AN INTERIOR POINT METHOD FOR CONSTRAINED SADDLE POINT PROBLEMS

Alfredo N. Iusem*

Instituto de Matemática Pura e Aplicada
Estrada Dona Castorina, 110
22460, Rio de Janeiro, RJ, Brazil

Markku Kallio

Helsinki School of Economics
Runeberginkatu 14–16
00100 Helsinki, Finland

Revised, November 1998

Abstract

We present an algorithm for the constrained saddle point problem with a convex-concave function L and convex sets with nonempty interior. The methods consists of moving away from the current iterate by choosing certain perturbed vectors. The values of gradients of L at these vectors provide an appropriate direction. Bregman functions allow us to define a curve which starts at the current iterate with this direction, and is fully contained in the interior of the feasible set. The next iterate is obtained by moving along such a curve with a certain step size. We establish convergence to a solution with minimal conditions upon the function L, weaker than Lipschitz continuity of the gradient of L, for instance, and including cases where the solution needs not be unique. We also consider the case in which the perturbed vectors are on certain specific curves starting at the current iterate, in which case another convergence proof is provided. In the case of linear programming, we obtain a family of interior point methods where all the iterates and perturbed vectors are computed with very simple formulae, without factorization of matrices or solution of linear systems, which makes the method attractive for very large and sparse matrices. The method may be of interest for massively parallel computing. Numerical examples for the linear programming case are given.

Keywords: Saddle point problems, variational inequalities, linear programming, interior point methods, Bregman distances.
1991 AMS Classification numbers: 90C25, 90C30.

---

*Research of this author was partially supported by CNPq grant N$^{\underline{o}}$ 301280/86.

1. Introduction

In this paper, we discuss methods for solving constrained saddle point problems. Given closed convex sets X, Y contained in $\mathbf{R}^n$, $\mathbf{R}^m$ respectively, and a function $L : X \times Y \to \mathbf{R}$, convex in the first variable and concave in the second one, the saddle point problem SPP(L,X,Y) consists of finding $(x^*, y^*) \in X \times Y$ such that

$$L(x^*, y) \le L(x^*, y^*) \le L(x, y^*) \tag{1}$$

for all $(x, y) \in X \times Y$. SSP(L,X,Y) is a particular case of the variational inequality problem, which we describe next. Given a closed convex set $C \subset \mathbf{R}^p$ and a maximal monotone operator $T : \mathbf{R}^p \to \mathcal{P}(\mathbf{R}^p)$, VIP($T$,$C$) consists of finding $z \in C$ such that there exists $u \in T(z)$ satisfying

$$\langle u, z' - z\rangle \ge 0 \tag{2}$$

for all $z' \in C$. If $\partial_x L$, $\partial_y L$ denote the subdifferential and superdifferential of L in each variable (i.e. the sets of subgradients and supergradients) respectively, and we define $T : \mathbf{R}^n \times \mathbf{R}^m \to \mathcal{P}(\mathbf{R}^n \times \mathbf{R}^m)$ as $T = (\partial_x L, -\partial_y L)$ then VIP(T,C) coincides with SSP(X,Y) if we take $p = m + n$ and $C = X \times Y$. It is easy to check that this T is maximal monotone.

It is well known that when $T = \partial f$ for a convex function $f(x)$ then VIP(T,C) reduces to min $f(x)$ s.t. $x \in C$. This suggests the extension of optimization methods to variational inequalities. In the unconstrained case $C = \mathbf{R}^n$, for which VIP(T,C) consists just of finding a zero of T, i.e. a z^* such that $0 \in T(z^*)$, one could attempt the natural extension of the steepest descent method, i.e.

$$z^{k+1} = z^k - \alpha u^k$$

with $u^k \in T(z^k)$ and $\alpha > 0$. This simple approach works only under very restrictive assumptions on T (strong monotonicity and Lipschitz continuity, see [**1**]). An alternative one, which works under weaker assumptions, is the following: move in a direction contained in $-T(z^k)$ finding an auxiliary (or perturbed) point w^k and then move from z^k in a direction contained in $-T(w^k)$, i.e.

$$w^k = z^k - \alpha_k u^k \tag{3}$$

$$z^{k+1} = z^k - \bar{\alpha}_k d^k \tag{4}$$

with $u^k \in T(z^k)$, $d^k \in T(w^k)$, and $\alpha_k,\bar{\alpha}_k > 0$. In the constrained case, one must project onto C in order to preserve feasibility, i.e., one takes

$$w^k = P_C(z^k - \alpha_k u^k) \tag{5}$$

$$z^{k+1} = P_C(z^k - \bar{\alpha}_k d^k) \tag{6}$$

where P_C is the orthogonal projection onto C. If T is Lipschitz continuous with constant π, then convergence of $\{z^k\}$ to a solution of VIP(T,C) is guaranteed by taking $\alpha_k = \bar{\alpha}_k = \alpha \in (0, 1/\pi)$. This result can be found in [**16**], where the method was first introduced. In the absence of Lipschitz continuity (or when the Lipschitz constant exists but is not known beforehand) it has been proven in [**11**] that the algorithm still converges to a solution of VIP(T,C) if α_k, $\bar{\alpha}_k$ are determined through a certain finite bracketing procedure. A somewhat similar method has been studied in [**15**].

When C has nonempty interior, it is interesting to consider interior point methods, i.e. such that $\{w^k\}$ and $\{z^k\}$ are contained in the interior C^o of C, thus doing away with the projection P_C. This can be achieved if a Bregman function g with zone C^o is available. Loosely speaking, such a g is a strictly convex function which is continuous in C, differentiable in C^o and such that its gradient diverges at the boundary of C. With g we construct a distance D_g on $C \times C^o$ as

$$D_g(w,z) = g(w) - g(z) - \langle \nabla g(z), w - z \rangle \tag{7}$$

and instead of the half line $\{z - tu : t \in \mathbf{R}_+\}$ we consider the curve $\{s_g(z,u,t) : t \in \mathbf{R}\}$ where $s_g(z,u,t)$ is the solution of $\min_{w \in C}\{u^t w + (1/t)D_g(w,z)\}$, or equivalently $s_g(z,u,t)$ is the solution w of

$$\nabla g(w) = \nabla g(z) - tu. \tag{8}$$

Under suitable assumptions on g (see Section 2), $s_g(z,u,0) = z$, the curve is uniquely defined for $t \in [0,\infty)$ and it is fully contained in C^o. The idea is to replace (5) and (6) by

$$w^k = s_g(z^k, u^k, \alpha_k) \tag{9}$$

$$z^{k+1} = s_g(z^k, d^k, \bar{\alpha}_k) \tag{10}$$

with u^k, d^k, α_k and $\bar{\alpha}_k$ as in (3) and (4). This algorithm was proposed in [**3**], where it is proved that convergence of $\{z^k\}$ to a solution of VIP(T,C) is guaranteed if α_k is chosen through a given finite bracketing procedure and $\bar{\alpha}_k$ solves a certain nonlinear equation in one variable. A slight improvement is presented in [**12**], where it is shown that convergence is preserved when $\bar{\alpha}_k$ is also found through another finite bracketing search. The main advantage of these interior point methods with Bregman functions as compared e.g. to Korpelevich's algorithm lies in the fact that no orthogonal projections onto the feasible set are needed, since all iterates belong automatically to the interior of this set. A detailed discussion of the computational effects of this advantage in the case in which C is an arbitrary polyhedron with nonempty interior can be found in Section 6 of [**3**]. For SPP(L,X,Y), when either X or Y is an arbitrary polyhedron with nonempty interior, the method introduced in this paper presents the same advantage over the algorithm in [**13**], discussed in the next paragraph, which requires orthogonal projections onto X and Y.

Going back to SPP(L,X,Y), another related method is presented in [**13**]. Given a point (x^k, y^k), a rather general perturbed vector $w^k = (\xi^k, \eta^k)$ is computed, and then

$$x^{k+1} = P_X(x^k - \bar{\alpha}_k d_x^k) \tag{11}$$

$$y^{k+1} = P_Y(y^k + \bar{\alpha}_k d_y^k) \tag{12}$$

with $d_x^k \in \partial_x L(x^k, \eta^k)$ and $d_y^k \in \partial_y L(\xi^k, y^k)$. In this case no search is required for the step size $\bar{\alpha}_k$, which is given by a simple formula in terms of the gap $L(x^k,\eta^k) - L(\xi^k, y^k)$. The difficulty is to some extent transferred to the selection of (ξ^k, η^k), which must satisfy certain conditions in order to ensure convergence to a solution of SPP(L,X,Y). One possibility is to adopt (5) so that $\xi^k = P_X(x^k - \alpha_k u_x^k)$, $\eta^k = P_Y(y^k + \alpha_k u_y^k)$, with $u_x^k \in \partial_x L(x^k, y^k)$, $u_y^k \in \partial_y L(x^k, y^k)$. In this case the method is somewhat similar to Korpelevich's (see [**16**]), but not identical: in linear programming, for instance, a closed expression for α_k is easily available for the method of [**13**] even when a Lipschitz constant for ∇L is not known beforehand. Other differences between the method in [**13**] with this choice of (ξ^k, η^k) and Korpelevich's are commented upon right after Example 6 in Section 3.

Besides (5), the method in [**13**] includes other options for the selection of (ξ^k, η^k). Taking advantage of this fact, we will combine the approaches in [**3**] and [**13**], producing an interior point method for the case of nonempty $X^o \times Y^o$.

