International Institute for Applied Systems Analysis
IIASA
www.iiasa.ac.at

Equilibrium Programming Using Proximal-Like Algorithms

Flam, S.D.

IIASA Working Paper

June 1995

Working Paper

Equilibrium Programming Using Proximal–Like Algorithms

Sjur Didrik Flåm

WP-95-55
June 1995

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

Equilibrium Programming Using Proximal–Like Algorithms

Sjur Didrik Flåm

WP-95-55
June 1995

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

Abstract

We consider problems where solutions – called equilibria – emerge as fixed points of an extremal mapping. Examples include convex programming, convex – concave saddle problems, many noncooperative games, and quasi – monotone variational inequalities. Using Bregman functions we develop proximal – like algorithms for finding equilbria. At each iteration we allow numerical errors or approximate solutions.

Key words: Proximal minimization, mathematical programming, Bregman functions.

EQUILIBRIUM PROGRAMMING USING PROXIMAL-LIKE ALGORITHMS

Sjur Didrik Flåm
Economics Department, Bergen University, 5007 Norway

1. INTRODUCTION Numerous problems in optimization and economics reduce to find a vector x^* satisfying the fixed point condition

$$x^* \in \text{argmin}\{F(x^*, x) : x \in X\}. \qquad (1.1)$$

Here X is a nonempty closed convex subset of some Euclidean space E, and the bivariate function $F: X\text{x}X \to R$ is convex in its second coordinate. E is endowed with the standard inner product $\langle\cdot,\cdot\rangle$, generating the customary norm $\|\cdot\|$.

Our purpose is to solve (1.1). Usually this is a well defined task since solutions - henceforth named *equilibria* - are indeed available under general conditions:

Proposition 1 (Existence of equilibrium). *Suppose* X *is nonempty compact convex, and* $F(x,y)$ *is jointly lower semicontinuous, separately continuous in* x *and convex in* y. *Then* (1.1) *admits at least one solution.*

Proof. The correspondence $X \ni x \to \text{argmin}\{F(x,y) : y \in X\}$ has nonempty convex values and closed graph. Hence by Kakutani´s theorem there exists a fixed point. ♥

For computational reasons we shall restrict attention to a certain class of equilibrium problems.

Definition *Problem* (1.1) *is said to be of* ***saddle type*** *if for every equilibrium* x^* *and* $x \in X$ *we have*

$$F(x, x^*) \leq F(x, x). \qquad (1.2)$$

Problems fitting format (1.1) and satisfying (1.2) abound, as illustrated by important examples in Section 2. A prominent case included there, namely *monotone variational inequalities,* helps to put the subsequent development in perspective. Indeed, given a mapping $X \ni x \rightarrow m(x) \in E$, let $F(x,y) = \langle m(x), y - x\rangle$. Then x^* solves (1.1) $\Leftrightarrow \langle m(x^*), x - x^*\rangle \geq 0, \ \forall x \in X$. Moreover, (1.2) would follow from the monotonicity: $\langle m(x) - m(x^*), x - x^*\rangle \geq 0$. Granted this last property, it is well known that *proximal point algorithms* (Rockafellar 1976), (Güler 1991) give good convergence, but they are often hard to execute.

This motivates us to consider here new versions of proximal-like algorithms, especially adapted to the unifying framework (1.1). Section 3 states the said algorithms, all inspired by the the iteration $x^{k+1} \in \operatorname{argmin}\{F(x^k,x) : x \in X\}$. In line with recent devlopments of Censor & Zenios (1992), Eckstein (1993), Chen & Teboulle (1993), Bertsekas & Tseng (1994) we shall accomodate Bregman functions and tolerate approximate evaluations. A main novelty is the procedure where regularization is done twice at every stage: first to predict the next iterate, thereafter to update the current point. Section 4 contains the convergence analysis.

