Mathematical Programming 78 (1997) 29–41

Equilibrium programming using proximal-like algorithms

Sjur Didrik Flåm [a,*], Anatoly S. Antipin [b]

[a] *Economics Department, Bergen University, 5007 Bergen, Norway*
[b] *Computing Centre, Russian Academy of Sciences, Moscow 117967, Russia*

Received 13 December 1994; revised 20 January 1996

Abstract

We compute constrained equilibria satisfying an optimality condition. Important examples include convex programming, saddle problems, noncooperative games, and variational inequalities. Under a monotonicity hypothesis we show that equilibrium solutions can be found via iterative convex minimization. In the main algorithm each stage of computation requires two proximal steps, possibly using Bregman functions. One step serves to predict the next point; the other helps to correct the new prediction. To enhance practical applicability we tolerate numerical errors.

Keywords: Equilibrium problems; Convex programming; Saddle problems; Non-cooperative games; Variational inequalities; Quasi-monotonicity; Proximal point algorithms; Bregman distances

1. Introduction

Numerous problems in physics, optimization, and economics reduce to find a vector x^* satisfying a fixed point condition

$$x^* \in \operatorname{argmin}\{F(x^*, x) \mid x \in X\}. \tag{1.1}$$

Here X is a nonempty compact convex subset of $E := \mathbb{R}^n$, this space being endowed with the standard inner product $\langle \cdot, \cdot \rangle$ and the associated norm $\|\cdot\|$. Moreover, the bivariate function $F: X \times X \to \mathbb{R}$ is convex in its second co-ordinate.

* Corresponding author. Email: sjur.flaam@econ.uib.no. Research supported partly by the Norwegian Research Council, project: Quantec 111039/401.

PII S0025-5610(96)00071-8

Our purpose is to solve (1.1). Usually that enterprise is well founded since solutions – henceforth named *equilibria* – are indeed available under general conditions:

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

Proof. The lower semicontinuity and convexity of $F(x,y)$ in y imply, since X is compact convex, that the correspondence $X \ni x \to A(x) := \operatorname{argmin}\{F(x,y) \mid y \in X\}$ has nonempty convex values. Given any sequence $(x^k, y^k) \to (x,y)$ with $y^k \in A(x^k)$, then for arbitrary $\xi \in X$, $F(x,y) \le \liminf F(x^k, y^k) \le \liminf F(x^k, \xi) = F(x, \xi)$. Therefore, $y \in A(x)$, i.e., A has closed graph, and, by Kakutani's theorem, there exists a fixed point $x^* \in A(x^*)$. □

For computational reasons, related to convergence, we shall consider only a distinguished subclass of equilibrium problems:

Definition 1.2. Problem (1.1) is said to be *monotone* if for every equilibrium x^* and vector $x \in X$ we have

$$F(x,x^*) \le F(x,x). \tag{1.2}$$

Problems fitting (1.1) and satisfying (1.2) abound, as illustrated by important examples in Section 2. A prominent case – included there – helps to advertize the subsequent development, and to provide a useful perspective. Namely, given a mapping $m: X \to E$, let $F(x,y) = \langle m(x), y - x \rangle$. Then x^* satisfies (1.1) iff it solves the variational inequality

$$\langle m(x^*), x - x^* \rangle \ge 0 \quad \text{for all } x \in X, \tag{1.3}$$

see [19]. Moreover, in this case (1.2) would follow from *quasi-monotonicity* with respect to equilibrium, this notion meaning that $\langle m(x), x - x^* \rangle \ge 0$, $\forall x \in X$. A fortiori, (1.2) holds under customary monotonicity: $\langle m(x) - m(x^*), x - x^* \rangle \ge 0$, $\forall x, x^* \in X$. Then *proximal point algorithms* converge well, but suffer frequently from being hard to execute [4,6,28].

This motivates us to consider here new versions of proximal-like algorithms, especially adapted to the unifying framework (1.1). The proposed procedures are naturally inspired by the iterative scheme

$$x^{k+1} \in \operatorname{argmin}\{F(x^k, x) \mid x \in X\}, \tag{1.4}$$

known to be rather unstable. Therefore, to stabilize the iterations – or to regularize the underlying data – we shall invoke proximal-type penalty terms. Appropriate terms of this sort – prominent in many recent studies [9,10,12] – are so-called *Bregman distances* [7]. Whatever penalty term we use, a prime feature of proximal point methods, in their original form, is the ambition to predict *and* update in one single shot. Clearly, doing so tends to be difficult or costly. Therefore, in the main algorithm, we shall divorce these two aspects from each other. Indeed, a main novelty here, extending earlier ideas introduced by Antipin [1–3], is the double regularization undertaken at every stage: First, we predict the next iterate; thereafter, using the new prediction, we update the current point. In both operations, for the sake of practical applicability, we shall tolerate approximate evaluations of the objectives. The algorithms are stated in Section 3, and Section 4 contains the convergence analysis.