We will consider two independent modifications upon the method in [**13**]. The first one to be discussed in Section 3 is the following: assuming that Bregman functions g_X, g_Y with zones X^o, Y^o respectively are available, we will use (10) instead of (11) and (12), setting

$$x^{k+1} = s_{g_X}(x^k, d_x^k, \tau_k) \tag{13}$$

$$y^{k+1} = s_{g_Y}(y^k, -d_y^k, \tau_k). \tag{14}$$

We will show that if the perturbation pair (ξ^k, η^k) satisfies conditions required in [**13**] and the stepsizes τ_k are chosen in a suitable way then convergence to a solution follows.

It is clear that the algorithm becomes practical only when s_{g_X}, s_{g_Y} have explicit formulae, i.e., in view of (8), when ∇g_X, ∇g_Y are easily invertible. Bregman functions with this property are available for the cases in which C is an orthant, a box, or certain polyhedra (in the latter case, inversion of the gradient requires solution of two linear systems) and are presented in [**3**]. Examples of realizations of this method are given are Section 4.

The second modification presented in Section 5 consists of an interior point procedure for the determination of (ξ^k, η^k), namely

$$\xi^k = \operatorname*{argmin}_{u \in X}\{L(u, y^k) + (1/\lambda) D_{g_X}(u, x^k)\} \tag{15}$$

$$\eta^k = \operatorname*{argmax}_{v \in Y}\{L(x^k, v) - (1/\mu) D_{g_Y}(v, y^k)\} \tag{16}$$

where λ, μ are positive constants. Definitions (15) and (16) imply

$$\xi^k = s_{g_X}(x^k, \tilde{d}_x^k, \lambda) \tag{17}$$

$$\eta^k = s_{g_Y}(y^k, -\tilde{d}_y^k, \mu) \tag{18}$$

with $\tilde{d}_x^k \in \partial_x L(\xi^k, y^k)$, $\tilde{d}_y^k \in \partial_y L(x^k, \eta^k)$. Equivalently with (17) and (18) we have

$$(1/\lambda)[\nabla g_X(x^k) - \nabla g_X(\xi^k)] \in \partial_x L(\xi^k, y^k) \tag{19}$$

$$(1/\mu)[\nabla g_Y(\eta^k) - \nabla g_Y(y^k)] \in \partial_y L(x^k, \eta^k). \tag{20}$$

Differently from (13) and (14), equations (17) and (18) are implicit in ξ^k, η^k, which appear in both sides of (19) and (20). These equations cannot be in general explicitly solved, even when ∇g_X, ∇g_Y are easily invertible. On the other hand, no search is needed for perturbation stepsizes λ and μ even when ∇L is not Lipschitz continuous. This choice of (ξ^k, η^k) does not satisfy the conditions in [**13**], so that the general convergence result for the sequence defined by (13) and (14) cannot be fully used. A special convergence proof will be given assuming that the Bregman functions are separable and the constraint sets X and Y are the nonnegative orthants.

It is immediate that (19) and (20) become explicit equations (up to the inversion of ∇g_X and ∇g_Y) when $\partial_x L(\cdot, y)$, $\partial_y L(x, \cdot)$ are constant. This happens in the case of linear programming. For this case, combining both modifications, we obtain an interior point method with explicit updating formulae both for the perturbed points and the primal and dual iterates. The curvilinear stepsize τ_k requires a simple search. We analyze the linear programming case in detail in Section 6, where we present also some numerical illustration of the method.

2. Bregman functions and distances

Let $C \subset \mathbf{R}^p$ be a closed and convex set and let C^o be the interior of C. Consider $g: C \to \mathbf{R}$ differentiable in C^o. For $w \in C$ and $z \in C^o$, define $D_g: C \times C^o \to \mathbf{R}$ by (7). D_g is said to be the Bregman distance associated to g.

We say that g is a *Bregman function with zone* C if the following conditions hold:

B1. g is continuous and strictly convex on C.

B2. g is twice differentiable on C^o, and in any bounded subset of C^o the eigenvalues of $\nabla^2 g(z)$ are bounded below by a strictly positive number.

B3. The level sets of $D_g(w, \cdot)$ and $D_g(\cdot, z)$ are bounded, for all $w \in C$, $z \in C^o$, respectively.

B4. If $\{z^k\} \subset C^o$ converges to z^*, then $\lim_{k\to\infty} D_g(z^*, z^k) = 0$.

B5. For all $v \in \mathbf{R}^n$ there exists $z \in C^o$ such that $\nabla g(z) = v$.

The definition of Bregman functions and distances originates in [**2**]. They have been widely used in convex optimization, e.g. [**4**], [**5**], [**6**], [**7**], [**8**]. Condition B2 in these references is weaker than here; only differentiability of g in C^o is required. On the other hand, such references included an additional condition which is now implied by our stronger condition B2, namely the result of Proposition 1 below. Condition B5 is called *zone coerciveness* in [**3**]. ¿From (7) and B1 it is immediate that for all $w \in C$, $z \in C^o$, $D_g(w, z) \geq 0$ and $D_g(w, z) = 0$ if and only if $w = z$.

We present next some examples of Bregman functions.

Example 1. $C = \mathbf{R}^p$, $g(x) = \|x\|^2$. Then $D_g(w, z) = \|w - z\|^2$.

Example 2. $C = \mathbf{R}^p_+$, $g(z) = \sum_{j=1}^p z_j \log z_j$, continuously extended to the boundary of C with the convention that $0 \log 0 = 0$. Then

$$D_g(w, z) = \sum_{j=1}^p [w_j \log(w_j/z_j) + z_j - w_j]. \tag{21}$$

D_g is the Kullback–Leibler information divergence, widely used in statistics.

Example 3. $C = \mathbf{R}^p_+$, $g(z) = \sum_{j=1}^p (z_j^\alpha - z_j^\beta)$, with $\alpha > 1$, $\beta \in (0,1)$.

Examples of Bregman functions satisfying these conditions for the cases of C being a box or a polyhedron can be found in [**3**]. We now proceed to a result which will be employed in the convergence analysis.

Proposition 1. *If g is a Bregman function with domain C^o, and $\{z^k\}$ and $\{w^k\}$ are sequences in C^o such that $\{z^k\}$ is bounded and $\lim_{k\to\infty} D_g(w^k, z^k) = 0$, then*

$$\lim_{k\to\infty} \parallel z^k - w^k \parallel = 0.$$

Proof: Let $B \subset R^n$ be a closed ball containing $\{z^k\}$ such that, for any z on the boundary of B, $\left\|z^k - z\right\| > 1$, for all k. For $\nu \in [0,1]$ define $z(\nu) = \nu w^k + (1-\nu)z^k$. Let ν_k be the largest $\nu \in [0,1]$ such that $z(\nu) \in B$. For each k, define $u^k = z(\nu_k)$. Since C^o is convex and $\nu \in [0,1]$, we get that $z(\nu_k) \in C^o$. By convexity of $D_g(\cdot, z^k)$,

$$0 \le D_g(u^k, z^k) \le D_g(w^k, z^k).$$

Because $D_g(w^k, z^k)$ converges to zero, so does $D_g(u^k, w^k)$. We shall show that $\parallel z^k - u^k \parallel$ converges to zero. Then, from the definition of B it follows that there exists k' such that, for all $k > k'$, u^k is in the interior of B, such that $\nu_k = 1$ and $u^k = w^k$; hence the assertion follows.

Define $u(\omega) = \omega u^k + (1-\omega)z^k$, for $\omega \in [0,1]$. By Taylor's Theorem,

$$g(u^k) = g(z^k) + \nabla g(z^k)(u^k - z^k) + \frac{1}{2}(u^k - z^k)^t[\nabla^2 g(u(\omega))](u^k - z^k) \tag{22}$$

for some $\omega \in [0,1]$. The quadratic term in (22) is equal to $D_g(u^k, z^k)$, which converges to zero. Because of boundedness of $\{z^k\}$ and $\{u^k\}$, B2 implies that $D_g(u^k, z^k)$ is bounded below by $\theta \left\|u^k - z^k\right\|^2$ for some $\theta > 0$ and independent of k. Therefore $\left\|u^k - z^k\right\|$ converges to zero. $\square$

3. An interior point method for saddle point computation

In this section we assume that X and Y have nonempty interior and that g_X and g_Y are Bregman functions with zones X and Y, respectively. $L : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is continuous on $X \times Y$, convex in the first variable and concave in the second one.

Following [**13**] we introduce perturbation sets $\Delta : X \times Y \to \mathcal{P}(X)$, $\Gamma : X \times Y \to \mathcal{P}(Y)$. We consider the following properties of Δ, Γ.

A1. If $(x,y) \in X^o \times Y^o$ and $(\xi, \eta) \in \Delta(x,y) \times \Gamma(x,y)$ then $L(x,\eta) - L(\xi, y) \ge 0$, with strict inequality if (x,y) is not a saddle point.

A2. If $E \subset X^o \times Y^o$ is bounded then $\cup_{(x,y)\in E}\Delta(x,y)$ and $\cup_{(x,y)\in E}\Gamma(x,y)$ are bounded.

A3. If $\{(u^\ell, v^\ell)\} \subset X \times Y$ converges to (u,v) as ℓ goes to ∞, $\xi^\ell \in \Delta(u^\ell, v^\ell)$, $\eta^\ell \in \Gamma(u^\ell, v^\ell)$ and $\lim_{\ell\to\infty}[L(u^\ell, \eta^\ell) - L(\xi^\ell, v^\ell)] = 0$ then (u,v) solves SPP(L,X,Y).

We give next a few examples of sets Δ, Γ, taken from [**13**].

Example 4. Assume that X and Y are compact and define $\Delta(x,y) = \mathrm{argmin}_{\xi \in X} L(\xi, y)$ and $\Gamma(x,y) = \mathrm{argmax}_{\eta \in Y} L(x, \eta)$.