2. EXAMPLES

This section offers a list of problems all fitting format (1.1). We begin with

Convex minimization Let $F(x,y) = f(y)$ with $f:X \rightarrow R$ convex. Then x^* solves (1.1) $\Leftrightarrow$ $x^* \in \operatorname{argmin}\{f(x): x \in X\}$. In this instance (1.2) is automatically satisfied. ♥

Convex-concave saddle problems Let $X = X_1 x X_2$ be a product of two nonempty closed convex sets, $F(x, y) = L(y_1,x_2) - L(x_1,y_2)$ with $x = (x_1,x_2)$, $y = (y_1,y_2)$, and L convex-concave. Then x^* solves (1.1) $\Leftrightarrow$ x^* is a saddle point of L. The saddle property (1.2) holds in this case as well. ♥

Noncooperative games with convex costs Generalizing the saddle problem, let individual $i \in I$, (I finite), incur convex cost $f_i(x_{-i}, x_i)$ in own decision $x_i \in X_i$, the latter set being nonempty closed convex. Here x_{-i} is short notation for actions taken by i's rivals. Let $X := \prod X_i$ and $F(x, y) := \Sigma_i f_i(x_{-i}, y_i)$. Then x^* solves (1.1) $\Leftrightarrow$ x^* is a Nash equilibrium. Property (1.2) is somewhat stringent in this case. In particular, it holds when $F(x,x) - F(x,y)$ is convex in x. For a discussion see Flåm & Ruszczynski (1994), Antipin & Flåm (1994). ♥

Variational inequalities Let $X \ni x \to G(x)$ be a correspondence with nonempty compact convex values. When $F(x,y) := \sup\{<g, y - x> : g \in G(x)\}$, we get that x^* solves (1.1) $\Leftrightarrow$ $\exists\, g^* \in G(x^*)$ such that $<g^*, x - x^*> \geq 0, \forall x \in X$. Here condition (1.2) holds if G is *quasi-monotone at equilibrium* x^* in the sense that for all $x \in X$

$$\sup\{<g^*, x - x^*> : g^* \in G(x^*)\} \geq 0 \quad \Rightarrow \quad \sup\{<g, x - x^*> : g \in G(x)\} \geq 0.$$ ♥

Successive approximations Related to variational inequalities is the following optimization procedure. Let $f:X \to R$ be convex and differentiable. Then, with $F(x,y) = f(x) + <f'(x), y - x>$, we have that x^* solves (1.1) $\Leftrightarrow$ $x^* \in \operatorname{argmin}\{f(x): x \in X\}$. In this instance (1.2) is automatically satisfied.
Likewise, if $X \ni x \to G(x)$ is differentiable with $G'(x)$ positive semidefinite, and $F(x,y) = <G(x), y -x> + <y-x,G'(x)(y-x>/2$, then x^* solves (1.1) $\Leftrightarrow$ $<G(x^*), x - x^*> \geq 0, \forall x \in X$. ♥

3. ALGORITHMS

This section proposes two procedures to solve (1.1). Both are ammendments of

$$x^{k+1} \in \operatorname{argmin}\{F(x^k, x) : x \in X\}. \qquad (3.1)$$

Our motivation stems from three deficiencies of (3.1). *Firstly*, it is unreasonable - at least in practice - to insist that argmin in (3.1) be computed exactly at every stage k. Rather one should tolerate some error $\varepsilon_k \geq 0$. *Secondly*, the argmin operation - whether executed exactly or not - may cause instabilities. In particular, this happens often when $F(x,y)$ is affine in y. (See the above examples on variational inequalities). *Thirdly*, (3.1) reflects some myopia in minimizing at the current outcome x^k in lieu of at some predicted point, henceforth denoted x^{k+}.

These considerations lead us to replace (3.1) by more stable and flexible algorithms. For their statement we need to recall the notion of a *Bregman function*.