2. Examples

To motivate and justify the study of (1.1) we offer here a list of problems, all ubiquitous, and all complying with the form (1.1).

Convex minimization. Let $F(x,y)=f(y)$ with $f\colon X\to\mathbb{R}$ convex. Then x^* solves (1.1) iff $x^*\in\operatorname{argmin}\{f(x)\mid x\in X\}$. In this instance (1.2) holds by definition.

Convex-concave saddle problems. Let $X=X_1\times X_2$ be a product of two nonempty compact convex sets, both contained in Euclidean spaces. Posit $F(x,y)=L(y_1,x_2)-L(x_1,y_2)$ with $x=(x_1,x_2)$, $y=(y_1,y_2)$, and suppose the "Lagrangian" $L\colon X\to\mathbb{R}$ is convex-concave. Then x^* solves (1.1) iff x^* is a min-max saddle point of L. The monotonicity condition (1.2) is automatically satisfied in this case as well [27].

Noncooperative games in strategic form. Generalizing saddle problems, suppose each individual $i\in I$ (I finite), seeks, without cooperation, to minimize his private cost $F_i(x)=F_i(x_{-i},xi)$ with respect to own strategy x_i in a compact convex subset X_i of some real Euclidean space. Here $x_{-i}:=(x_\iota)_{\iota\in I-i}$ is short notation for actions taken by i's adversaries, and his cost $F_i(x_{-i},x_i)$ is convex in x_i. Let $F(x,y)=\sum_i F_i(x_{-i},y_i)$. Then $x^*=(x_i^*)\in X:=\prod X_i$ solves (1.1) iff x^* is a *Nash equilibrium*, that is, iff

$$x_i^* \text{ minimizes } F_i(x_{-i}^*,x_{-i}) \text{ s.t. } x_i\in X_i \quad \text{for all } i\in I.$$

This concept dominates in game theory [25], and provides a framework for exploring collective consequences of individual rationality. Clearly, it is quite demanding, requiring correct predictions about rival behavior, and optimal responses. So, rather than taking equilibrium for granted, a sound defence for the use of this concept must rest on arguments showing that such outcomes will eventually emerge under repeated play somehow. The many mechanisms at work then – including imitation, learning, evolution, and stepwise adjustments – are far from being understood, see [14,15,17,29]. It is clear, however, that any relevant process would hinge upon two key parts: the formation of beliefs, and the updating of strategies. As will be seen, one algorithm, proposed below, has the merit of treating these two things separately and explicitly.

Admittedly the monotonicity hypothesis (1.2) is fairly stringent in the context of games. Nonetheless, the class satisfying (1.2) is sufficiently large to merit special attention, and more rich than might first be imagined.

Definition 2.1 (Convex-concave games). A noncooperative strategic form game, as described above, is said to be *convex-concave* if its Ky Fan function

$$K(x,y) := F(x,x) - F(x,y) = \Sigma_i\{F_i(x) - F_i(x_{-i},y_i)\}$$

is convex-concave.

In such games equilibria may be found using saddle point algorithms, see [16,30]. What is important here is that these games tend to satisfy (1.2).

Theorem 2.2 (Convex-concave games are monotone). *Assume the game is convex-concave with $F(x,y)$ continuous in y. Then condition* (1.2) *holds.*

Proof. We claim that

$$-\partial K(x,x)/\partial y \subset \partial K(x,x)/\partial x \quad \text{for every } x \in X, \tag{2.1}$$

where the partial differentials are taken in the sense of convex analysis. To verify (2.1) pick any two points x^0, $x^1 \in X$, together with a number $\alpha \in]0,1[$, and let $x^\alpha := (1-\alpha)x^0 + \alpha x^1$. The convexity of K in its first argument yields

$$(1-\alpha)K(x^0,x^\alpha) + \alpha K(x^1,x^\alpha) \geq K(x^\alpha,x^\alpha) = 0.$$

Divide by α and let $\alpha \downarrow 0$ to obtain

$$\lim_{\alpha \downarrow 0} \alpha^{-1} K(x^0,x^\alpha) + K(x^1,x^0) \geq 0.$$

By concavity of K in the second argument, for every $g \in \partial K(x^0,x^0)/\partial y$ it holds that

$$\lim_{\alpha \downarrow 0} \alpha^{-1} K(x^0,x^\alpha) \leq \langle g, x^1 - x^0 \rangle.$$

Thus, combining the last two inequalities, $K(x^1,x^0) \geq \langle -g, x^1 - x^0 \rangle$, saying that $-g \in \partial K(x^0,x^0)/\partial x$. This verifies (2.1). Returning now to the main argument, suppose x^* is a Nash equilibrium. Since $K(x^*,x^*) = 0$, it follows that $K(x^*,x^*) = \sup_{y \in X} K(x^*,y) = 0$.