Example 5. Let $\lambda > 0$, $\mu > 0$ and define $\Delta(x,y) = \mathrm{argmin}_{\xi \in X}[L(\xi, y) + \lambda \|\xi - x\|^2]$ and $\Gamma(x,y) = \mathrm{argmax}_{\eta \in Y}[L(x,\eta) - \mu \|\eta - y\|^2]$, for $x \in X$, $y \in Y$. Due to strict convexity (concavity) of the minimand (maximand), $\Delta(x,y)$ $(\Gamma(x,y))$ is uniquely determined.

Example 6. Assume that $\nabla_x L$ and $\nabla_y L$ are Lipschitz continuous with constant π and let $0 < \alpha < 1/\pi$. Define $\Delta(x,y) = \{P_X(x - \alpha \nabla_x L(x,y))\}$ and $\Gamma(x,y) = \{P_Y(y + \alpha \nabla_y L(x,y))\}$. It corresponds to the perturbed vector w^k of Korpelevich's method ([**16**]) in (5). But we must point out that when we use the method in [**13**] (i.e. (11)–(12)) with this perturbation we obtain something quite different from Korpelevich's method, because Korpelevich's method, with $T = (\nabla_x L, -\nabla_y L)$, would then choose $d_x^k = \nabla_x L(\xi^k, \eta^k)$ and $d_y^k = \nabla_y L(\xi^k, \eta^k)$, while the method in [**13**] takes $d_x^k = \nabla_x L(x^k, \eta^k)$ and $d_y^k = \nabla_y L(\xi^k, y^k)$. This choice of d_x^k, d_y^k is possible only in the case of a saddle point problem, where we have two variables x and y, while Korpelevich's method is devised for a general variational inequality problem, where we only have the vector w^k, instead of the pair (ξ^k, η^k).

It is not difficult to prove that the choices of Δ and Γ made in Examples 4–6 satisfy conditions A1–A3.

Next we present the algorithm generating the sequence $\{(x^k, y^k)\}$. We denote $z^k = (x^k, y^k)$, $C = X \times Y$ and $C^o = X^o \times Y^o$, and define $D_g : C \times C^o \to \mathbf{R}$ employing

$$g(z) = g_X(x) + g_Y(y). \tag{23}$$

It is a matter of routine to check that g is a Bregman function with zone C. The algorithm of this section is defined as follows:

Step 0: Initialization. Choose parameters β and γ, $1 > \beta \geq \gamma > 0$, and take

$$z^0 \in C^o. \tag{24}$$

Step 1: Perturbation. Begin iteration with $z^k = (x^k, y^k)$. Choose perturbation vectors $(\xi^k, \eta^k) \in \Delta(x^k, y^k) \times \Gamma(x^k, y^k)$ and define

$$\sigma_k = L(x^k, \eta^k) - L(\xi^k, y^k). \tag{25}$$

Step 2: Stopping test. If $\sigma_k = 0$ then stop.

Step 3: Direction. Choose $d_x^k \in \partial_x L(x^k, \eta^k)$, and $d_y^k \in \partial_y L(\xi^k, y^k)$, and denote $d^k = (d_x^k, -d_y^k)$.

Step 4: Stepsize. For $t \geq 0$ define functions

$$z(t) = s_g(z^k, d^k, t) \tag{26}$$

and

$$\phi_k(t) = t\sigma_k - D_g(z^k, z(t)). \tag{27}$$

Search for a step size τ_k satisfying

$$\gamma \sigma_k \tau_k \leq \phi_k(\tau_k) \leq \beta \sigma_k \tau_k. \tag{28}$$

Step 5: Update. Set

$$z^{k+1} = z(\tau_k), \tag{29}$$

increment k by one and return to Step 1.

Theorem 1 below shows that $\phi_k(t)$ is a concave function with $\phi_k(0) = 0$ and $\phi'_k(0) = \sigma_k > 0$. For iteration k and a potential stepsize t, this function provides a lower bound for the decrease in the Bregman distance to a solution. Thus the left inequality in (28) aims to guarantee that such a distance is suitably decreased while the right inequality aims to bound the stepsize from below. To shorten the search for an acceptable stepsize, one may begin with the stepsize employed in the preceding iteration. Also, we will show that the stopping criterion of Step 2 is appropriate, in the sense that when termination occurs z^k is a solution of the problem. The convergence properties of the algorithm are formalized as follows.

Theorem 1. *Assume that SPP(L,X,Y) has solutions. Let g be a Bregman function with zone C^o and let $\{z^k\}$ be a sequence generated by the algorithm of this section, where $\xi^k \in \Delta(z^k)$ and $\eta^k \in \Gamma(z^k)$, with Δ and Γ satisfying A1 and A2. If the algorithm stops at iteration k, then z^k solves SPP(L,X,Y). If the sequence $\{z^k\}$ is infinite, then*

i) $z^k \in C^o$, *for all k,*

ii) *for any solution z^* of SPP(L,X,Y), the Bregman distance $D_g(z^*, z^k)$ is monotonically decreasing,*

iii) $\{z^k\}$ *is bounded,*

iv) $\lim_{k\to\infty} \sigma_k = 0$;

v) $\lim_{k\to\infty} \left\| z^{k+1} - z^k \right\| = 0$;

vi) *if additionally A3 is satisfied, then the sequence $\{z^k\}$ converges to a solution of SPP(L,X,Y).*

Proof: We prove first (i) by induction. z^0 belongs to C^o by (24). Assuming that z^k belongs to C^o, note that (26) is equivalent to

$$\nabla g(z(t)) = \nabla g(z^k) - t d^k, \tag{30}$$

so that by B1 and B5, $z(t)$ is uniquely defined by (30) and belongs to C^o for all $t \geq 0$. By (29), z^{k+1} belongs to C^o and (i) holds. Now we consider the case of finite termination, i.e. when, according to Step 2, $\sigma_k = 0$. In such a case, since z^k belongs to C^o by (i), it follows from A1 that z^k is a saddle point. From now on we assume that $\{z^k\}$ is infinite. We proceed to prove (ii). Since $z(t) \in C^o$ for all $t \geq 0$, using (7) and (30) we get

$$D_g(z^*, z^k) - D_g(z^*, z(t)) = -t\langle d^k, z^* - z^k\rangle - D_g(z^k, z(t)). \tag{31}$$

By convexity/concavity of L, since (x^*, y^*) is a saddle point,

$$-\langle d^k, z^* - z^k\rangle \geq -L(x^*, \eta^k) + L(x^k, \eta^k) + L(\xi^k, y^*) - L(\xi^k, y^k)$$

$$= \sigma_k - L(x^*, \eta^k) + L(\xi^k, y^*) \geq \sigma_k \tag{32}$$

Combining (27), (31) and (32) yields

$$D_g(z^*, z^k) - D_g(z^*, z(t)) \geq \phi_k(t). \tag{33}$$

Differentiating (30) yields

$$\nabla^2 g(z(t)) \frac{d}{dt} z(t) = -d^k$$

so that the first two derivatives of $\phi_k(t)$ in (27) are

$$\phi_k'(t) = \sigma_k + \langle d^k, z(t) - z^k \rangle \tag{34}$$

and

$$\phi_k''(t) = -\varphi_k(t), \tag{35}$$

with φ_k defined as

$$\varphi_k(t) = (d^k)^t [\nabla^2 g(z(t))]^{-1} d^k. \tag{36}$$

Since $z(0) = z^k$, $\phi_k(0) = 0$. By A1-A2 and (34), $\phi_k'(0) = \sigma_k > 0$, in view of the stopping criterion, and so (32) implies $d^k \neq 0$. Consequently by B2, (35) and (36), $\phi_k''(t) < 0$. Therefore $\phi_k(t)$ is a concave function of $t \in [0, \infty)$ with a positive slope at $t = 0$. Inequality (33) implies that $\phi_k(t)$ is bounded above. Hence there exists a stepsize τ_k satisfying (28) and therefore item (i) follows from (30). Furthermore, for some $\gamma_k \in [\gamma, \beta]$,

$$\phi_k(\tau_k) = \gamma_k \tau_k \sigma_k \tag{37}$$

which together with (33) implies item (ii). Denote by B^o the set $\{z \in C^o : D_g(z^*, z) \leq D_g(z^*, z^0)\}$. By B3, B^o is bounded. Item (ii) implies that $z^{k+1} \in B^o$ if $z^k \in B^o$. Therefore, item (iii) follows then by induction from (24).

Since $\{z^k\}$ is bounded, we get from A2 that $\{(\xi^k, \eta^k)\}$ is bounded. The operator $(\partial_x L, -\partial_y L)$ is maximal monotone, therefore bounded over bounded sets contained in the interior of its domain, which in this case is $\mathbf{R}^n \times \mathbf{R}^m$ (e.g. [**18**]). It follows that $\{d^k\}$ is bounded.

