:

$$\sigma_k = \sum_{\ell=1}^{k} \lambda_\ell, \tag{10}$$

$$x^k = \sum_{\ell=1}^{k} \rho_{\ell,k} y^\ell, \tag{11}$$

$$\sum_{\ell=1}^{k} \rho_{\ell,k} = 1, \tag{12}$$

with

$$\rho_{\ell,k} = \frac{\lambda_\ell}{\sigma_k} \qquad (1 \le \ell \le k). \tag{13}$$

Observe that $\lim_{k\to\infty} \sigma_k = +\infty$, and that $\lim_{k\to\infty} \rho_{\ell,k} = 0$ for any fixed ℓ. It follows that if y^k converges, say to $\bar{y}$, then $\{x^k\}$ also converges to $\bar{y}$. Hence, the convergence analysis of $\{x^k\}$ is relevant only when convergence of $\{y^k\}$ cannot be ensured.

Assuming maximal monotonicity of T and existence of solutions of NCP(T), we will prove that the ergodic sequence is bounded and all its cluster points are solutions of NCP(T). We will also prove that the the proximal sequence converges to a solution of NCP(T), but with a stronger assumption on the problem, namely the cut property of NCP(T) (or paramonotonicity of T). Observe that, since x^{k-1} is not required for the computation of y^k, under the assumptions which imply the convergence of $\{y^k\}$ one can exclude the sequence $\{x^k\}$ from the algorithm (i.e., eliminate Step 2). However, as already mentioned, this exclusion will have a negligible effect on the computational performance of the method.

We prove next that Algorithm EPNLC is well defined.

Theorem 3.1. *Assume that T is maximal monotone and $D(T) \cap \mathbb{R}^n_{++} \neq \emptyset$. Then, for any given sequence $\{\lambda_k\} \subseteq \mathbb{R}_{++}$ and any initial point $x^0 \in \mathbb{R}^n_{++}$, Algorithm EPNLC is well defined, i.e., at iteration $k \ge 1$, given $x^{k-1}, y^{k-1} \in \mathbb{R}^n_{++}$, there exists a unique $y^k \in \mathbb{R}^n_{++}$ and $v^k \in \mathbb{R}^n$ satisfying*

(4) *and* (5) *and, a unique* $x^k \in \mathbb{R}^n_{++}$ *satisfying* (6). *Moreover, if* $y^k = y^{k-1}$ *then* $v^k = 0$ *and* y^k *is a zero of* T*, and, in particular, a solution of NCP*(T).

Proof. We proceed by induction, assuming that the method has found $y^{k-1} \in \mathbb{R}^n_{++}$. We must prove that there exists a unique $y^k \in \mathbb{R}^n_{++}, v^k \in T(y^k)$ satisfying (4) and (5). As we have observed, this is equivalent to proving that $T_k^{-1}(0)$ is a singleton in $\mathbb{R}^n_{++}$, with T_k as in (7). Define now $h_k : \mathbb{R}^n \to \mathbb{R}$ as

$$h_k(y) = \begin{cases} \sum_{j=1}^n y_j^{k-1} \varphi\left(\frac{y_j}{y_j^{k-1}}\right) & \text{if } y \in \mathbb{R}^n_{++} \\ +\infty & \text{otherwise.} \end{cases}$$

Note that h_k is convex by convexity of φ, and therefore its subdifferential ∂h_k is maximal monotone. Since λ_k and μ are positives, T is maximal monotone and $D(T) \cap \mathbb{R}^n_{++} \neq \emptyset$, a well known result in convex analysis ensures that the operator $\widehat{T}_k : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ defined as $\widehat{T} = \lambda_k T + \mu \partial h_k$ is maximal monotone.

It is easy to check that the equation $y = T_k^{-1}(0)$ is equivalent to the inclusion

$$\nu y^{k-1} \in (\widehat{T}_k + \nu I)^{-1}(y). \tag{14}$$

We invoke now Minty's Theorem (see [26]), which states that if $U : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ is maximal monotone and I is the identity operator in $\mathbb{R}^n$, then the operator $U + \gamma I$ is onto and its inverse is point-to-point for all $\gamma \in \mathbb{R}_{++}$. Since $\nu > 0$ and $\widehat{T}_k$ is maximal monotone, Minty's Theorem implies that there exists a unique y satisfying (14), and so y^k is uniquely determined by (4)-(5). Moreover, the definition of φ ensures that y^k belongs to $\mathbb{R}^n_{++}$, and so does x^k, in view of (6). Hence, x^k is well defined and belongs to $\mathbb{R}^n_{++}$. Finally, $\varphi'(1) = 0$, so that $y^k = y^{k-1}$ implies $v^k = 0$. □