Definition *Let* S *be an open convex subset of the ambient Euclidean space* E. *Then* $\psi : \mathrm{cl}S \to R$ *is baptized a* ***Bregman function*** *with* ***zone*** S *and "distance"*

$$D(x,y) := \psi(x) - \psi(y) - \langle \psi'(y), x - y \rangle$$

if the following conditions hold:

(i) ψ *is continuously differentiable on* S;

(ii) ψ *is strictly convex continuous on* clS;

(iii) *for any number* $r \in R$ *and points* $x \in \mathrm{cl}S$, $y \in S$ *the two level sets*

$$\{x \in \mathrm{cl}S : D(x,y) \le r\} \quad \textit{and} \quad \{y \in S : D(x,y) \le r\}$$

are both bounded;

(iv) $S \ni y^k \to y \Rightarrow D(y, y^k) \to 0$;

(v) *if* $\{x^k\}$ *and* $\{y^k\}$ *are bounded sequences such that* $y^k \to y \in \mathrm{cl}S$ *and* $D(x^k,y^k) \to 0$, *then* $x^k \to y$.

Examples of such functions are given by Censor &Zenios (1992), Teboulle (1992), Eckstein (1993), Chen & Teboulle (1993). Generalizations are found in Kiwiel (1994a). (Of particular importance and convenience is the instance $\psi = \|\cdot\|^2/2$, yielding $D(x,y) = \|x-y\|^2/2$). Since X is bounded condition (iii) is not needed in the sequel. Now, with this notion in hand, employing a fixed Bregman function ψ we shall consider iterative procedures of the type

$$x^{k+1} \in \varepsilon_k\text{-argmin}\{\alpha_k F(x^{k+},x) + D(x,x^k) : x \in X\}, \tag{3.2}$$

the initial point $x^o \in X$ being arbitrary. An explanation of (3.2) is in order. The parameter $\varepsilon_k \ge 0$ there is an error tolerance. For asymptotic accuracy we invariably posit that

$$\Sigma_k \varepsilon_k^{1/2} < +\infty. \tag{3.3}$$

The other parameter $\alpha_k > 0$ in (3.2) is a matter of relative free choice. It should be bounded away from 0 and $+\infty$. More will be said about appropriate specifications later. The penalty term

$$D(x,x^k) = \psi(x) - \psi(x^k) - \langle \psi'(x^k), x - x^k \rangle$$

in (3.2), being the "distance" associated to a fixed *Bregman function* ψ with *zone* S $\supset$ X, is intended to lend some inertia and stability to the adjustment process. Finally, the vector x^{k+} in (3.2) stands for a "prediction" or approximation of x^{k+1} to be defined in two alternative manners. One simply requires $x^{k+} = x^{k+1}$. The other makes for a special step to find x^{k+}, going as follows

$$x^{k+} \in \varepsilon_k\text{-argmin}\{\alpha_k F(x^k, x) + D(x,x^k) : x \in X\} \tag{3.4}$$

Algorithms of the sort (3.2-4), or akin to this procedure, have been studied recently by Antipin & Flåm (1994), Bertsekas & Tseng (1994), Kiwiel (1994b), Chen & Teboulle (1993), Eckstein (1993). However, none of these accomodate as much generality as done here. Typically these studies focus on convex minimization, or make the choice $\varepsilon_k = 0$, or employ $\psi = \|\cdot\|^2/2$. Our purpose is to lift these restrictions.

4. CONVERGENCE

Throughout the rest *we assume that the hypotheses of Proposition 1 and condition (1.2) are all in vigour.* Also, we posit that *the Bregman function* ψ *has a zone* S *containing* X, *with Lispschitz continuous gradient.* Specifically, there exists some constant $L > 0$, such that for any error tolerance ε used in the sequel it holds

$$\|\psi'(x) - \psi'(y)\| \le L\|x - y\|. \qquad (4.1)$$

whenever $x \in X$ and $\text{dist}(y, X) \le \varepsilon^{1/2}$. Three auxiliary results are needed.