By Taylor's theorem and (28), for some $\hat{\tau} \in [0, \tau_k]$,

$$\phi_k(\tau_k) = \tau_k \sigma_k - \frac{\tau_k^2}{2} \varphi_k(\hat{\tau}) \leq \beta \tau_k \sigma_k.$$

Consequently, and because B2, A2 and item (iii) imply $\varphi_k(\hat{\tau}) \leq \theta$, for some $\theta > 0$, we have

$$\tau_k \geq \frac{2(1-\beta)\sigma_k}{\varphi_k(\hat{\tau})} \geq \frac{2(1-\beta)\sigma_k}{\theta}.$$

By (28) and (33),

$$\sum_k \tau_k \sigma_k < \infty. \tag{38}$$

By (38), $\lim_{k\to\infty} \tau_k \sigma_k = 0$, and then, multiplying (37) by σ_k, we conclude that $\lim_{k\to\infty} \sigma_k = 0$, i.e. (iv) holds. ¿From (27) and (28) we have $0 \le D_g(z^k, z^{k+1}) \le \tau_k \sigma_k$, where the right side converges to zero. Therefore, item (v) follows from item (iii) and Proposition 1. By item (iii) the sequence $\{z^k\}$ has an accumulation point $\bar{z}$. By item (iv) and A3, $\bar{z}$ solves SPP(L,X,Y). By item (ii) the Bregman distance $D_g(\bar{z}, z^k)$ is non-increasing. If $\{z^{j_k}\}$ is a subsequence of $\{z^k\}$ which converges to $\bar{z}$, by B4 $\lim_{k\to\infty} D_g(\bar{z}, z^{j_k}) = 0$. Since $\{D_g(\bar{z}, z^k)\}$ is a nondecreasing and nonnegative sequence with a subsequence which converges to 0, we conclude that the whole sequence converges to 0. Now we apply Proposition 1 with $w^k = \bar{z}$, and conclude that $\lim_{k\to\infty} z^k = \bar{z}$. □

4. Particular realizations of the algorithm

In the unrestricted case ($X = \mathbf{R}^n$, $Y = \mathbf{R}^m$) we can take g as in Example 1 so that

$$z(t) = z^k - \frac{t}{2} d^k$$

and

$$\phi_k(t) = t\sigma_k - \frac{t^2}{4} \left\| d^k \right\|^2.$$

In this case, with $\beta = \gamma = 0.5$, we may choose τ_k to maximize $\phi_k(t)$ yielding

$$\tau_k = \frac{2\sigma_k}{\|d^k\|^2}.$$

and

$$z^{k+1} = z^k - \frac{\sigma_k}{\|d^k\|^2} d^k,$$

which is a special case of the method considered in [**13**].

Next, consider the case in which X, Y are the nonnegative orthants. For g as in Example 2 we have

$$\nabla g(z)_j = 1 + \log z_j \tag{39}$$

$$\nabla^2 g(z)^{-1} = \operatorname{diag}(z_1, \ldots, z_n) \tag{40}$$

$$z(t)_j = z_j^k \exp(-t d_j^k) \tag{41}$$

$$\phi_k(t) = t\sigma_k - \sum_j z_j^k [t d_j^k - 1 + \exp(-t d_j^k)] \tag{42}$$

In this case, a search is needed for determining the stepsize τ_k in Step 3 of the algorithm. For further illustration, see the case of linear programming in Section 6.

5. An interior point procedure for perturbation

In this section we analyze a selection rule for specifying ξ^k and η^k which extends the proposal of Example 6 in Section 3. Let g_X and g_Y be Bregman functions with zones X^o and Y^o, respectively. Take $\lambda > 0$ and $\mu > 0$, and for $(x, y) \in X^o \times Y^o$ define

$$\xi = \xi(x, y) = \underset{u \in X}{\operatorname{argmin}} \{L(u, y) + \frac{1}{\lambda} D_{g_X}(u, x)\} \tag{43}$$

$$\eta = \eta(x, y) = \underset{v \in Y}{\operatorname{argmax}} \{L(x, v) - \frac{1}{\mu} D_{g_Y}(v, y)\}. \tag{44}$$

We shall restrict the discussion to the following special case of SPP(L,X,Y): $X = \mathbf{R}^n_+$, $Y = \mathbf{R}^m_+$, L twice continuously differentiable on $\mathbf{R}^n \times \mathbf{R}^m$ and g separable, i.e. $g_X(x) = \sum_{j=1}^n \tilde{g}_j(x_j)$, $g_Y(y) = \sum_{i=1}^m \hat{g}_i(y_i)$. We will show that the perturbation sets defined by (43)-(44) satisfy A1 and A2. A3 does not hold in general for this perturbation, but we will establish convergence of the method under a nondegeneracy assumption on the problem, and, in the case of linear programming, without such assumption. Our proof can be extended to nondifferentiable L and to box, rather than positivity, constraints, at the cost of some minor technical complications.

Proposition 2. *Assume that L is continuously differentiable, $X \times Y = \mathbf{R}^n_+ \times \mathbf{R}^n_+$, and that g is a separable Bregman functions with zone $X^o \times Y^o$. Consider $(x, y) \in X^o \times Y^o$ and (ξ, η) as in (43)–(44). Then*

i) (ξ, η) is uniquely determined and $(\xi, \eta) \in X^o \times Y^o$;
ii) if (x, y) is not a saddle point, then $L(x, \eta) - L(\xi, y) > 0$;
iii) if $B \subset X^o \times Y^o$ is nonempty and bounded, then the set $\{(\xi(x, y), \eta(x, y)) : (x, y) \in B\}$ is bounded.

Proof: Let $f(u) = L(u, y) + 1/\lambda D_{g_X}(u, x)$. Take $w = \nabla_x L(x, y)$ and define $\bar{f}(u) = L(x, y) + \langle w, u - x\rangle + 1/\lambda D_{g_X}(u, x)$. By convexity of $L(\cdot, y)$ we have

$$\bar{f}(u) \leq f(u) \tag{45}$$

for all $u \in X$. Consider the unrestricted minimization problem $\min \bar{f}(u)$, whose optimality conditions are $\nabla g_X(u) = \nabla g(x) - \lambda w$. By B5, this equation in u has a unique solution $\bar{u}$ which belongs to X^o. It follows easily from (7) that $D_{g_X}(\cdot, x)$ is strictly convex, and the same holds for $\bar{f}$, so that $\bar{u}$ minimizes $\bar{f}$ in $\mathbf{R}^n$. A strictly convex function which attains its minimum has bounded level sets, and it follows from (45) that f also has bounded level sets (see [**19**, Corollary 8.7.1]), and therefore it attains its minimum in X. Being the sum of a convex function and a strictly convex one, f is strictly convex and so the minimizer is unique, by convexity of X. The fact that this minimizer belongs to X^o follows from B5 and has been established in [**10**, Theorem 4.1]. By definition of f, this minimizer is $\xi(x, y, \lambda)$. A similar proof holds for $\eta(x, y, \mu)$ so that the proof of item (i) is complete.

For item (ii), assume that (x, y) is not a saddle point. Then $\nabla_x L(x, y) \neq 0$ or $\nabla_y L(x, y) \neq 0$. Assume that the former case holds; the latter one can be analyzed similarly. Observe that

$$\nabla f(u) = \nabla_x L(u, y) + \frac{1}{\lambda}[\nabla g_X(u) - \nabla g_X(x)] \tag{46}$$

so that $\nabla f(x) = \nabla_x L(x, y) \neq 0$. Hence x is not a solution of (43) and consequently

$$f(x) = L(x, y) > L(\xi, y) + \frac{1}{\lambda} D_{g_X}(\xi, x) > L(\xi, y).$$

This implies $L(x, y) - L(\xi, y) > 0$. Similarly, $L(x, \eta) - L(x, y) \geq 0$ and therefore $L(x, \eta) - L(\xi, y) > 0$. This completes the proof of item (ii).

To prove item (iii), we will prove boundedness of $\{\xi(x, y) : (x, y) \in B\}$. A similar argument holds for $\eta(x, y)$. The proof relies on successive relaxations for optimization problems, whose optimal objective function values are upper bounds for $\|\xi\|^2$. Since $\nabla f(\xi) = 0$ by (46), ξ is unique by item (i), and we have

$$\|\xi\|^2 = \max\{\|u\|^2 \mid u \geq 0, \nabla g_X(u) - \nabla g_X(x) = -\lambda \nabla_x L(u, y)\}.$$

Note that

$$\begin{aligned} &\{u \geq 0 |\ \nabla g_X(u) - \nabla g_X(x) = -\lambda \nabla_x L(u, y)\} \subset \\ &\big\{u \geq 0 |\ u^t \big(\nabla g_X(u) - \nabla g_X(x)\big) = -\lambda u^t \nabla_x L(u, y)\big\} \subset \\ &\big\{u \geq 0 |\ u^t \big(\nabla g_X(u) - \nabla g_X(x)\big) \leq -\lambda u^t \nabla_x L(u, y)\big\} \subset \\ &\big\{u \geq 0 |\ g_X(u) - g_X(0) \leq u^t \big(\nabla g_X(x) - \lambda \nabla_x L(0, y)\big)\big\}, \end{aligned} \tag{47}$$

using the facts that $u^t\big(\nabla_x L(u, y) - \nabla_x L(0, y)\big) \geq 0$ and $g_X(u) - g_X(0) \leq u^t \nabla g_X(u)$, which result from convexity of g_X and $L(\cdot, y)$, in the last inclusion of (47).

Let now $\rho = \max_j \{\sup_{(x,y)\in B}[\tilde{g}_j'(x_j) - \lambda \nabla_x L(0, y)_j]\}$. We claim that $\rho < \infty$. It suffices to show that $\sup_{(x,y)\in B}[\tilde{g}_j'(x_j) - \lambda \nabla_x L(0, y)_j] < \infty$ for all j, which follows from the facts that $\nabla_x L(0, y)$ is continuous as a function of y, $\tilde{g}_j'$ is continuous in $\mathbf{R}_{++}$, $\lim_{t\to 0} \tilde{g}_j'(t) = -\infty$, because of B5, and B is bounded. The claim is established.