4 Convergence Analysis

We recall next a result on the regularized Kullback-Leibler entropy functional $E : \mathbb{R}^n_+ \times \mathbb{R}^n_{++}$, defined as

$$E(x, y) = \frac{\nu}{2} \|x - y\|^2 + \mu \sum_{j=1}^p y_j \psi(x_j / y_j), \tag{15}$$

where $\psi : \mathbb{R}_+ \to \mathbb{R}$ is defined as $\psi(r) = rln(r) - r + 1$ if $r > 0$, $\psi(0) = 1$, so that ψ is continuous on $\mathbb{R}_+$.

Lemma 4.1. *For any* $x \in \mathbb{R}^n_+$ *and* $y,\ z \in \mathbb{R}^n_{++}$ *it holds that*

$$E(x, z) - E(x, y) \geq \nu \langle x - y, y - z \rangle + \frac{\nu}{2} \|y - z\|^2 + \mu \sum_{j=1}^p \left[x_j \varphi'(y_j / z_j) + z_j - y_j \right].$$

Proof. See equations (15) and (16) in Theorem 6.1 of [1] □

We state next some basic properties of the sequences generated by Algorithm EPNLC.

Proposition 4.2. *Assume that $\{x^k\}, \{y^k\}$ and $\{v^k\}$ are the sequences generated by the Algorithm EPNLC. Define*

$$w^k = \sum_{\ell=1}^{k} \rho_{\ell,k} v^\ell, \tag{16}$$

with $\rho_{\ell,k}$ as defined by (13). Then,

i)

$$E(x, y^{k-1}) - E(x, y^k) \geq \lambda_k \langle v^k, y^k - x\rangle + \frac{\nu}{2} \left\| y^k - y^{k-1} \right\|^2 \geq \lambda_k \langle v^k, y^k - x\rangle, \tag{17}$$

$$\langle w^k, x\rangle \geq \sum_{\ell=1}^{k} \rho_{\ell,k} \langle v^\ell, y^\ell\rangle - \frac{1}{\sigma_k} E(x, x^0) \tag{18}$$

for all $x \in \mathbb{R}^n_+$.

ii) If T is monotone then

$$E(x, y^{k-1}) - E(x, y^k) \geq \lambda_k \langle v, y^k - x\rangle, \tag{19}$$

$$\langle v, x^k - x\rangle \leq \frac{1}{\sigma_k} E(x, x^0) \tag{20}$$

for all $x \in \mathbb{R}^n_+$ and all $v \in T(x)$.

iii) If T is maximal monotone then $w^k \in T^{\beta_k}(x^k)$, with T^{β_k} as in Definition 2.6, where

$$\beta_k = \sum_{\ell=1}^{k} \rho_{\ell,k} \langle w^k - v^\ell, x^k - y^\ell\rangle \geq 0. \tag{21}$$

Proof. Multiplying the j-th equation in (5) by $x_j - y^k_j$ and summing with j between 1 and n, we get from (5) that

$$\lambda_k \langle v^k, x - y^k\rangle + \nu \langle y^k - y^{k-1}, x - y^k\rangle + \mu \sum_{j=1}^{n} \varphi'(y^k_j / y^{k-1}_j)(x_j - y^k_j) = 0 \tag{22}$$

for all $x \in \mathbb{R}^n$. Using (22) and Lemma 4.1 with the choice $y = y^k$ and $z = y^{k-1}$, we obtain (17). Looking now at (17) with ℓ substituting for k, and summing then with ℓ between 1 and k, we obtain (18), taking into account the nonnegativity of $E(x, y^k)$, (10), (13) and (16), thus completing the proof of item (i).