Lemma 1 *Suppose a function* f *is finite-valued convex near some nonempty closed convex subset* X *of the ambient Euclidean space. For fixed* $\xi \in X$, *and error tolerance* $\varepsilon \ge 0$ *let*

$$x^+ \in \varepsilon\text{-argmin}\{f(x) + D(x, \xi) : x \in X\}.$$

Then, for some $\delta \in [o, \varepsilon]$ *and all* $x \in X$,

$$f(x) + D(x,\xi) \ge f(x^+) + D(x^+,\xi) + D(x,x^+) - \delta - (L+1)(\varepsilon - \delta)^{1/2} \|x - x^+\|.$$

Proof. The ε-optimality of x^+ implies that

$$0 \in \varepsilon\text{-}\partial[f + D(\cdot,\xi) + I_X](x^+)$$

where ε-∂ denotes the ε-subdifferential operator, and I_X is the convex indicator of X (i.e., I_X equals 0 on X, and $+\infty$ elsewhere). By Hiriart-Urruty & Lemarechal (1991, Thm. XI 3.1.1) there exist "subgradients"

$$s_1 \in \varepsilon_1\text{-}\partial f(x^+), \qquad s_2 \in \varepsilon_2\text{-}\partial D(\cdot,\xi)(x^+), \qquad s_3 \in \varepsilon_3\text{-}\partial I_X(x^+)$$

with $\varepsilon_1, \varepsilon_2, \varepsilon_3 \ge 0$ such that

$$\varepsilon_1 + \varepsilon_2 + \varepsilon_3 = \varepsilon \quad \text{and} \quad 0 = s_1 + s_2 + s_3. \tag{4.2}$$

Now, $s_1 \in \varepsilon_1\text{-}\partial f(x^+)$ implies

$$f(x) \geq f(x^+) + \langle s_1, x - x^+\rangle - \varepsilon_1 \quad \text{for all } x \in X.$$

Adding the *three-point identitity* (see Chen & Teboulle 1993)

$$D(x,\xi) = D(x^+,\xi) + D(x,x^+) + \langle\psi'(x^+) - \psi'(\xi), x - x^+\rangle$$

to the above subgradient inequality, we obtain

$$f(x) + D(x,\xi) \geq f(x^+) + D(x^+,\xi) + D(x,x^+) + \langle s_1 + \psi'(x^+) - \psi'(\xi), x - x^+\rangle - \varepsilon_1 \tag{4.3}$$

In turn, $s_2 \in \varepsilon_2\text{-}\partial D(\cdot,\xi)(x^+)$ implies $s_2 = S_2 - \psi'(\xi)$ for some $S_2 \in \varepsilon_2\text{-}\partial\psi'(x^+)$. By the Brönsted-Rockafellar theorem (see Hiriart-Urruty & Lemarechal 1993, Thm. XI, 4.2.1) there exists $y \in B(x^+, \varepsilon_2^{1/2})$ such that $\|\psi'(y) - S_2\| \leq \varepsilon_2^{1/2}$. Drawing upon these facts and (4.2) we have

$$s_1 + \psi'(x^+) - \psi'(\xi) = s_1 + s_2 + s_3 + \psi'(x^+) - S_2 - s_3$$

$$= \psi'(x^+) - S_2 - s_3 \qquad = \psi'(x^+) - \psi'(y) + \psi'(y) - S_2 - s_3$$

so, using $\langle s_3, x - x^+\rangle \leq \varepsilon_3$, it follows that $\langle s_1 + \psi'(x^+) - \psi'(\xi), x - x^+\rangle$

$$= \langle\psi'(x^+) - \psi'(y) + \psi'(y) - S_2 - s_3, x - x^+\rangle$$

$$\geq -(\|\psi'(x^+) - \psi'(y)\| + \|\psi'(y) - S_2\|)\|x - x^+\| - \langle s_3, x - x^+\rangle$$

$$\geq -(L\varepsilon_2^{1/2} + \varepsilon_2^{1/2})\|x - x^+\| - \varepsilon_3 .$$

Using this last inequality in (4.3) the desired conclusion follows immediately with $\delta = \varepsilon_1 + \varepsilon_3$ and $\varepsilon_2 = \varepsilon - \delta$.♥

Lemma 2 *Suppose a function* f *is finite-valued convex near some nonempty closed convex subset* X *of the ambient Euclidean space. Then*

$$x^* \in \operatorname{argmin}\{f(x) + D(x,x^*) : x \in X\} \Leftrightarrow x^* \in \operatorname{argmin}\{f(x) : x \in X\}.$$