Consequently, there exists a partial supergradient $g^* \in \partial K(x^*,x^*)/\partial y$ such that $\langle g^*, x - x^* \rangle \leq 0$ for all $x \in X$. Now (2.1) implies $-g^* \in \partial K(x^*,x^*)/\partial x$. Finally, the desired condition (1.2) follows from the subgradient inequality

$$K(x,x^*) \geq K(x^*,x^*) + \langle -g^*, x - x^* \rangle = \langle -g^*, x - x^* \rangle \geq 0. \qquad \square$$

Proposition 2.3. *Suppose $F_i(x) = \Sigma_{j \neq i} \langle A_{ij} x_j, x_i \rangle + f_i(x_i) + h_i(x_i)$ with $f_i(x_i)$ twice differentiable. If the block matrix having $f_i''(x_i)$ in diagonal entry i and A_{ij} in entry (i,j) is positive semidefinite, then the game is convex-concave.*

Proof. Simply observe that $K(x,y) = \langle Ax,x \rangle + \Sigma_i\{f_i(x_i) - f_i(y_i)\} - \langle Ax,y \rangle$ where A is the block matrix (A_{ij}) with $A_{ii} = 0$ for all $i \in I$. Since this function K has a positive semidefinite Hessian in x, it must be convex with respect to that variable. Concavity in y is immediate. □

Example 2.4. An important instance of noncooperative games is the classical oligopoly model of Cournot [11]. That model remains a workhorse within modern theories of industrial organisation [32]. Generalizing it to comprise n different goods, it goes as follows: Firm $i \in I$ produces the commodity bundle $x_i \in \mathbb{R}^n$, thus incurring convex production cost $c_i(x_i)$ and gaining market revenues $\langle p(\Sigma_j x_j), x_i \rangle$. Here $p(\Sigma_j x_j)$ is the price vector at which total demand equals the aggregate supply $\Sigma_j x_j$. Suppose this inverse demand curve is affine and "slopes downwards" in the sense that $p(Q) = \alpha - SQ$ where $\alpha \in \mathbb{R}^n$ and S is an $n \times n$ positive semidefinite matrix. Then, letting $F_i = c_i(x_i) - \langle p(\Sigma_j x_j), x_i \rangle$, the resulting Cournot oligopoly is convex-concave. □

Proposition 2.3 and Example 2.4 point to important games having *bilinear interaction*. By this notion is meant that individual cost has the format mentioned in Proposition 2.3, with $f_i: \mathbb{R}^n i \to (-\infty, +\infty]$ proper convex lower semicontinuous, and h_i arbitrary.

Proposition 2.5. *Suppose the game has bilinear interaction with* $\Sigma_i\{\Sigma_{j \neq i}\langle A_{ij}x_j, x_i \rangle + f_i(x_i)\}$ *convex. Then the game itself is convex-concave. In particular, this happens when* $f_i(x_i) = \langle A_{ii}x_i, x_i \rangle /2 + \langle b_i, x_i \rangle$ *with* $A = (A_{ij})$ *positive semidefinite.* □

As is well known, noncooperative games – subsuming single-agent optimization and saddle problems as special cases – are related to:

Variational inequalities. Let $X \ni x \to M(x) \subset E$ be a correspondence with nonempty compact convex values. Then, letting $F(x,y) = \sup\{\langle m, y - x \rangle \mid m \in M(x)\}$, we get via von Neumann's minimax theorem that x^* solves (1.1) iff $\exists\, m(x^*) \in M(x^*)$ such that (1.3) holds. In this case (1.2) would follow if M were *quasi-monotone at equilibrium* x^* in the sense that $\inf\{\langle m, x - x^* \rangle \mid m \in M(x)\} \geq 0$ for all $x \in X$, see [18].

Successive approximations. Related to variational inequalities (1.3) is the following optimization procedure. Given $m: X \to \mathbb{R}^n$, a nonnegative number μ, and an $n \times n$ positive semidefinite matrix $H(x)$ for every $x \in X$. In this case, letting $F(x,y) = \langle m(x), y - x \rangle + \mu \langle y - x, H(x)(y - x) \rangle$, we have that x^* solves (1.1) iff $\langle m(x^*), x - x^* \rangle + \mu \langle x - x^*, H(x^*)(x - x^*) \rangle \geq 0$ for all $x \in X$ iff (1.3) holds. Indeed, if $\langle m(x^*), x^0 - x^* \rangle$ for some $x^0 \in X$, consider the point $x^\alpha := (1 - \alpha)x^0 + \alpha x^*$ with $\alpha \in [0,1[$, thus obtaining

$$\begin{aligned} &\langle m(x^*), x^\alpha - x^* \rangle + \mu \langle x^\alpha - x^*, H(x^*)(x^\alpha - x^*) \rangle \\ &\quad = (1 - \alpha)\{\langle m(x^*), x^0 - x^* \rangle + (1 - \alpha)\mu \langle x^0 - x^*, H(x^*)(x^0 - x^*) \rangle\} \end{aligned}$$

for α sufficiently close to 1. We note that (1.2) is satisfied iff

$$\langle m(x), x-x^* \rangle \geq \mu \langle x-x^*, H(x)(x-x^*) \rangle \text{ for all } x \in X. \tag{2.2}$$