Note that (19) is a direct consequence of the monotonicity of T, because $v^k \in T(y^k)$ and $v \in T(x)$, so that $\langle v^k - v, y^k - x\rangle \geq 0$ and hence $\langle v^k, y^k - x\rangle \geq \langle v, y^k - x\rangle$. Substituting ℓ for k in (19) and summing with ℓ from 1 to k, we get

$$E(x, x^0) - E(x, y^k) \geq \sum_{\ell=1}^{k} \lambda_\ell \langle v, y^\ell - x\rangle = \left\langle v, \sum_{\ell=1}^{k} \lambda_\ell y^\ell \right\rangle - \sigma_k \langle v, x\rangle. \tag{23}$$

It follows from (23) and (11) that

$$\frac{1}{\sigma_k} E(x, x^0) \geq \left\langle v, \sum_{\ell=1}^{k} \rho_{\ell,k} y^\ell \right\rangle - \langle v, x\rangle = \langle v, x^k - x\rangle,$$

so that (20) holds, completing the proof of item (ii). Finally, we prove item (iii). Since $v^\ell \in T(y^\ell) = T^0(y^\ell)$, we apply the transportation formula of Proposition 2.7 to the set of k triplets $\{(0, y^\ell, v^\ell)\}_{\ell=1}^{k}$, with $\alpha = (\rho_{1,k} \cdots, \rho_{k,k}) \in \Delta_k$, and we conclude that $w^k \in T^{\beta_k}(x^k)$ with

$$\beta_k = \sum_{\ell=1}^{k} \rho_{\ell,k} \left[0 + \langle w^k - v^\ell, x^k - y^\ell\rangle \right] \geq 0.$$

□

Now we will show that the algorithm generates bounded sequences whenever NCP(T) has solutions and, as it is usual with proximal point methods, the difference between consecutive iterates $y^k - y^{k-1}$ converges to zero. Moreover, the entropy functional of y^k with respect to any solution x^*, namely $E(x^*, y^k)$, is nonincreasing. We will denote by S the set of solutions of NCP(T).

Corollary 4.3. *Assume that T is maximal monotone and $S \neq \emptyset$. Then, the sequences $\{y^k\}, \{x^k\}$ generated by Algorithm EPNLC are bounded, the sequence $\{y^k - y^{k-1}\}$ converges to zero and the sequence $\{E(x^*, y^k)\}$ is nonincreasing and convergent for any $x^* \in S$. Moreover, for any $v^* \in T(x^*)$,*

$$0 \leq \sum_{k=1}^{\infty} \lambda_k \langle v^*, y^k - x^*\rangle \leq \sum_{k=1}^{\infty} \lambda_k \langle v^k, y^k - x^*\rangle < \infty$$

and

$$\sum_{k=1}^{\infty} \left\| y^k - y^{k-1} \right\|^2 < \infty.$$

Proof. Take any $x^* \in S$ and any $v^* \in T(x^*)$, and apply Proposition 4.2(i) and monotonicity of T in order to obtain

$$E(x^*, y^{k-1}) - E(x^*, y^k) \geq \lambda_k \langle v^k, y^k - x^*\rangle + \frac{\nu}{2} \left\| y^k - y^{k-1} \right\|^2 \geq \lambda_k \langle v^*, y^k - x^*\rangle \geq 0, \tag{24}$$

where the last inequality follows from the facts that $y^k \in \mathbb{R}^n_+$, $x^* \in S$ and $\lambda_k > 0$. Then, $E(x^*, y^{k-1}) \geq E(x^*, y^k) \geq \frac{\nu}{2} \left\|x^* - y^k\right\|^2 \geq 0$ for all k. Therefore, $\{E(x^*, y^k)\}$ is nonincreasing and nonnegative, hence convergent, and $\{y^k\}$ is bounded, since $\left\|x^* - y^k\right\|^2 \leq \frac{2}{\nu} E(x^*, y^0)$. Boundedness of $\{x^k\}$ follows then from (11) and (12). Finally, we invoke (24) in order to get

$$E(x^*, y^0) - \lim_{k\to\infty} E(x^*, y^k) \geq \sum_{k=1}^{\infty} \left[\lambda_k \langle v^k, y^k - x^* \rangle + \frac{\nu}{2} \left\|y^k - y^{k-1}\right\|^2\right] \geq 0.$$

Thus,

$$\sum_{k=1}^{\infty} \lambda_k \langle v^*, y^k - x^* \rangle \leq \sum_{k=1}^{\infty} \lambda_k \langle v^k, y^k - x^* \rangle < \infty$$

and

$$\sum_{k=1}^{\infty} \left\|y^k - y^{k-1}\right\|^2 < \infty.$$

Summability of $\left\|y^k - y^{k-1}\right\|^2$ implies then that the sequence $\{y^k - y^{k-1}\}$ converges to 0. □

Next we prove convergence of the proximal sequence $\{y^k\}$, assuming that the cut property holds.