It follows from (47) that

$$\{u \geq 0 |\ \nabla g_X(u) - \nabla g_X(x) = -\lambda \nabla_x L(u, y)\} \subset$$

$$\{u \geq 0 |\ g_X(u) - g_X(0) \leq \rho s^t u\} = \{u \geq 0 |\ g_X(u) - \rho s^t u \leq g_X(0)\}.$$

where $s^t = (1, 1, \ldots, 1)$. We claim that the set $\{u \geq 0 | g_X(u) - \rho s^t u \leq g_X(0)\}$ is bounded. Note that this set is a level set of the convex function $\bar{g}_X(u) = g_X(u) - \rho s^t u$ and that $\nabla \bar{g}_X(u) = \nabla g_X(u) - \rho s$. By B5, there exists $z > 0$ such that $\nabla g_X(z) = \rho s$, i.e. $\nabla \bar{g}_X(z) = 0$, so that z is an unrestricted minimizer of $\bar{g}_X$. Since $\nabla^2 \bar{g}_X = \nabla^2 g_X$, we get from B2 that $\nabla^2 \bar{g}_X(u)$ is positive definite for all $u > 0$, and hence $\bar{g}_X$ is strictly convex and z is its unique minimizer. Thus, the level set of $\bar{g}_X$ corresponding to the value $\bar{g}_X(z)$ is the singleton $\{z\}$. It is well known that if a convex function has a bounded level set, then all its nonempty level sets are bounded, which establishes the claim. It follows that

$$\begin{aligned} \|\xi\|^2 = \max\{\|u\|^2 \mid u \geq 0, \nabla g_X(u) - \nabla g_X(x) = -\lambda \nabla_x L(u, y)\} \leq \\ \max\{\|u\|^2 \mid u \geq 0, g_X(u) - \rho s^t u \leq g_X(0)\} < \infty, \end{aligned} \tag{48}$$

where the last inequality in (48) results from boundedness of the feasible set and continuity of the objective of the corresponding optimization problem. □

We use now (43) and (44) to define perturbation sets Δ and Γ consisting of single elements ξ and η, respectively. In view of Proposition 2, this rule of perturbation satisfies conditions A1 and A2. However, the following example demonstrates that A3 may not hold.

Take $m = n = 1$, $\lambda = \mu = 1$, g as in Example 2 and $L(x, y) = (x-1)^2 - (y-1)^2$, whose only saddle point is $(1, 1)$. In this case the optimality conditions for (43)–(44) become

$$2(\xi - 1) + \log \xi = \log x \tag{49}$$

$$2(\eta - 1) + \log \eta = \log y. \tag{50}$$

Take a sequence $\{(x^k, y^k)\}$ convergent to $(0, 0)$. Then the right hand sides of (49)–(50) diverge to $-\infty$, implying that $\log \xi^k$ and $\log \eta^k$ also diverge to $-\infty$; i.e. $\{(\xi^k, \eta^k)\}$ also converges to $(0, 0)$. It follows that $\lim_{k\to\infty}[L(x^k, \eta^k) - L(\xi^k, y^k)] = L(0, 0) - L(0, 0) = 0$, but $(0, 0)$ is not a saddle point of L.

This example shows that the convergence argument of Theorem 1 is no longer valid for our algorithm with perturbation sets chosen according to (43)–(44). Instead, we provide another convergence argument, which can be more easily formulated in terms of variational inequalities and nonlinear complementarity problems. We need two results on these problems. The first one is well known, while the second one is, to our knowledge, new, and of some interest on its own.

Proposition 3.

i) If $T : \mathbf{R}^p \to \mathbf{R}^p$ is monotone and continuous, and $C \subset \mathbf{R}^p$ is closed and convex, then the solution set of VIP(T,C), when nonempty, is closed and convex.

ii) If $C = \mathbf{R}^p_+$ then VIP(T,C) becomes the nonlinear complementarity problem NCP(T), consisting of finding $z \in \mathbf{R}^p$ such that

$$T(z)_j \geq 0 \qquad (1 \leq j \leq p) \tag{51}$$

$$z_j \geq 0 \qquad (1 \leq j \leq p) \tag{52}$$

$$T(z)_j z_j = 0 \qquad (1 \leq j \leq p) \tag{53}$$

Proof: Elementary (see, e.g. [**9**]). □

Corollary 1. *If $X \times Y = \mathbf{R}^n_+ \times \mathbf{R}^m_+$ then $(\bar{x}, \bar{y})$ is a solution of SPP(L,X,Y) if and only if*

$$\bar{x} \in X, \bar{y} \in Y \tag{54}$$

$$\nabla_x L(\bar{x}, \bar{y})_j = 0 \quad \text{for} \quad \bar{x}_j > 0 \tag{55}$$

$$\nabla_x L(\bar{x}, \bar{y})_j \geq 0 \quad \text{for} \quad \bar{x}_j = 0 \tag{56}$$

$$\nabla_y L(\bar{x}, \bar{y})_i = 0 \quad \text{for} \quad \bar{y}_i > 0 \tag{57}$$

$$\nabla_y L(\bar{x}, \bar{y})_i \leq 0 \quad \text{for} \quad \bar{y}_i = 0 \tag{58}$$

Proof: Follows from Proposition 3 with $T = (\nabla_x L, -\nabla_y L)$. □

Before presenting our new result, we need some notation.

Definition: We say that NCP(T) satisfies the *strict complementarity assumption* (SCA from now on) if for all solutions $\hat{z}$ of NCP(T) and for all j between 1 and p it holds that either $\hat{z}_j > 0$ or $T(\hat{z})_j > 0$.

For $z \in \mathbf{R}^p$, let $J(z) = \{j \in \{1, \ldots, p\} : z_j = 0\}$ and $I(z) = \{1, \ldots, p\} \setminus J(z)$.

Proposition 4. *Assume that T is monotone and continuous and that NCP(T) satisfies SCA. If $\bar{z} \in \mathbf{R}^p$ satisfies (52) and (53), and $J(\bar{z}) \subset J(z^*)$ for some solution z^* of NCP(T), then $\bar{z}$ solves NCP(T).*

Proof: Let $z(\alpha) = \bar{z} + \alpha(z^* - \bar{z})$. We will prove that $z(\alpha)$ satisfies (52)-(53) for all $\alpha \in [0, 1]$, and also (51) for α close to 1; then we will show that if (51) is violated for some $\alpha \in [0, 1]$ then SCA will be violated too. In order to prove that $z(\alpha)$ satisfies (53) for all $\alpha \in [0, 1]$ we must make a detour. Let $U = \{z \in \mathbf{R}^p_+ : z_j = 0 \text{ for all } j \in J(\bar{z})\}$. A vector $\hat{z}$ solves VIP(T,U) if and only if $\hat{z} \in U$ and

$$\sum_{j=1}^{p} T(\hat{z})_j (z_j - \hat{z}_j) \geq 0 \tag{59}$$

for all $z \in U$. Note that $\bar{z} \in U$. Since $J(\bar{z}) \subset J(z^*)$, z^* also belongs to U. We claim that both $\bar{z}$ and z^* satisfy (59). This is immediate for z^* which satisfies (59) for all $z \in \mathbf{R}^p_+$, since it solves NCP(T), equivalent, by Proposition 3(ii), to VIP(T,$\mathbf{R}^p_+$). Regarding $\bar{z}$, we have $\bar{z}_j = z_j = 0$ for all $z \in U$ and all $j \in J(\bar{z})$, so that terms with $j \in J(\bar{z})$ in the summation of (59) vanish. For $j \notin J(\bar{z})$, we have $T(\bar{z})_j = 0$ because $\bar{z}$ satisfies (53), so that terms with $j \notin J(\bar{z})$ in the summation of (59) also vanish. Therefore both $\bar{z}$ and z^* solve VIP(T,U), and so $z(\alpha)$ solves VIP(T,U) for all $\alpha \in [0, 1]$ by Proposition 3(i). Thus $z(\alpha)$ satisfies (52) for all $\alpha \in [0, 1]$. We look now at (53). Since $z(\alpha) \in U$, we have $z(\alpha)_j = 0$ for $j \in J(\bar{z})$, i.e. (53) holds for such j's. If $j \in I(\bar{z})$, we take z^1 and z^2 defined as $z^1_\ell = z^2_\ell = 0$ for $\ell \neq j$, $z^1_j = 2z(\alpha)_j$, $z^2_j = (1/2)z(\alpha)_j$, so that z^1 and z^2 belong to U. Looking at (59) with $\hat{z} = z(\alpha)$, $z = z^1$ or z^2 we get $T(z(\alpha))_j z(\alpha)_j \geq 0$, $-(1/2)T(z(\alpha))_j z(\alpha) j \geq 0$, so that $z(\alpha)$ satisfies (53) for all $\alpha \in [0, 1]$. Finally we look at (51). For $j \in I(\bar{z})$, we have $z(\alpha)_j > 0$ for $\alpha \in [0, 1)$, and therefore, since $z(\alpha)$ satisfies (53), $T(z(\alpha))_j = 0$, i.e. (51) holds for $j \in I(\bar{z})$, $\alpha \in [0, 1)$. If $j \in J(\bar{z})$ then $j \in J(z^*)$, i.e. $z^*_j = 0$. By SCA, $0 < T(z^*)_j = T(z(1))_j$. By continuity of T, $T(z(\alpha))_j > 0$ for $j \in J(\bar{z})$ and α close to 1, i.e., defining $A = \{\alpha \in [0, 1) : T(z(\alpha))_j > 0 \text{ for all } j \in J(\bar{z})\}$, we get that $A \neq \emptyset$. It follows that for $\alpha \in A$, $z(\alpha)$ satisfies (51)-(53), i.e. it solves NCP(T). Let $\bar{\alpha} = \inf A$. By continuity of T, $z(\bar{\alpha})$ also solves NCP(T). We claim now that $\bar{\alpha} = 0$. If $\bar{\alpha} \neq 0$, then, by the infimum property of $\bar{\alpha}$, there exists $j \in J(\bar{z})$ such that $T(z(\bar{\alpha}))_j = 0$, and also $z(\bar{\alpha})_j = 0$ because $z(\bar{\alpha}) \in U$, in which case SCA is violated at $z(\bar{\alpha})$. The claim is established and therefore $\bar{\alpha} = 0$ and $\bar{z} = z(0)$ solves NCP(T). □

We will also use the following elementary result for bilinear L.