Suppose therefore, that m is differentiable and $m'(\xi) - \mu H(x)$ is positive semidefinite for all $\xi \in [x, x^*]$. Then (1.2) holds because, using the mean-value theorem, (2.2) follows from

$$\begin{aligned}
&\langle m(x), x-x^* \rangle \\
&\quad = \langle m(x^*), x-x^* \rangle + \langle m(x) - m(x^*), x-x^* \rangle \\
&\quad \geq \langle m(x) - m(x^*), x-x^* \rangle = \langle x-x^*, m'(\xi)(x-x^*) \rangle \\
&\qquad \text{for some } \xi \in [x, x^*] \\
&\quad = \langle x-x^*, [m'(\xi) - \mu H(x) + \mu H(x)](x-x^*) \rangle \\
&\qquad \geq \langle x-x^*, \mu H(x)(x-x^*) \rangle.
\end{aligned}$$

3. The algorithm

This section advocates two procedures to solve (1.1). Both amend (1.4). Our motivation is that (1.4) – derived directly from (1.1) – has three potential drawbacks, all begging to be rectified.

First, it is unreasonable, and not very practical, to insist that argmin in (1.4) be located exactly at every stage k. Rather one should tolerate small errors $\varepsilon_k \geq 0$ in extremal values – at least during early phases of the computation.

Second, there is some myopia in (1.4): Specifically, at stage k, instead of considering $F(x^k, x)$, one might form a prediction x^{k+} of the upcoming point x^{k+1}, and rather minimize $F(x^{k+}, x)$.

Third, but no less important, the argmin operation – whether executed exactly or not – may cause instabilities. To mitigate this we add a nonnegative penalty for x^{k+1} deviating from the x^k.

These considerations lead us to replace (1.4) by a more stable and flexible algorithm:

Algorithm. Start at arbitrary $x^0 \in X$, and **update** iteratively, for $k = 0, 1, \ldots$ until convergence, by the rule

$$x^{k+1} \in \varepsilon_k - \operatorname{argmin}\{\alpha_k F_k(x^{k+}, x) + D(x, x^k) \mid x \in X\}. \tag{3.1}$$

Several remarks are needed to make (3.1) well defined and comprehensible:

As said, $\varepsilon_k \geq 0$ is an error tolerated when evaluating minimal values at iteration k. For asymptotic accuracy we require that

$$\textstyle\sum_k \epsilon_k^{1/2} < +\infty. \tag{3.2}$$

Such errors ε_k not withstanding, the constraint $x \in X$ is always enforced, and tacitly assumed to be easy.

· The other real parameter α_k in (3.1) is positive and bounded away from 0 and $+\infty$. It permits judicious weighing of the objective against the penalty term. Both sequences $\{\varepsilon_k\}$, $\{\alpha_k\}$ can be selected rather freely. Indeed, all hypotheses concerning these sequences are purely technical, and satisfied by construction.

· The intermediate and "predictive" point x^{k+} figuring in (3.1) is determined in two possible ways, depending on the quality of predictions. One, utterly demanding scenario, is naturally named *perfect foresight*, meaning that

$$x^{k+} := x^{k+1}. \tag{3.3}$$

Another, more reasonable and less taxing arrangement, amounts to tolerate *imperfect foresight*. Specifically, we select then

$$x^{k+} \in \varepsilon_k - \operatorname{argmin}\left\{\alpha_k F_k(x^k, x) + D(x, x^k) \mid x \in X\right\}. \tag{3.4}$$

· The penalty function D in (3.1), (3.4) is a so-called *Bregman distance* [7] given by

$$D(x,y) := \psi(x) - \psi(y) - \langle \psi'(y), x - y \rangle \tag{3.5}$$

where ψ is a differentiable strictly convex function defined on a neighborhood of X. We shall require that ψ has a Lipschitz continuous gradient. Specifically, there must exist a positive constant L such that for every error $\varepsilon \geq 0$ considered in the sequel we have

$$x \in X,\ dist(X,y) \leq \varepsilon^{1/2} \quad \Rightarrow \quad \|\psi'(x) - \psi' y)\| \leq L\|x - y\|. \tag{3.6}$$

Since X is compact, (3.6) holds whenever ψ is twice continuously differentiable.