Theorem 4.4. *Assume that T is maximal monotone, that $S \neq \emptyset$, that $\mathbb{R}^n_+ \subset int(D(T))$ and that the cut property in Definition 2.5 holds. Then, the sequence $\{y^k\}$ generated by Algorithm EPNLC converges to a solution of NCP(T).*

Proof. The sequences $\{v^k\}, \{y^k\}$ and $\{x^k\}$ are well defined by Theorem 3.1. Corollary 4.3, together with the assumptions of monotonicity of T and existence of solutions, ensures that $\{y^k\}$ is bounded, that $\{E(x^*, y^k)\}$ converges for all $x^* \in S$ and that

$$0 \leq \sum_{k=1}^{\infty} \lambda_k \langle v^k, y^k - x^* \rangle < \infty. \tag{25}$$

Since $\lambda_k \geq \bar{\lambda} > 0$ for all k, we get from (25) that

$$\lim_{k\to\infty} \langle v^k, y^k - x^* \rangle = 0. \tag{26}$$

Let now y^∞ be a cluster point of the bounded sequence $\{y^k\}$, say the limit of a subsequence $\{y^{k_i}\}$. Note that $v^{k_i} \in T(y^{k_i})$ for all i and that $y^\infty \in \mathbb{R}^n_+ \subset \mathrm{int}(D(T))$. Since maximal monotone operators are locally bounded in the interior of their domains, we conclude that $\{v^{k_i}\}$ is bounded and has therefore some cluster point v^∞. Since maximal monotonicity of T also implies that $G(T)$ is closed, we get that v^∞ belongs to $T(y^\infty)$. Moreover, in view of equation (26), $\langle v^\infty, x^* - y^\infty \rangle = 0$. Thus, by the cut property assumption, y^∞ is a solution of NCP(T). Hence, $\{E(y^\infty, y^k)\}$ converges too, and

$$\lim_{k\to\infty} E(y^\infty, y^k) = \lim_{i\to\infty} E(y^\infty, y^{k_i}) = \lim_{i\to\infty} \frac{\nu}{2} \left\| y^\infty - y^{k_i} \right\|^2 + \mu \sum_{j=1}^{n} y_j^\infty \psi(y_j^\infty / y_j^{k_i}) = 0,$$

which easily implies that $\lim_{k\to\infty} \left\| y^\infty - y^k \right\|^2 = 0$, so that the whole sequence $\{y^k\}$ converges to a unique cluster point which solves NCP(T). □

The next proposition shows that our convergence result holds for the wider class of pseudomonotone operators if we assume, instead of maximal monotonicity of T and $\mathbb{R}^n_+ \subset \text{int}(D(T))$, that Algorithm EPNLC is well defined and that
(AC): If $\{z^k\}$ is bounded and $(z^k, u^k) \in G(T)$ then for any accumulation point z^∞ of $\{z^k\}$ there exists some accumulation point u^∞ of $\{u^k\}$ such that $(z^\infty, u^\infty) \in G(T)$.

Note that assumption (AC) holds true, e.g., when the range of T is compact or when T is locally bounded in bounded sets.

Proposition 4.5. *Assume that T is pseudomonotone and verifies (AC), that $S \neq \emptyset$, that the cut property in Definition 2.5 holds and that Algorithm EPNLC generates an infinite sequence (y^k, v^k) satisfying* (4) *and* (5)*. Then, the sequence $\{y^k\}$ converges to a solution of NCP(T).*

Proof. Note that (17) in Proposition 4.2(i) demands no monotonicity assumptions on T. Thus, for any $x^* \in S$ and some $v^* \in T(x^*)$ we have

$$E(x^*, y^{k-1}) - E(x^*, y^k) \geq \lambda_k \langle v^k, y^k - x^* \rangle + \frac{\nu}{2} \left\| y^k - y^{k-1} \right\|^2 \geq \lambda_k \langle v^k, y^k - x^* \rangle, \tag{27}$$

and, since $y^k \in \mathbb{R}^n_{++} \subseteq \mathbb{R}^n_+$,

$$\langle v^*, y^k - x^* \rangle \geq 0. \tag{28}$$

Now, (4) and pseudomonotonicity of T imply

$$\langle v^k, y^k - x^* \rangle \geq 0. \tag{29}$$

From (27)

$$E(x^*, y^{k-1}) - E(x^*, y^k) \geq \lambda_k \langle v^k, y^k - x^* \rangle + \frac{\nu}{2} \left\| y^k - y^{k-1} \right\|^2 \geq \lambda_k \langle v^k, y^k - x^* \rangle \geq 0. \tag{30}$$