Proposition 5. *For a bilinear function $L(x, y)$ and for any (x, y) and (x', y') in $\mathbf{R}^n \times \mathbf{R}^m$, it holds that*

$$(\nabla_x L(x', y') - \nabla_x L(x, y))(x' - x) - (\nabla_y L(x', y') - \nabla_y L(x, y))(y' - y) = 0.$$

Proof: Elementary. □

We will consider SCA for SPP(L,$\mathbf{R}^n_+$,$\mathbf{R}^m_+$), which is just NCP(T) with $T = (\nabla L_x. - \nabla L_y)$. It is clear that in this case SCA means that, for all solutions (x, y) of SPP(L,$\mathbf{R}^n_+$,$\mathbf{R}^m_+$), $x_j = 0$ implies $\partial L(x, y)/\partial x_j > 0$ and $y_i = 0$ implies $\partial L(x, y)/\partial y_i < 0$.

Theorem 2. *Assume that SPP(L,X,Y) has solutions, L is continuously differentiable, $X \times Y = \mathbf{R}_+^n \times \mathbf{R}_+^n$, and g is a separable Bregman function with zone $\mathbf{R}_+^n \times \mathbf{R}_+^m$, satisfying B1–B5. Consider the algorithm of Section 3 with the perturbation vectors ξ^k and η^k defined according to (43) and (44). If either SCA is satisfied or L is bilinear then the sequence $\{(x^k, y^k)\}$ generated by the algorithm converges to a solution $(\bar{x}, \bar{y})$ of SPP(L,X,Y).*

Proof: Proposition 2 implies that our choice of (ξ^k, η^k) satisfies A1 and A2, so that the results of Theorem 1(i)-(v) hold. Therefore $\{(x^k, y^k)\}$ has an accumulation point $\bar{z} = (\bar{x}, \bar{y}) \in X \times Y$, i.e. (54) holds for $(\bar{x}, \bar{y})$. We prove next that $(\bar{x}, \bar{y})$ satisfies also (55) and (57). Under SCA, we then show that $(\bar{x}, \bar{y})$ solves SPP(L,X,Y). For the bilinear case we show directly that $\{z^k\}$ converges to $(\bar{x}, \bar{y})$.

First we show that $\|\xi^k - x^k\|$ and $\|\eta^k - y^k\|$ converge to zero. Using (7), (19) and (20), we have

$$\frac{1}{\lambda}[D_{g_X}(x^k, \xi^k) + D_{g_X}(\xi^k, x^k)] = \frac{1}{\lambda}\langle \nabla g_X(x^k) - \nabla g_X(\xi^k), x^k - \xi^k \rangle$$

$$= \langle \nabla_x L(\xi^k, y^k), x^k - \xi^k \rangle \leq L(x^k, y^k) - L(\xi^k, y^k), \tag{60}$$

$$\frac{1}{\mu}[D_{g_Y}(y^k, \eta^k) + D_{g_Y}(\eta^k, y^k)] = \frac{1}{\mu}\langle \nabla g_Y(y^k) - \nabla g_Y(\eta^k), y^k - \eta^k \rangle$$

$$= \langle -\nabla_y L(x^k, \eta^k), y^k - \eta^k \rangle \leq L(x^k, \eta^k) - L(x^k, y^k). \tag{61}$$

Add (60)–(61) to get

$$0 \leq \frac{1}{\lambda}[D_{g_X}(x^k, \xi^k) + D_{g_X}(\xi^k, x^k)] + \frac{1}{\mu}[D_{g_Y}(y^k, \eta^k) + D_{g_Y}(\eta^k, y^k)] \leq L(x^k, \eta^k) - L(\xi^k, y^k) \tag{62}$$

By Theorem 1(iv), the rightmost expression in (62) tends to 0 as k goes to ∞ so that

$$\lim_{k\to\infty} D_{g_X}(\xi^k, x^k) = \lim_{k\to\infty} D_{g_Y}(\eta^k, y^k) = 0. \tag{63}$$

Since $\{(x^k, y^k)\}$ is bounded, by Proposition 1 and (63),

$$\lim_{k\to\infty} \|\xi^k - x^k\| = 0 \tag{64}$$

and

$$\lim_{k\to\infty} \|\eta^k - y^k\| = 0. \tag{65}$$

We check now that (55) holds at $(\bar{x}, \bar{y})$. In the case of separable g_X and differentiable L, (43) with $(x, y) = (x^k, y^k)$ and $\xi = \xi^k$ implies

$$\tilde{g}_j'(x_j^k) - \tilde{g}_j'(\xi_j^k) = \lambda \nabla_x L(\xi^k, y^k)_j. \tag{66}$$

Consider now j such that $\bar{x}_j > 0$. Let $\{x_k^{\ell_k}\}$, $\{\xi_j^{\ell_k}\}$ be subsequences of $\{x_j^k\}$, $\{\xi_j^k\}$ respectively such that $\lim_{k\to\infty} x_j^{\ell_k} = \lim_{k\to\infty} \xi_j^{\ell_k} = \bar{x}_j > 0$. Since g_j and L are continuously differentiable for $x_j > 0$, we can take limits in (66) along the subsequence and obtain

$$0 = \tilde{g}_j'(\bar{x}_j) - \tilde{g}_j'(\bar{x}_j) = \lambda \nabla_x L(\bar{x}, \bar{y})_j. \tag{67}$$

Hence (55) follows. A similar argument shows that (57) holds. In terms of NCP(T), we have proved that $\bar{z} = (\bar{x}, \bar{y})$ satisfies (52) and (53). It remains to be proved that $\bar{z}$ satisfies (51), i.e. that $(\bar{x}, \bar{y})$ satisfies (56) and (58).

¿From now on we will consider separately the case of general L under SCA, for which we will use Proposition 4, and the case of bilinear L, which will result from Proposition 5.

Consider first the case in which SCA holds. In view of Proposition 4, it is enough to check that $J(\bar{z}) \subset J(z^*)$ for some solution z^* of SPP(L,X,Y). Take any solution z^* of SPP(L,X,Y). If the result does not hold, there exists j such that $\bar{z}_j = 0$ and $z_j^* > 0$. By Theorem 1(ii)

$$g_j(z_j^*) - g_j(z_j^k) - g_j'(z_j^k)(z_j^* - z_j^k) \le \sum_{\ell=1}^{p} g_\ell(z_\ell^*) - g_\ell(z_\ell^k) - g_\ell'(z_\ell^k)(z_\ell^* - z_\ell^k) =$$

$$D_g(z^*, z^k) \le D_g(z^*, z^0) < \infty. \tag{68}$$

Taking limits in (68) along a subsequence $\{z^{i_k}\}$ converging to $\bar{z}$, we get

$$g_j(z_j^*) - g_j(\bar{z}_j) - z_j^* \lim_{k\to\infty} g_j'(z_j^{i_k}) < \infty. \tag{69}$$

Since $z_j^* > 0$ and $\lim_{k\to\infty} z_j^{i_k} = \bar{z}_j = 0$, (69) implies that $\lim_{t\to 0^+} g_j'(t) > -\infty$. On the other hand, B5 in the case of separable g means that $g_j'(0,\infty) = (-\infty, +\infty)$, which in turn implies, since g_j' is increasing by B1, that $\lim_{t\to 0^+} g_j'(t) = -\infty$. This contradiction establishes that $J(\bar{z}) \subset J(z^*)$. Thus, all the hypotheses of Proposition 4 hold for $\bar{z}$ and so $\bar{z}$ solves SPP(L,X,Y).

Now we proceed exactly as at the end of the proof of Theorem 1. Since $\bar{z}$ is a solution, $\{D_g(\bar{z}, z^k)\}$ is nonincreasing by Theorem 1(ii). By B4 this sequence has a subsequence (corresponding to the subsequence of $\{z^k\}$ which converges to $\bar{z}$) which converges to 0, so that the whole sequence converges to 0, and therefore $\lim_{k\to\infty} z^k = \bar{z}$ by Proposition 1.

Finally, consider a bilinear L, i.e., $L(x,y) = c^t x + b^t y - y^t A x$ with $c \in \mathbf{R}^n$, $b \in \mathbf{R}^m$ and $A \in \mathbf{R}^{m\times n}$. In this case we will invert the procedure of the previous case: we will establish first convergence of the whole sequence to a unique cluster point and then use this uniqueness to prove optimality. In this case L is the Lagrangian of the linear programming problem $\min_x\{c^t x | Ax \ge b, x \ge 0\}$. It is well known that the saddle points of L and the optimal primal-dual pairs (x,y) of the linear programming problem coincide. Denote the gradient components of L by $d_x(y) = \nabla_x L(x,y) = c - A^t y$, $d_y(x) = \nabla_y L(x,y) = b - Ax$, and $d(z) = (d_x(y), -d_y(x))$ with $z = (x,y)$. If a saddle point exists, Tucker [**20**] shows that if $z_j = 0$ in all optimal solutions then there exists an optimal solution z such that $d_j(z) > 0$.