Proof. $\Rightarrow$ By Lemma 1, $f(x) + D(x,x^*) \geq f(x^*) + D(x^*,x^*) + D(x,x^*)$ for all $x \in X$, whence $f(x) \geq f(x^*)$ for all $x \in X$. Conversely, when $f(x) \geq f(x^*)$ for all $x \in X$, it holds that $f(x) + D(x,x^*) \geq f(x^*) + D(x^*,x^*)$ for all $x \in X$. ♥

Lemma 3 *Suppose* $\{a_k\},\{b_k\}, \{c_k\}$ *are sequences of nonnegative numbers such that* $\Sigma_k b_k <+\infty$, *and*

$$a_{k+1} \leq a_k + b_k - c_k.$$

Then $\{a_k\}$ *converges, and* $\Sigma_k c_k < +\infty$.

Proof. From $a_K + \Sigma_{k<K} c_k \leq a_o + \Sigma_{k<K} b_k$ it follows that $\{a_k\}$ is bounded and $\Sigma_k c_k <+\infty$. Let a be any cluster point of $\{a_k\}$. The inequality $a_\kappa \leq a_K + \Sigma_{k\geq K} b_k$ valid for all $\kappa > K$, implies that $\{a_k\}$ has no cluster point $> a$, whence $\{a_k\}$ converges. ♥

Theorem 1 (Convergence under "correct" predictions). *For arbitrary initial* $x^o \in X$, *the process* (3.2) *with* $x^{k+} = x^{k+1}$ *converges to equilibrium.*

Proof. For any equilibrium x^* Lemma 1 yields $\alpha_k F(x^{k+1},x^*) + D(x^*,x^k) \geq$

$$\alpha_k F(x^{k+1},x^{k+1}) + D(x^*,x^{k+1}) + D(x^{k+1},x^k) - \delta_k - (L+1)(\varepsilon_k - \delta_k)^{1/2} \|x^* - x^{k+1}\|$$

for some $\delta_k \in [o,\varepsilon_k]$. Invoking now the saddle property $F(x^{k+1},x^*) \leq F(x^{k+1},x^{k+1})$ we have

$$D(x^*,x^k) \geq D(x^*,x^{k+1}) + D(x^{k+1},x^k) - \delta_k - (L+1)\varepsilon_k^{1/2} \|x^* - x^{k+1}\|$$

Using here (3.3), the boundedness of X, and Lemma 3 it obtains from the last inequality that $D(x^*,x^k)$ converges, and $\Sigma_k D(x^{k+1},x^k) < +\infty$. In particular, $D(x^{k+1},x^k) \to 0$. Let x^* be an acumulation point of $\{x^k\}$. Then, for some subsequence K, $\lim_{k\in K} x^k = \lim_{k\in K} x^{k+1} = x^*$, and $\lim_{k\in K} \alpha_k = \alpha > 0$ Passing to the limit along this subsequence in (3.2) we obtain

$$x^* \in \operatorname{argmin}\{\alpha F(x^*, x) + D(x,x^*) : x \in X\}$$

which by Lemma 2 is equivalent to (1.1). Thus $\{x^k\}$ clusters to an equilibrium x^*, and $\{D(x^*,x^k)\}$ converges to zero. It follows that the entire sequence $\{x^k\}$ converges to x^*. ♥

When $F(x,y)$ is subdifferentiable in y near X, $M(x) := \partial_y F(x,x) + \partial I_X(x)$, $\varepsilon_k = 0$, and $\psi = \|\cdot\|^2/2$, the procedure of Thm.1 is tantamount to the exact proximal point algorithm of Rockafellar (1976). To wit, the iteration in Thm. 1 then comes in the form $x^{k+1} = (I + \alpha_k M)^{-1}(x^k)$, recently generalized by Eckstein (1993). The requirement $x^{k+} = x^{k+1}$ in Thm.1, may make however, for laborious iterations (3.2). Essentially, the difficulty stems from the fact that (1.1) has two related features, namely: prediction in the first variable and optimization in the second. (3.4) serves to separate these two aspect from each other. For success in these matters we need some smoothness of F, and the parameters α_k must not be too large. Specifically, we assume there exists a constant $\Lambda > 0$ such that on X we have

$$\|F(x+\Delta x,y+\Delta y) - F(x,y+\Delta y) - F(x+\Delta x,y) + F(x,y)\| \le 2\Lambda\{D(x,x+\Delta x)D(y+\Delta y,y)\}^{1/2} \quad (4.4)$$