From this point on we can follow the proof of Corollary 4.3 in order to ensure that $\{E(x^*, y^k)\}$ is nonincreasing and convergent, that $\{y^k\}$ is bounded, that the sequence $\{y^k - y^{k-1}\}$ converges to zero, and that

$$0 \leq \sum_{k=1}^{\infty} \lambda_k \langle v^k, y^k - x^* \rangle < \infty \qquad \sum_{k=1}^{\infty} \left\| y^k - y^{k-1} \right\|^2 < \infty.$$

Thus,

$$\lim_{k\to\infty} \langle v^k, y^k - x^* \rangle = 0.$$

Choose now any cluster point y^∞ of the bounded sequence $\{y^k\}$. Then, by assumption (AC), there exists a cluster point v^∞ of $\{v^k\}$ with $v^\infty \in T(y^\infty)$. Thus, $\langle v^\infty, y^\infty - x^* \rangle = 0$, and by the cut property we have $y^\infty \in S$. Hence, as in Theorem 4.4, we can conclude that $\{E(y^\infty, y^k)\}$ converges to 0, implying that $\{y^k\}$ converges to a solution of NCP(T). □

Next we establish the convergence properties of the ergodic sequence $\{x^k\}$, without assuming the cut property.

Theorem 4.6. *Assume that T is maximal monotone, that $S \neq \emptyset$, and that $\mathbb{R}^n_+ \subset int(D(T))$. Then the sequences $\{x^k\}$, $\{w^k\}$ generated by Algorithm EPNLC are bounded and every cluster point (x^∞, w^∞) of $\{(x^k, w^k)\}$ provides a solution of NCP(T), i.e., $x^\infty, w^\infty \in \mathbb{R}^n_+$, $\langle x^\infty, w^\infty \rangle = 0$ and $w^\infty \in T(x^\infty)$.*

Proof. The sequences $\{v^k\}, \{y^k\}$ and $\{x^k\}$ are well defined by Theorem 3.1. By (4) and (5), we have $v^k \in T(y^k)$ and

$$\lambda_k v_j^k + \nu(y_j^k - y_j^{k-1}) + \mu\left(1 - y_j^{k-1}/y_j^k\right) = 0 \quad (1 \le j \le n),$$

implying

$$v_j^k y_j^k = \lambda_k^{-1}\left(y_j^{k-1} - y_j^k\right)\left(\mu + \nu y_j^k\right) \quad (1 \le j \le n). \tag{31}$$

Corollary 4.3, together with the maximal monotonicity of T and the nonemptiness of S, ensures that $\{y^k\}$ and $\{x^k\}$ are bounded and that $\{y^k - y^{k-1}\}$ converges to zero. Since $\lambda_k \ge \bar{\lambda} > 0$, and $\lim_{k\to\infty}(y_j^{k-1} - y_j^k) = 0$, (31) implies that

$$\lim_{k\to\infty} v_j^k y_j^k = 0 \quad (1 \le j \le n). \tag{32}$$

By (32)

$$\lim_{k\to\infty} \langle v^k, y^k \rangle = 0 \quad \text{and} \quad \lim_{k\to\infty} \sum_{\ell=1}^{k} \rho_{\ell,k} \langle v^\ell, x^\ell \rangle = 0. \tag{33}$$