Therefore, we can choose the solution $z^* = (x^*, y^*)$ above with the following additional property

$$z_j^* = 0 \Leftrightarrow d_j(z^*) > 0.$$

¿From (31), (32) and (37) we obtain

$$\Delta_k^* \equiv D_g(z^*, z^k) - D_g(z^*, z^{k+1})$$

$$= -\tau_k d^k(z^* - z^k) - D_g(z^k, z^{k+1})$$

$$= -\tau_k d^k(z^* - w^k) - \tau_k d^k(w^k - z^k) - (1 - \gamma_k)\tau_k \sigma_k,$$

where $w^k = (\xi^k, \eta^k)$. For a bilinear L,

$$d_x^k \equiv \nabla_x L(x^k, \eta^k) = \nabla_x L(\xi^k, \eta^k) \tag{70}$$

and

$$d_y^k \equiv \nabla_y L(\xi^k, y^k) = \nabla_y L(\xi^k, \eta^k), \tag{71}$$

so that

$$-d^k(w^k - z^k) = -[L(\xi^k, \eta^k) - L(x^k, \eta^k)] + [L(\xi^k, \eta^k) - L(\xi^k, y^k)] = \sigma_k \geq 0 \tag{72}$$

and, by Propositions 5 and 3,

$$-d^k(z^* - w^k) = -d^t(z^*)(z^* - w^k) = d^t(z^*)w^k.$$

Combining the equations above we obtain

$$\Delta_k^* = \tau_k d^t(z^*)w^k + \gamma_k \tau_k \sigma_k. \tag{73}$$

By Theorem 1(iii) $\{z^k\}$ is bounded, which implies easily that $\sum_{k=0}^{\infty} \Delta_k^* < \infty$. Consequently, by nonnegativity of all terms in (73),

$$\sum_k \tau_k w_j^k < \infty \tag{74}$$

for all j such that $d_j(z^*) > 0$. Let $\bar{z}$ be a cluster point of $\{z^k\}$. Next, we use (74) to show that the absolute values of increments of the sequence $\{D_g(\bar{z}, z^k)\}$ have a finite sum.

As in (73), using (27) and (37), we obtain

$$\bar{\Delta}_k \equiv D_g(\bar{z}, z^k) - D_g(\bar{z}, z^{k+1})$$

$$= -\tau_k d^k(\bar{z} - z^k) - D_g(z^k, z^{k+1})$$

$$= -\tau_k d^k(\bar{z} - w^k) - \tau_k d^k(w^k - z^k) - (1 - \gamma_k)\tau_k \sigma_k$$

$$= \tau_k d^t(\bar{z})w^k + \gamma_k \tau_k \sigma_k$$

so that

$$|\bar{\Delta}_k| \leq \sum_j \tau_k w_j^k |d_j(\bar{z})| + \gamma_k \tau_k \sigma_k. \tag{75}$$

Consider j such that $d_j(\bar{z}) \neq 0$. Since we have already proved that $\bar{z}$ satisfies (54), (55) and (57), we get $\bar{z}_j = 0$. $\{D_g(z^*, z^k)\}$ is bounded by Theorem 1(ii), so that $\{D_{\tilde{g}_j}(z_j^*, z_j^k)\}$ is bounded. By B5 in the one dimensional case, $\lim_{t\to 0^+} \tilde{g}_j'(t) = -\infty$. Since there exists a subsequence of $\{z_j^k\}$ which converges to $\bar{z}_j = 0$, boundedness of $\{D_{\tilde{g}_j}(z_j^*, z_j^k)\}$ implies that $z_j^* = 0$. It follows from the special choice of z^* that $d_j(z^*) > 0$. It follows from (74) that $\sum_{k=0}^{\infty} \tau_k w_j^k < \infty$. By (38), $\sum_{k=0}^{\infty} \tau_k \sigma_k < \infty$. Hence we have proved that the rightmost expression of (75) is summable in k, and therefore $\{D_g(\bar{z}, z^k)\}$ is a Cauchy sequence which converges to 0 by B4. Hence $\lim_{k\to\infty} z^k = \bar{z}$ by Proposition 1 with $w^k = \bar{z}$.

We have established that the whole sequence $\{z^k\}$ converges to a unique point $\bar{z}$. In view of Corollary 1, since we have shown above that $\bar{z}$ satisfies (54), (55) and (57), it only remains to prove that $\bar{z} = (\bar{x}, \bar{y})$ satisfies (56) and (58). We start with (56). Take j such that $\bar{x}_j = 0$. Suppose that (56) does not hold; i.e. $\nabla_x L(\bar{x}, \bar{y})_j < 0$. By (64), $\lim_{k\to\infty}(x^k, \eta^k) = (\bar{x}, \bar{y})$ so that by continuous differentiability of L there exists $\bar{k}$ such that $\nabla_x L(x^k, \eta^k)_j < 0$ for all $k \geq \bar{k}$. In the separable case (30) becomes

$$\tilde{g}_j'(x^{k+1}) = \tilde{g}_j'(x^k) - \tau_k (d_x^k)_j \tag{76}$$

and Step 2, under the differentiability assumption, is just $d_x^k = \nabla_x L(x^k, \eta^k)$ so that $(d_x^k)_j$ is negative for $k \geq \bar{k}$. Since $\tau_k > 0$ it follows from (76) that $g_j'(x^{k+1}) > g_j'(x^k)$, and, since g_j is strictly increasing by B2, we get $x_j^{k+1} > x_j^k$, for all $k \geq \bar{k}$. Therefore, we have $\bar{x}_j = \lim_{k\to\infty} x_j^k > x_j^{\bar{k}} \geq 0$. This is a contradiction, so that (56) holds for $\bar{x}_j$. The proof of (58) is virtually identical. □

6. An application to linear programming

In this section we present the iterative formulae of the algorithm in the case of linear programming problems, with ξ^k and η^k chosen as in Section 5, and with g chosen as in Example 2. Consider the problem in the form

$$\min c^t x \tag{77}$$

$$\text{s.t.} \quad Ax \geq b \tag{78}$$

$$x \geq 0 \tag{79}$$

with $c \in \mathbf{R}^n$, $b \in \mathbf{R}^m$ and $A \in \mathbf{R}^{m\times n}$, so that the Lagrangian is $L(x, y) = c^t x + b^t y - y^t Ax$ and the constraint sets are $X = \mathbf{R}_+^n$, $Y = \mathbf{R}_+^m$.

The Bregman functions are defined as $g_X(x) = \sum_{j=i}^n x_j \log x_j$, $g_Y(y) = \sum_{i=1}^m y_i \log y_i$. Note that in this case the perturbation vectors ξ and η defined by (43) and (44) can be explicitly solved, and, given the stepsize τ_k, the updated primal and dual solutions can be stated explicitly:

$$\xi_j^k = x_j^k \exp[\lambda(A^t y^k - c)_j] \tag{80}$$

$$\eta_i^k = y_i^k \exp[\mu(b - Ax^k)_i] \tag{81}$$

$$x_j^{k+1} = x_j^k \exp[\tau_k(A^t\eta^k - c)_j] \tag{82}$$

$$y_i^{k+1} = y_i^k \exp[\tau_k(b - A\xi^k)_i]. \tag{83}$$

With $z^k = (x^k, y^k)$, employing $\phi_k(t)$ in (42) and

$$\sigma_k = c^t(x^k - \xi^k) + b^t(\eta^k - y^k) - (\eta^k)^t Ax^k + (y^k)^t A\xi^k,$$

the stepsize τ_k satisfying (28) is obtained via Armijo search.

Based on our method, an experimental code called *Bregman* was written. It is not intended to fully explore the computational capabilities of our method, or its competitiveness with other algorithms, but just its plausibility, showing that its performance is compatible with other methods. Thus we chose to test on a well known and easily accessible collection of LP problems, though we don't claim that our method will be competitive with algorithms specific for LP; in fact, our article is intended to address primarily nonlinear saddle point problems.

Computational results are compared with those obtained from the saddle point method of [**13**]. The latter code is called *Saddle* and it employs the following iterative scheme:

$$x_j^{k+1} = \max[0, x_j^k + \tau_k(A^t\eta^k - c)_j] \tag{84}$$

$$y_i^{k+1} = \max[0, y_i^k + \tau_k(b - A\xi^k)_i], \tag{85}$$

with τ_k proportional to σ_k.

As is the case with *Saddle*, scaling is crucial for efficiency of *Bregman*. Our experience indicates that static (initial) data scaling via equilibration, which traditionally have been employed in LP codes, is not helpful for *Bregman*. That is why the dynamic scaling procedure in *Saddle* adopted from Kallio and Salo [**14**] was further developed for *Bregman*. In *Saddle*, auxiliary *reference quantity* and *value* vectors $\epsilon^k = (\epsilon_i^k) \in R^m$ and $\delta^k = (\delta_j^k) \in R^n$ are defined in each iteration so that

$$\epsilon_i^k = \sum_j |x_j^k a_{ij}| \tag{86}$$

$$\delta_j^k = \sum_i |y_i^k a_{ij}|, \tag{87}$$

where x_j^k and y_i^k denote primal and dual solutions at the beginning of the iteration k. If either ϵ_i^k or δ_j^k is smaller than certain safeguard values, then the safeguard value is used instead of the reference quantity smaller than it. These safeguard values are equal to one tenth of the average of ϵ_i^k and δ_j^k over all variables. The scaling heuristic of *Saddle* employs two phases as follows. In phase one, each row i of A is divided by ϵ_i^k and each column j by δ_j^k. Let $A' = (a'_{ij})$ be the resulting matrix. Let $\bar{a}_i$ be the mean of nonzero values $|a'_{ij}|$ in row i and let $\bar{a}_j$ be the corresponding column average, for each column j. Then, in the second phase of scaling, each element a'_{ij} is divided by the geometric mean of $\bar{a}_i$ and $\bar{a}_j$.