Note that $D(x,y)$, as defined in (3.5), is non-negative, and strictly convex differentiable in its first argument whenever x is sufficiently close to X and $y \in X$. Moreover, $D(x,y) = 0 \Leftrightarrow x = y$. We assume that for any sequence $\{x^k\}$ in X it holds

$$x^k \to x \quad \Leftrightarrow \quad D(x, x^k) \to 0 \tag{3.7}$$

as well as

$$x^k \to x,\ \xi^k \in X,\ D(\xi^k, x^k) \to 0 \quad \Rightarrow \quad \xi^k \to x. \tag{3.8}$$

Evidently, the customary and convenient choice $\psi = \|\cdot\|^2/2$ yields $L = 1$, and $D(x,y) = \|x - y\|^2/2$, satisfying (3.7) and (3.8). Other examples of Bregman distances and similar penalty terms are found in [5,9,10,12,13,21,22,31]. Many of these papers explore algorithms of the sort (3.1) and (3.3), or akin to this procedure.

4. Convergence

For the convergence analysis of (3.1) we shall need three lemmata.

Lemma 4.1. *Suppose the function $h: E \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous proper convex. Let X be a nonempty closed convex subset of E such that* ri dom $h \cap$ ri $X \neq \emptyset$, *and select an arbitrary fixed vector $\xi \in X$. Suppose the function ψ in (3.5) is* *real-valued convex on an open set containing X, with Lipschitz continuous gradient. Specifically, for a given number $\varepsilon \geq 0$, we suppose that* (3.6) *holds. Then, with*

$$x^+ \in \varepsilon - \operatorname{argmin}\{h(x) + D(x,\xi) \mid x \in X\}, \tag{4.1}$$

we have for some $\delta \in [0,\varepsilon]$ and all $x \in X$,

$$\begin{aligned} &h(x) + D(x,\xi) \\ &\quad \geq h(x^+) + D(x^+,\xi) + D(x,x^+) - \delta - (L+1)(\varepsilon-\delta)^{1/2}\|x - x^+\|. \end{aligned} \tag{4.2}$$

Proof. The ε-optimality (4.1) of x^+ implies that

$$0 \in \varepsilon - \partial\big[h + D(\cdot,\xi) + I_X\big](x^+), \tag{4.3}$$

where $\varepsilon - \partial$ denotes the ε-subdifferential operator, and I_X is the convex indicator of X, that is, $I_X(x) = 0$ when $x \in X$, $+\infty$ otherwise. Applying the ε-subdifferential calculus, excellently exposed in [20], to (4.3), there exist, by the domain qualification ri dom $h \cap$ ri $X \neq \emptyset$, approximate subgradients

$$s_1 \in \varepsilon_1 - \partial h(x^+), \quad s_2 \in \varepsilon_2 - \partial D(\cdot,\xi)(x^+), \quad s_3 \in \varepsilon_3 - \partial I_X(x^+),$$

with

$$\varepsilon_1,\varepsilon_2,\varepsilon_3 \geq 0, \quad \varepsilon_1 + \varepsilon_2 + \varepsilon_3 = \varepsilon, \quad \text{and} \quad s_1 + s_2 + s_3 = 0. \tag{4.4}$$

In turn,

$$s_1 \in \varepsilon_1 - \partial h(x^+) \quad \Rightarrow \quad h(x) \geq h(x^+) + \langle s_1, x - x^+\rangle - \varepsilon_1 \quad \text{for all } x; \tag{4.5}$$

$$s_2 \in \varepsilon_2 - \partial D(\cdot,\xi)(x^+) \quad \Rightarrow \quad s_2 = S_2 - \psi'(\xi) \quad \text{for some } S_2 \in \varepsilon_2 - \partial\psi(x^+); \tag{4.6}$$

$$s_3 \in \varepsilon_3 - \partial I_X(x^+) \quad \Rightarrow \quad \langle s_3, x - x^+\rangle \leq \varepsilon_3 \quad \text{for all } x \in X. \tag{4.7}$$

(3.5) yields the *three-point identity* [10]

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

Adding the latter to inequality (4.5) we arrive at

$$\begin{aligned} h(x) + D(x,\xi) \geq{}& h(x^+) + D(x^+,\xi) + D(x,x^+) \\ &+ \langle s_1 + \psi'(x^+) - \psi'(\xi), x - x^+\rangle - \varepsilon_1, \end{aligned} \tag{4.8}$$

valid for all $x \in X$. Considering now the inclusion $S_2 \in \varepsilon_2 - \partial\psi(x^+)$ figuring in (4.6), a theorem of Brønsted and Rockafellar [8] (see alternatively Theorem XI.4.2.1 in [20] ensures the existence of a vector y satisfying

$$\|x^+ - y\| \leq \varepsilon_2^{1/2} \quad \text{and} \quad \|\psi'(y) - S_2\| \leq \varepsilon_2^{1/2}. \tag{4.9}$$

We shall use these facts to underestimate the next to last term of (4.8) – that is, the inner product mentioned there – as follows: First observe via (4.4) and (4.6) that