Since T is maximal monotone, it is locally bounded in the interior of its domain, and hence the sequences $\{v^k\}$ and $\{w^k\}$, as defined by (16), are bounded too, taking into account (12). Apply now inequality (18) in Proposition 4.2(i) to (33) in order to obtain that

$$\liminf_{k\to\infty} \langle w^k, x \rangle \ge 0 \tag{34}$$

for all $x \in \mathbb{R}^n_+$. Let w^∞ be a cluster point of $\{w^k\}$. By (34), $\langle w^\infty, x \rangle \ge 0$ for all $x \in \mathbb{R}^n_+$, and consequently

$$w^\infty \in \mathbb{R}^n_+. \tag{35}$$

Note that $x^k \in \mathbb{R}^n_+$ by Theorem 3.1 and (12). Moreover, since $\{E(x^k, x^0)\}$ is bounded too, we can apply inequality (18) in Proposition 4.2(i) with $x = x^k \in \mathbb{R}^n_+$, and conclude that

$$\liminf_{k\to\infty} \langle w^k, x^k\rangle \geq 0. \tag{36}$$

Recall that β_k as defined in (21), is nonnegative by Proposition 4.2(iii). Thus, in view of (11), (12) and (16),

$$0 \leq \beta_k = \sum_{\ell=1}^{k} \rho_{\ell,k}\langle w^k - v^\ell, x^k - y^\ell\rangle \ = \sum_{\ell=1}^{k} \rho_{\ell,k}\left[\langle w^k, x^k\rangle - \langle w^k, y^\ell\rangle - \langle v^\ell, x^k\rangle + \langle v^\ell, x^\ell\rangle\right]$$

$$= \langle w^k, x^k\rangle \sum_{\ell=1}^{k} \rho_{\ell,k} + \sum_{\ell=1}^{k} \rho_{\ell,k}\langle v^\ell, y^\ell\rangle - 2\langle w^k, x^k\rangle = \sum_{\ell=1}^{k} \rho_{\ell,k}\langle v^\ell, y^\ell\rangle - \langle w^k, x^k\rangle. \tag{37}$$

By (37), $\langle w^k, x^k\rangle \leq \sum_{\ell=1}^k \rho_{\ell,k}\langle v^\ell, y^\ell\rangle$, and therefore, in view of (33),

$$\limsup_{k\to\infty} \langle w^k, x^k\rangle \leq 0. \tag{38}$$

Hence, combining (36) and (38), we have

$$\lim_{k\to\infty} \langle w^k, x^k\rangle = 0. \tag{39}$$

From (39), (33) and (37) we obtain that

$$\lim_{k\to\infty} \beta_k = 0. \tag{40}$$

Consider now any cluster point (x^∞, w^∞) of the bounded sequence $\{(x^k, w^k)\}$. By Proposition 4.2(iii), $w^k \in \widetilde{T}(\beta_k, x^k)$ for all k. As mentioned in Section 2, the graph of $\widetilde{T}$ is closed, so that, taking limits along an appropriate subsequence, we get, in view of (40),

$$w^\infty \in \widetilde{T}(0, x^\infty) = T(x^\infty). \tag{41}$$

Taking limits along the same subsequence in (39), we get

$$\langle w^\infty, x^\infty\rangle = 0. \tag{42}$$

On the other hand, since $x^k \in \mathbb{R}^n_{++}$ for all k, we get that

$$x^\infty \in \mathbb{R}^n_+. \tag{43}$$

This fact, together with (41), (42) and (35), entails that x^∞ is a solution of NCP(T), as required. $\square$

We consider next the case in which $D(T) \subseteq \mathbb{R}^n_+$, excluded in the hypotheses of Theorem 4.6, and show that the cluster points of the ergodic sequence are also solutions of NCP(T); actually, in this case, they are zeroes of T. Note that if T is maximal monotone and $D(T) \subset C$ then any solution of VIP(T, C) is a zero of T, so that $S \neq \emptyset$ implies $T^{-1}(0) \neq \emptyset$.

Theorem 4.7. *Assume that T is maximal monotone, that $S \neq \emptyset$, and that $D(T) \subseteq \mathbb{R}^n_+$. Then the sequence $\{x^k\}$ generated by Algorithm EPNLC is bounded and all its cluster points are zeroes of T.*

Proof. By Corollary 4.3, $\{x^k\}$ is bounded. Moreover, in view of (20) in Proposition 4.2(ii), we have

$$\langle v, x^k - x \rangle \leq \frac{1}{\sigma_k} E(x, x^0)$$

for all $x \in \mathbb{R}^n_+$ and all $v \in T(x)$. Thus, since $\lim_{k\to\infty} \sigma_k = +\infty$, it follows that

$$\limsup_{k\to\infty} \langle v, x^k - x \rangle \leq 0$$

for all $x \in \mathbb{R}^n_+$ and all $v \in T(x)$, and hence we conclude that

$$\langle v, x^\infty - x \rangle \leq 0$$

for all $x \in \mathbb{R}^n_+$, all $v \in T(x)$ and all cluster points x^∞ of $\{x^k\}$. Thus, since $D(T) \subseteq \mathbb{R}^n_+$ by assumption, we get that

$$\langle v - 0, x - x^\infty \rangle \geq 0$$

for all $(x, v) \in G(T)$. Finally, we apply maximal monotonicity of T in order to conclude that $0 \in T(x^\infty)$. □