This seemingly strange condition simplifies, when $\psi = \|\cdot\|^2/2$, to

$$\|F(x+\Delta x,y+\Delta y) - F(x,y+\Delta y) - F(x+\Delta x,y) + F(x,y)\| \le \Lambda\|\Delta x\| \, \|\Delta y\|,$$

which holds when X is compact and F is continuously differentiable.

Theorem 2 (Convergence under regularized predictions). *Suppose* $\{\alpha_k\Lambda\}$ *is contained in a closed subinterval of* $]0,1[$ *with* Λ *satisfying* (4.4). *Then for arbitrary initial* $x^o \in X$, *the process* (3.2- 4) *converges to equilibrium.*

Proof. Applying Lemma 1 to situation (3.4) we get $\alpha_k F(x^k,x^{k+1}) + D(x^{k+1},x^k) \ge$

$$\alpha_k F(x^k,x^{k+}) + D(x^{k+},x^k) + D(x^{k+1},x^{k+}) - \delta_k - (L+1)(\varepsilon_k - \delta_k)^{1/2} \|x^{k+1} - x^{k+}\|.$$

The same Lemma 1 applied to (3.2) yields, when x^* is any equilibrium,

$$\alpha_k F(x^{k+},x^*) + D(x^*,x^k) \ge$$

$$\alpha_k F(x^{k+},x^{k+1}) + D(x^{k+1},x^k) + D(x^*,x^{k+1}) - \delta_k - (L+1)(\varepsilon_k - \delta_k)^{1/2} \|x^* - x^{k+1}\|.$$

Adding these two inequalities we have

$$\alpha_k[F(x^k,x^{k+1}) - F(x^{k+},x^{k+1}) - F(x^k,x^{k+}) + F(x^{k+},x^*)] \ge$$

$$D(x^{k+},x^k) + D(x^{k+1},x^{k+}) - \delta_k - (L+1)(\varepsilon_k - \delta_k)^{1/2} \|x^{k+1} - x^{k+}\| +$$

$D(x^*,x^{k+1}) - \delta_k - (L+1)(\varepsilon_k - \delta_k)^{1/2} \|x^* - x^{k+1}\| - D(x^*,x^k)$

Now invoke the saddle property $F(x^{k+},x^*) \le F(x^{k+},x^{k+})$ and the Lipschitz condition (4.4) to have $2\alpha_k\Lambda\{D(x^{k+},x^k)D(x^{k+1},x^{k+})\}^{1/2} \ge$

$$\alpha_k[F(x^k,x^{k+1}) - F(x^{k+},x^{k+1}) - F(x^k,x^{k+}) + F(x^{k+},x^{k+})] \ge$$

$$\alpha_k[F(x^k,x^{k+1}) - F(x^{k+},x^{k+1}) - F(x^k,x^{k+}) + F(x^{k+},x^*)].$$

Combining the two last strings of inequalities we get

$$D(x^*,x^k) \ge D(x^*,x^{k+1}) + \{D(x^{k+},x^k)^{1/2} - \alpha_k\Lambda D(x^{k+1},x^{k+})^{1/2}\}^2 +$$

$$[1-(\alpha_k\Lambda)^2](D(x^{k+1},x^{k+}) - 2\delta_k - (L+1)(\varepsilon_k-\delta_k)^{1/2}\{\|x^{k+1}-x^{k+}\| + \|x^*-x^{k+1}\|\}.$$

This yields - by (3.3), the boundedness of X, and Lemma 3 - that $D(x^*,x^k)$ converges and

$$\sum_k \{D(x^{k+},x^k)^{1/2} - \alpha_k\Lambda D(x^{k+1},x^{k+})^{1/2}\}^2 + [1-(\alpha_k\Lambda)^2](D(x^{k+1},x^{k+}) < +\infty.$$