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

Therefore, invoking (4.9) and (4.7) in that order:

$$\begin{aligned}&\langle s_1 + \psi'(x^+) - \psi'(\xi), x - x^+\rangle\\&= \langle \psi'(x^+) - \psi'(y) + \psi'(y) - S_2 - s_3, x - x^+\rangle\\&\geq -\{\|\psi'(x^+) - \psi'(y)\| + \|\psi'(y) - S_2\|\}\|x - x^+\| - \langle s_3, x - x^+\rangle\\&\geq -\{L\varepsilon_2^{1/2} + \varepsilon_2^{1/2}\}\|x - x^+\| - \varepsilon_3.\end{aligned}$$

Finally, returning to (4.8), the desired inequality (4.2) follows with $\delta = \varepsilon_1 + \varepsilon_3$ and $\varepsilon_2 = \varepsilon - \delta$. □

Lemma 4.2. *Suppose the function $h: E \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous proper convex. Let X be a nonempty closed convex subset of E such that* $\text{ri}\,\text{dom}\, h \cap \text{ri}\, X \neq \emptyset$. *Suppose the function ψ in* (3.5) *is real-valued convex on an open set containing X, with Lipschitz continuous gradient. Then,*

$$x^* \in \text{argmin}\{h(x) + D(x, x^*) \mid x \in X\} \quad \Leftrightarrow \quad x^* \in \text{argmin}\{h(x) \mid x \in X\}.$$

Proof. For the implication "⇒" use Lemma 4.1 with $\varepsilon = 0$ and $\xi = x^* = x^+$ to have $h(x) + D(x,x^*) \geq h(x^*) + D(x^*,x^*) + D(x,x^*)$, whence $h(x) \geq h(x^*)$ for all $x \in X$.

Conversely, when $h(x) \geq h(x^*)$ for all $x \in X$, (3.5) and the convexity of ψ yields $D(x,x^*) \geq D(x^*,x^*) = 0$ so that $h(x) + D(x,x^*) \geq h(x^*) + D(x^*,x^*)$ for all $x \in X$. □

The final lemma is easily proven, or derived from more general result of Robbins and Siegmund [26].

Lemma 4.3. *Let $\{a_k\}$, $\{b_k\}$, $\{c_k\}$ be three sequences of nonnegative numbers such that $\sum_k b_k < +\infty$ and $a_{k+1} \leq a_k + b_k - c_k$ for all k. Then $\{a_k\}$converges, and $\sum_k c_k < +\infty$.* □

After these preparations we are ready to state a first convergence result. For this assume henceforth that *the hypotheses of Proposition* 1.1 *and condition* (1.2) *are all satisfied.*

Theorem 4.4 (Convergence under perfect foresight). *Any sequence $\{x^k\}$ generated according to algorithm* (3.1), (3.3) *converges to a solution of* (1.1).

Proof. The assumptions of Lemma 4.1 hold for any function $h(x) = \alpha_k F(x^{k+1}, x)$. Therefore, invoking (4.2) in situation (3.1), letting $x = x^*$ be any equilibrium, $x^{k+} = x^{k+1}$, and $\xi = x^k$, we get

$$\begin{aligned}&\alpha_k F(x^{k+1}, x^*) + D(x^*, x^k)\\&\geq \alpha_k F(x^{k+1}, x^{k+1}) + D(x^{k+1}, x^k) + D(x^*, x^{k+1}) - \delta_k\\&\qquad - (L+1)(\varepsilon_k - \delta_k)^{1/2}\|x^* - x^{k+1}\|\end{aligned}$$

for some $\delta_k \in [0, \varepsilon_k]$. Appealing to the monotonicity condition $F(x^{k+1}, x^*) \le F(x^{k+1}, x^{k+1})$ it follows that

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

Use now the boundedness of X, condition (3.2), and Lemma 4.3 with

$$a_k := D(x^*, x^k), \quad b_k := \delta_k + (L+1)(\varepsilon_k - \delta_k)^{1/2} \|x^* - x^{k+1}\|,$$
$$c_k := D(x^{k+1}, x^k),$$

to assert that $D(x^*, x^k)$ converges, and $\sum D(x^{k+1}, x^k) < +\infty$.

Let x be any accumulation point of $\{x^k\}$. Then, for some subsequence K of integers we have $\lim_{k \in K} \alpha_k = \alpha > 0$, and $\lim_{k \in K} x^k = x$. Since $D(x^{k+1}, x^k) \to 0$, we also get, via (3.8), that $\lim_{k \in K} x^{k+1} = x$. Thus, passing to the limit along K in (3.1) we obtain

$$x \in \operatorname{argmin}\{\alpha F(x, y) + D(y, x) \mid y \in X\}. \tag{4.10}$$

By Lemma 4.2 this implies $x \in \operatorname{argmin}\{F(x, y) \mid y \in X\}$, whence x solves (1.1).