Remark 4.8. We remark that the statement of Theorem 4.6 regarding the sequence $\{x^k\}$ (i.e., the fact that all its cluster points solve NCP(T)), can be obtained as a simple corollary of Theorem 4.7: it suffices to apply Theorem 4.7 to the operator $\bar{T}$ defined as $\bar{T} = T + N_C$ with $C = \mathbb{R}^n_+$, where N_C denotes the normal cone operator associated to a closed and convex $C \subset \mathbb{R}^n$. Indeed $T + N_C$ is maximal monotone by maximal monotonicity of T, its domain is contained in $\mathbb{R}^n_+$, and it coincides with T in $\mathbb{R}^n_{++}$, so that the sequences obtained by applying Algorithm EPNLC to T coincide with the corresponding sequences for $\bar{T}$, because, since $\{y^k\} \subset \mathbb{R}^n_{++}$, we have that the condition $v^k \in T(y^k)$ given by (4) is equivalent to $v^k \in \bar{T}(y^k)$. Theorem 4.7 applied to $\bar{T}$ ensures that the cluster points of $\{x^k\}$ are zeroes of $\bar{T}$, but it is elementary to check that the zeroes of $\bar{T}$ are the x-components of the solutions of NCP(T). On the other hand, Theorem 4.6 gives also information on the behavior of the sequence $\{w^k\}$, and provides a solution of NCP(T) in the form of the full pair (x^∞, w^∞), instead of just the x-component of the pair. Thus, we considered it worthwhile to present both theorems, with their proofs.

Remark 4.9. We comment that our convergence analysis can be easily extended to an inexact version of Algorithm EPNLC, allowing for an approximate solution of the system (4)–(5). This approximation consists of replacing T_k by a suitable enlargement, in the sense of Definition 2.6. More precisely, we can replace (4) by $v^k \in T^{\varepsilon_k}(y^k)$. In view of Proposition 2.7, the results in Theorems 4.6 and 4.7 will remain valid for a sufficiently small ε_k, for instance such that $\varepsilon_k \leq \nu/(2\lambda_k) \left\|y^k - y^{k-1}\right\|^2$.

Remark 4.10. We make now some comments on the issue of the convergence properties of the proximal sequence $\{y^k\}$ in the absence of the cut property. Proposition 4.2 establishes that $\{y^k\}$ is bounded, contained in $\mathbb{R}^n_{++}$, that $\{v^k\}$ is bounded and that $\lim_{k\to\infty}(y^k - y^{k-1}) = 0$. It is easy to check that if (y^∞, v^∞) is a cluster point of $\{(y^k, v^k)\}$, then it holds that $y^\infty \geq 0$ and $\langle y^\infty, v^\infty\rangle = 0$. However, it has not been possible to prove that v^∞ is nonnegative. Indeed, if the sequence $\{y^k\}$ has a unique cluster point, then it can be established that $v^\infty \geq 0$, but the possibility that $\{v^k\}$ oscillates between two cluster points laying outside the positive orthant has not been excluded. Our analysis shows that this improper behavior cannot occur for the ergodic sequences $\{(x^k, w^k)\}$, due to the averaging effect resulting from (16). As observed above, the possible lack of convergence of $\{y^k\}$ is irrelevant for any practical purposes.

Remark 4.11. Another open issue is the uniqueness of the cluster points of $\{x^k\}$, again in the absence of the cut property. We conjecture that the whole sequence is indeed convergent to a solution of NCP(T), but we have found no proof of this fact.

References

[1] Attouch, H., Teboulle, M. Regularized Lotka-Volterra dynamical system as a continuous proximal-like method in optimization. *Journal of Optimization Theory and Applications* **121** (2004) 541–570.

[2] Auslender, A., Teboulle, M., Ben-Tiba, S. A logarithmic-quadratic proximal method for variational inequalities. *Computational Optimization and Applications* **12** (1999) 31–40.