This procedure defines row and column scaling factors for rows and columns in A as well as for rows of b and columns of c^t. Primal and dual variables are scaled by the inverses of these factors. Formulae (80)–(83) are applied in the scaled space. For *Bregman*, the scaling factors of *Saddle* are adjusted in such a way, that the update direction at $(x^k, y^k) > 0$ (obtained by differentiating (82)–(83) with respect to τ_k, at $\tau_k = 0$) is the same as the update direction of *Saddle* obtained from (84)–(85). For example, if D_i is the scaling factor of row i obtained for *Saddle* and we denote by G_i the scaling factor for that row to be applied in *Bregman*, then (83) yields $y_i^{k+1} = y_i^k \exp[\tau_k G_i(b - A\xi^k)_i]$ and (85) yields $y_i^{k+1} = \max[0, y_i^k + \tau_k D_i^2(b - A\xi^k)_i]$. Differentiating these at $(x^k, y^k) > 0$ with respect to τ^k, and evaluating the derivatives at $\tau^k = 0$, yields directional components, which we require to be equal. This implies that $G_i = D_i^2/y_i^k$.

As a termination test for *Bregman* and *Saddle*, we use $V(x^k, y^k) \leq \phi |c^t x^k|$, where

$$V(x^k, y^k) = \sum_i |y_i^k(b - Ax^k)_i| + \sum_j |x_j^k(c - A^T y^k)_j|$$

is an error measure, and $\phi > 0$ is a stopping parameter. If $(b - Ax^k)_i > 0$, then $|y_i^k(b - Ax^k)_i|$ is the value (in units of the objective function) of the primal error; otherwise $|y_i^k(b - Ax^k)_i|$ is the value of complementarity violation. Terms $|x_j^k(c - A^T y^k)_j|$ have similar interpretation, so that $V(x^k, y^k)$ is interpreted as the total value of primal, dual and complementarity violations. Our experience indicates that, if $\phi = 10^{-k}$, then we may expect $k+1$ significant digits in the optimal objective function value. Using our termination test, final infeasibilities are usually small, although they do not enter the stopping criterion directly (see Table 4 below).

Problem	Rows	Columns	Non-zeros
stocfor2	2157	2031	9492
sctap3	1480	2480	10734
ship12l	1151	5427	21597
ship12s	1151	2763	10941
sctap2	1090	1880	8124
ship08l	778	4283	17085
agg2	516	302	4515
degen2	444	534	4449
scsd8	397	2750	11334
sctap1	300	480	2052

Table 1. *The number of rows, columns and non-zeros in the test problems.*

For computational illustration, *Bregman* was tested against *Saddle* on ten problems from the Netlib library [**17**]. For the purpose of this comparison, each problem was first cast in the form given in (77)–(79). Problem names and dimensions are given in Table 1.

All primal and dual variables are set to one initially. We set the perturbation step size parameters in (43)–(44) to $\lambda = \mu = 0.5$, the stepsize test parameters in (28) to $\gamma = 0.3$ and $\beta = 0.7$. Two values for the stopping parameter were applied: $\phi = 10^{-4}$ and $\phi = 10^{-6}$.

Problem	*Bregman* $\phi = 10^{-4}$	*Bregman* $\phi = 10^{-6}$	*Saddle* $\phi = 10^{-4}$	*Saddle* $\phi = 10^{-6}$
stocfor2	29697	63289	11973	40605
sctap3	1761	3907	2726	5623
ship12l	6703	49554	4842	73510
ship12s	6973	43230	4489	51034
sctap2	7282	17456	8143	19574
ship08l	8931	39012	5576	39106
agg2	47023	49856	42997	46769
degen2	8260	31714	6119	32293
scsd8	8419	83400	45413	146610
sctap1	17860	38041	10946	30513

Table 2. *Number of iterations.*

Problem	*Bregman* $\phi = 10^{-4}$	*Bregman* $\phi = 10^{-6}$	*Saddle* $\phi = 10^{-4}$	*Saddle* $\phi = 10^{-6}$
stocfor2	1E-05	6E-07	1E-06	4E-07
sctap3	1E-06	1E-08	4E-06	1E-10
ship12l	2E-05	2E-07	2E-06	2E-07
ship12s	1E-05	5E-08	1E-05	2E-07
sctap2	7E-05	4E-07	2E-05	4E-07
ship08l	6E-05	3E-07	4E-05	6E-08
agg2	3E-06	5E-07	4E-06	8E-08
degen2	3E-05	4E-08	2E-05	2E-08
scsd8	2E-05	2E-08	3E-06	3E-08
sctap1	1E-05	2E-07	8E-06	2E-08

Table 3. *Relative error in the optimal objective function value.*

Problem	*Bregman*	*Saddle*
stocfor2	3E-05	3E-06
sctap3	2E-05	1E-05
ship12l	1E-05	3E-05
ship12s	3E-05	6E-06
sctap2	1E-05	9E-06
ship08l	8E-06	4E-04
agg2	1E-04	6E-04
degen2	1E-06	2E-06
scsd8	8E-05	4E-05
sctap1	1E-06	4E-06

Table 4. *Maximal final infeasibilities of Bregman and Saddle;* $\phi = 10^{-6}$.

Table 2 shows the iteration count for $\phi = 10^{-6}$ and $\phi = 10^{-6}$, both for *Bregman* and *Saddle*. Table 3 shows the corresponding relative objective errors $|c^t\bar{x} - c^t x^*|/|c^t x^*|$, where $\bar{x}$ is the primal solution obtained at the end of iterations and x^* is a true optimal solution.

We observe that the iteration counts for both methods are of the same order of magnitude. Also the precision in both cases is approximately the same (five and seven significant digits as predicted for $\phi = 10^{-4}$ and $\phi = 10^{-6}$, respectively). Due to function evaluations (for the exponent functions) and due to the search for step sizes, work per iteration for *Bregman* is somewhat larger than for *Saddle*. However, at this early stage of development we consider *Bregman* highly promising. In particular, development of a suitable scaling method for *Bregman* is expected to improve its performance significantly.

For each iteration k, define relative primal and dual infeasibilities, denoted by e_P^k and e_D^k, respectively, as follows:

$$e_P^k = \max_i\{0, (b - Ax^k)_i/(|b_i| + \epsilon_i^k)\},$$

$$e_D^k = \max_j\{0, (A^t y^k - c)_j/(|c_j| + \delta_j^k)\}.$$

Table 4 reports the final infeasibilities in terms of $\max\{e_P^k, e_D^k\}$ for *Bregman* and *Saddle*, with $\phi = 10^{-6}$.

To see the importance of scaling, we ran *Bregman* for a dozen of small *Netlib* problems without scaling for $\phi = 10^{-4}$. Five out our twelve test problems failed to converge in one million iterations, while for the others the number of iterations was increased by a factor ranging from 3 to 22 as compared with the scaled version.

As noted before, we may update the scaling factors during a prespecified number of iterations only. To demonstrate, that it pays off to update scaling until the end, we run our smallest problem *sctap1* fixing scaling after 5000 iterations. As a result, the number of iterations is 26088, for $\phi = 10^{-4}$, and 63631, for $\phi = 10^{-6}$. Thus, freezing of scaling almost doubles the number of iterations as compared with the figures in Table 2 for *sctap1*. Hence, the frozen scaling strategy, for which also we have computational convergence, is the one to which our theoretical results apply, but updating of scaling factors to the end is plainly more efficient.

Finally, it is important to note that for small problems both of these approaches often turn out to be inefficient. Their comparative advantage is expected to appear for large-scale problems, and in particular, our method may be of interest for massively parallel computing.

References.

[1] Bertsekas, D., Tsitsiklis, J.N. *Parallel and Distributed Computation: Numerical Methods*. Prentice Hall, New Jersey (1989).

[2] Bregman, L.M. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics* **7**,3 (1967) 200-217.

[3] 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.

[4] Censor, Y., Lent, A. An iterative row-action method for interval convex programming. *Journal of Optimization Theory and Applications* **34** (1981) 321-353.

[5] Censor, Y., Zenios, S.A. The proximal minimization algorithm with D-functions. *Journal of Optimization Theory and Applications* **73** (1992) 451-464.

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

[7] De Pierro, A.R., Iusem, A.N. A relaxed version of Bregman's method for convex programming. *Journal of Optimization Theory and Applications* **51** (1986) 421-440.

[8] Eckstein, J. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Mathematics of Operations Research* **18** (1993) 202-226.

[9] Harker, P.T., Pang, J.S. Finite dimensional variational inequalities and nonlinear complementarity problems: a survey of theory, algorithms and applications. *Mathematical Programming* **48** (1990) 161-220.

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

[11] Iusem, A.N. An iterative algorithm for the variational inequality problem. *Computational and Applied Mathematics* **13** (1994) 103-114.

[12] Iusem, A.N. An interior point method for the nonlinear complementarity problem. *Applied Numerical Mathematics* **24** (1997) 469-482.

[13] Kallio, M., Ruszczyński, A. Perturbation methods for saddle point computation. Working Paper 94-38, Institute for Applied Systems Analysis, Laxenburg, Austria (1994).

[14] Kallio, M., Salo, S. Tatonnement procedures for linearly constrained convex optimization. *Management Science* **40** (1994) 788-797.

[15] Khobotov, E.N. Modifications of the extragradient method for solving variational inequalities and certain optimization problems. *USSR Computational Mathematics and Mathematical Physics* **27**,5 (1987) 120-127.

[16] Korpelevich, G.M. The extragradient method for finding saddle points and other problems. *Ekonomika i Matematcheskie Metody* **12** (1976) 747-756.

[17] Netlib, LP Test Problems, Bell Laboratories.

[18] Rockafellar, R.T. Local boundedness of nonlinear monotone operators. *Michigan Mathematical Journal* **16** (1969) 397-407.

[19] Rockafellar, R.T. *Convex Analysis*. Princeton University Press, New Jersey (1970).

[20] Tucker, A.W. Dual systems of homogeneous linear relations. In *Linear Inequalities and Related Systems* (H.W. Kuhn and A.W. Tucker, editors). *Annals of Mathematics Studies*, Princeton University Press, Princeton **38** (1956) 3-18.