The upshot is that $\{x^k\}$ clusters to a solution $x^* := x$ of (1.1). Knowing already that $D(x^*, x^k)$ converges, we obtain, via (3.7), that $D(x^*, x^k) \to 0$. Consequently, again relying on (3.7), the entire sequence $\{x^k\}$ converges to x^*. □

If $F(x, y) = f(y)$ with $f: X \to \mathbb{R}$ convex, and all $\varepsilon_k = 0$, then the implicit algorithm (3.1) and (3.3) – first developed by Martinet [24] – requires few assumptions, but is, in general, rather hard to implement. By contrast, (3.1) and (3.4) being far more tractable, needs a

Hypothesis on smoothness. There exists a constant $\Lambda > 0$ such that on X it obtains

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

This rather strange smoothness 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 F is twice continuously differentiable. We can now state the main result:

Theorem 4.5 (Convergence under imperfect foresight). *Suppose that* $0 < \inf \alpha_k \le \sup \alpha_k < 1/\Lambda$ *where* Λ *satisfies the smoothness condition. Then any sequence* $\{x^k\}$ *generated according to algorithm* (3.1), (3.4) *converges to a solution of* (1.1).

Proof. The assumptions of Lemma 4.1 hold when applied to $h(x)=\alpha_k F(x^k,x)$. Therefore, invoking (4.2) in situation (3.4) we get

$$\begin{aligned}&\alpha_k F(x^k,x^{k+1})+D(x^{k+1},x^k)\\&\quad\geq \alpha_k F(x^k,x^{k+})+D(x^{k+},x^k)+D(x^{k+1},x^{k+})-\delta_{k+}\\&\qquad-(L+1)(\varepsilon_k-\delta_{k+})^{1/2}\|x^{k+1}-x^{k+}\|\end{aligned}$$

for some $\delta_{k+}\in[0,\varepsilon_k]$. Likewise, using now $h(x)=\alpha_k F(x^{k+},x)$, situation (3.1) implies, again via (4.2), that for any solution x^* to (1.1) we have

$$\begin{aligned}&\alpha_k F(x^{k+},x^*)+D(x^*,x^k)\\&\quad\geq \alpha_k F(x^{k+},x^{k+1})+D(x^{k+1},x^k)+D(x^*,x^{k+1})-\delta_k\\&\qquad-(L+1)(\varepsilon_k-\delta_k)^{1/2}\|x^*-x^{k+1}\|\end{aligned}$$

for some $\delta_k\in[0,\varepsilon_k]$. Adding the last two inequalities we have

$$\begin{aligned}&\alpha_k\{F(x^k,x^{k+1})-F(x^{k+},x^{k+1})-F(x^k,x^{k+})+F(x^{k+},x^*)\}+D(x^*y,x^k)\\&\quad\geq D(x^*,x^{k+1})+D(x^{k+},x^k)+D(x^{k+1},x^{k+})-\delta_{k+}-\delta_k\\&\qquad-(L+1)\{(\varepsilon_k-\delta_{k+})^{1/2}\|x^{k+1}-x^{k+}\|+(\varepsilon_k-\delta_k)^{1/2}\|x^*-x^{k+1}\|\}.\end{aligned}\tag{4.11}$$

Now invoke the smoothness condition and (1.2), in that order, to get

$$\begin{aligned}&2\Lambda\alpha_k\{D(x^{k+},x^k)D(x^{k+1},x^{k+})\}^{1/2}\\&\quad\geq \alpha_k\{F(x^k,x^{k+1})-F(x^{k+},x^{k+1})-F(x^k,x^{k+})+F(x^{k+},x^{k+})\}\\&\quad\geq \alpha_k\{F(x^k,x^{k+1})-F(x^{k+},x^{k+1})-F(x^k,x^{k+})+F(x^{k+},x^*)\}.\end{aligned}$$

Combining this last string of inequalities with (4.11) it follows that

$$\begin{aligned}&2\Lambda\alpha_k\{D(x^{k+},x^k)D(x^{k+1},x^{k+})\}^{1/2}+D(x^*,x^k)\\&\quad\geq D(x^*,x^{k+1})+D(x^{k+},x^k)+D(x^{k+1},x^{k+})-\delta_{k+}-\delta_k\\&\qquad-(L+1)\{(\varepsilon_k-\delta_{k+})^{1/2}\|x^{k+1}-x^{k+}\|+(\varepsilon_k-\delta_k)^{1/2}\|x^*-x^{k+1}\|\},\end{aligned}$$

whence

$$\begin{aligned}D(x^*,x^k)\geq{}& D(x^*,x^{k+1})+\{D(x^{k+},x^k)^{1/2}-\Lambda\alpha_k D(x^{k+1},x^{k+})^{1/2}\}^2\\&+\{1-(\Lambda\alpha_k)^2\}D(x^{k+1},x^{k+})-\delta_{k+}-\delta_k-(L+1)\\&\times\{(\varepsilon_k-\delta_{k+})^{1/2}\|x^{k+1}-x^{k+}\|+(\varepsilon_k-\delta_k)^{1/2}\|x^*-x^{k+1}\|\}.\end{aligned}$$