[3] Bruck, R.E. Corrigendum: An iterative solution of a variational inequality for certain monotone operators in Hilbert space. *Bulletin of the American Mathematical Society* **82** 2 (1976) 353.

[4] Burachik, R., Dutta J. Inexact proximal point methods for variational inequality problems. *SIAM Journal on Optimization* **20** (2010) 2653–2678.

[5] Burachik, R.S., Iusem, A.N. A generalized proximal point algorithm for the nonlinear complementarity problem. *RAIRO Recherche Opérationalle* **33** (1999) 447–479.

[6] Burachik, R.S., Iusem, A.N., Svaiter, B.F. Enlargement of maximal monotone operators with application to variational inequalities. *Set Valued Analysis* **5** (1997) 159–180.

[7] Burachik, R.S., Svaiter, B.F. $\varepsilon-$ enlargement of maximal monotone operators in Banach spaces. *Set-Valued Analysis* **7** (1999) 117–132.

[8] Censor, Y., Iusem, A.N., Zenios, S.A. An interior point method with Bregman functions for the variational inequality problem with paramonotone operators. *Mathematical Programming* **81** (1998) 373–400.

[9] Censor, Y., Zenios, S.A. Proximal minimization algorithm with D-functions. *Journal of Optimation Theory and Applications* **73** (1992) 451–464.

[10] Chen, G., Teboulle, M. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization* **3** (1993) 538–543.

[11] Crouzeix, J.P., Marcotte, P., Zhu, D. Conditions ensuring the applicability of cutting-plane methods for solving variational inequalities. *Mathematical Programing* **88** (2000) 521–539.

[12] Eggermont, P.P.B. Multiplicative iterative algorithms for convex programming. *Linear Algebra and Its Applications* **130** (1990) 25–42.

[13] Erlander, S. Entropy in linear programs. *Mathematical Programming* **21** (1981) 137–151.

[14] Güler, O. Ergodic convergence in proximal point algorithms with Bregman functions. In *Advances in Optimization and Approximation*, Kluwer, Dordrecht (1994) 155–165.

[15] Hadjisavvas, N., Schaible, S. On a generalization of paramonotone maps and its application to solving the Stampacchia variational inequality. *Optimization* **55** (2006) 593–604.

[16] Iusem, A.N. Some properties of generalized proximal point methods for quadratic and linear programming. *Journal of Optimization Theory and Applications* **85** (1995) 593–612.

[17] Iusem, A.N. On some properties of paramonotone operators. *Journal of Convex Analysis* **5** (1998) 269–278.

[18] Iusem, A.N. Augmented Lagrangian methods and proximal point methods for convex optimization. *Investigación Operativa* **8** (1999) 12–49.

[19] Iusem, A.N., Svaiter, B.F., Teboulle, M. Entropy-like proximal methods in convex programming. *Mathematics of Operations Research* **19** (1994) 790–814.

[20] Iusem, A.N., Teboulle, M. On the convergence rate of entropic proximal optimization algorithms. *Computational and Applied Mathematics* **12** (1993) 153–168.

[21] Iusem, A.N., Teboulle, M. Convergence rate analysis of nonquadratic proximal and augmented Lagrangian methods for convex and linear programming. *Mathematics of Operations Research* **20** (1995) 657–677.

[22] Kaplan, A., Tichatschke, R. Interior proximal method for variational inequalities: case of nonparamonotone operators. *Set-Valued Analysis* **12** (2004) 357–382.

[23] Kiwiel, K.C. Proximal minimization methods with generalized Bregman functions. *SIAM Journal on Control and Optimization* **35** (1997) 1142–1168.

[24] Langenberg, N. On the cutting plane property and the Bregman proximal point algorithm. *Journal of Convex Analysis* **18** (2011) 601–619.

[25] Martinet, B. Régularisation d'inéquations variationnelles par approximations successives. *Revue Française de Informatique et Recherche Opérationnelle* **4** (1970) 154–158.

[26] Minty, G. A theorem on monotone sets in Hilbert spaces. *Journal of Mathematical Analysis and Applications* **11** (1967) 434-439.

[27] Rockafellar, R.T. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization* **14** (1976) 877-898.

[28] Silva, P.J.S., Eckstein, J., Humes, C.Jr. Rescaling and stepsize selection in proximal methods using separable generalized distances. *SIAM Journal on Optimization* **12** (2001) 238–261.