It follows that $D(x^{k+},x^k) \to 0$ and $D(x^{k+1},x^{k+}) \to 0$. Let x^* be an acumulation point of $\{x^k\}$. Then, for some subsequence K, $\lim_{k\in K} x^k = \lim_{k\in K} x^{k+} = \lim_{k\in K} x^{k+1} = x^*$, and $\lim_{k\in K} \alpha_k = \alpha > 0$. Passing to the limit along this subsequence in (3.2) we obtain

$$x^* \in \operatorname{argmin}\{\alpha F(x^*, x) + D(x,x^*) : x \in X\}$$

which by Lemma 2 is equivalent to (1.1). Thus $\{x^k\}$ clusters to an equilibrium x^*, and $\{D(x^*,x^k)\}$ converges to zero. It follows that the entire sequence $\{x^k\}$ converges to x^*. ♥

Clearly, in (3.4) one might use a sequence $\{\varepsilon_{k+}\}$ of errors different from $\{\varepsilon_k\}$ but also satisfying (3.3).

When f is convex differentiable on X, $\varepsilon_k = 0$, and $F(x,y) = \langle f'(x), y - x\rangle$, the steps of Thm. 2 assume the form: $\langle f'(x^{k+}), x - x^{k+1}\rangle \ge 0$ for all $x \in X$, reminiscent of the extragradient method of Korpelevich (1976).

It appears interesting to incorporate variational convergence of functions $F^k \to F$, and sets $X^k \to X$, as done by Alart & Lemaire (1991). However, this falls outside the scope of this paper.

REFERENCES

P. Alart and B. Lemaire, "Penalization in non-classical convex programming via variational convergence," *Mathematical Programming* 51, 307-331 (1991).

A.S. Antipin, "Controlled proximal differential systems for saddle problems, " *Differential Equations* vol. 28, no. 11, 1498-1510 (1992).

A.S. Antipin and S.D. Flåm,"Noncooperative games; Iterative approaches to Nash equilibrium, " manuscript (1994).

D.P. Bertsekas and P. Tseng, "Partial proximal point algorithms for convex programming, " *SIAM J. Optimization* vol. 4, no. 3, 551-572(1994).

Y. Censor and S. A. Zenios, "Proximal minimization algorithms with D-functions," *J. Opt. Theory and Applications* vol. 27, no.3, 451-464 (1992).

G. Chen and M. Teboulle, " Convergence analysis of a proximal-like minimization algorithm using Bregman functions," *SIAM J. Optimization* vol. 3, no. 3, 538-543 (1993).

J. Eckstein, " Nonlinear proximal point algorithms using Bregman projections, with applications to convex programming, " *Mathematics of Operations Research* vol. 18, no. 1, 202-226 (1993).

S.D. Flåm and A. Ruszczynski, "Noncooperative games: Computing equilibrium by partial regularization," Working paper IIASA 42 (1994).

J-B. Hiriart-Urruty and C. Lemarechal, *Convex Analysis and Minimization Algorithms,* Springer-Verlag, Berlin (1993).

K. C. Kiwiel, "Free-steering relaxation methods for problems with strictly convex costs and linear constraints," Tech. rep., Systems Research Institute, Warsaw (1994a).

K. C. Kiwiel, "Proximal minimization methods with generalized Bregman functions," Tech. rep., Systems Research Institute, Warsaw (1994b).

G.M. Korpelevich, "The extragradient method for finding saddle points and other problems," *Ekon. i Mat. Metody* 12, 747-756 (1976).

B. Lemaire, "The proximal point algorithm," *International Series of Numerical Mathematics,* Birkhaüser-Verlag, Basel, vol. 87, 73-87 (1989).

R.T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM J. Control and Optimization* 14, 5, 877-898 (1976).

M. Teboulle, "Entropic proximal mappings with applications to nonlinear programming," *Mathematics of Operations Research* 17, 670-690 (1992).