We rely now on the boundedness of X, condition (3.2), and Lemma 4.3, letting

$$\begin{aligned}a_k&:=D(x^*,x^k),\\b_k&:=\delta_{k+}+\delta_k+(L+1)\{(\varepsilon_k-\delta_{k+})^{1/2}\|x^{k+1}-x^{k+}\|\\&\qquad+(\varepsilon_k-\delta_k)^{1/2}\|x^*-x^{k+1}\|\},\end{aligned}$$

and

$$c_k := \left\{D(x^{k+},x^k)^{1/2} - \Lambda\alpha_k D(x^{k+1},x^{k+})^{1/2}\right\}^2 + 1 - (\Lambda\alpha_k)^2 D(x^{k+1},x^{k+}),$$

to assert, since $\Lambda\alpha_k < 1$ for all large k, that $D(x^*,x^k)$ converges, and

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

In particular, $D(x^{k+1},x^{k+}) \to 0$ and $D(x^{k+},x^k) \to 0$. Let x be any accumulation point of $\{x^k\}$. Then, for some integer subsequence K we have $\lim_{k\in K}\alpha_k = \alpha > 0$, and via (3.8),

$$\lim_{k\in K} x^k = \lim_{k\in K} x^{k+} = \lim_{k\in K} x^{k+1} = x.$$

Passing to the limit along K in (3.1) we arrive at (4.10). From there onwards the same arguments furnishes the desired conclusion. □

Clearly, in (3.4) one might use a another sequence $\{\varepsilon_{k+}\}$ of nonnegative errors, possibly different from $\{\varepsilon_k\}$, but also satisfying (3.2).

When $f: X \to \mathbb{R}$ is convex differentiable, $\varepsilon_k = 0$, and $F(x,y) = \langle f'(x), y - x\rangle$, step (3.1) assumes the form $\langle f'(x^{k+}), x - x^{k+1}\rangle \geq 0$ for all $x \in X$, reminiscent of the extra-gradient method of Korpelevich [23].

References

[1] A.S. Antipin, Controlled proximal differential systems for saddle problems, *Differential Equations* 28 (1992) 1498–1510.

[2] A.S. Antipin, Feed-back controlled saddle gradient processes, *Automat. Remote Control* 55 (1994) 311–320.

[3] A.S. Antipin, Convergence and estimates of the rate of convergence of proximal methods to fixed points of extremal mappings, *Zh. Vychisl. Mat. Fiz.* 35 (1995) 688–704.

[4] A. Auslender, J.P. Crouzeix and P. Fedit, Penalty-proximal methods in convex programming, *Journal of Optimization Theory and Applications* 55 (1987) 1–21.

[5] A. Auslender and M. Haddou, An interior-proximal method for convex linearly constrained problems and its extension to variational inequalities, *Mathematical Programming* 71 (1995) 77–100.

[6] D.P. Bertsekas and P. Tseng, Partial proximal minimization algorithms for convex programming, *SIAM Journal of Optimization* 4 (1994) 551–572.

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

[8] A. Brønsted and R.T. Rockafellar, On the subdifferentiability of convex functions, *Proceedings of the American Mathematical Society* 16 (1965) 605–611.

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

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

[11] A. Cournot, *Recherches sur les Principes Mathématiques de la Théorie des Richesses* (Paris, 1838); English translation: *Researches into the Mathematical Principles of the Theory of Wealth* (N. Bacon, ed.) (Macmillan, New York, 1897).

[12] J. Eckstein, Nonlinear proximal point algorithms using Bregman functions with applications to convex programming, *Mathematics of Operations Research* 18 (1) (1993) 202–226.

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

[14] Yu.M. Ermol'ev and S.P. Uryas'ev, Nash equilibrium in n-person games, *Kibernetika* 3 (1982) 85–88 (in Russian).

[15] S.D. Flåm, Paths to constrained Nash equilibria, *Applied Mathematics and Optimization* 27 (1993) 275–289.

[16] S.D. Flåm and A. Ruszczynski, Noncooperative games: Computing equilibrium by partial regularization, Working Paper IIASA 42, 1994.

[17] S.D. Flåm, Approaches to economic equilibrium, *Journal of Economic Dynamics and Control* 20 (1996) 1505–1522.

[18] N. Hadjisavvas and S. Schaible, On strong monotonicity and (semi)strict quasimonotonicty, *Journal of Optimization Theory and Applications* 79 (1993) 139–155.

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

[20] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms* (Springer, Berlin, 1993).

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

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

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

[24] B. Martinet, Perturbations des methodes d'optimisation, *RAIRO Anal. Numer.* 12 (1978) 153–171.

[25] M.J. Osborne and A. Rubinstein, *A Course in Game Theory* (MIT Press, Cambridge, MA, 1994).

[26] H. Robbins and D. Siegmund, A convergence theorem for nonnegative almost surmartingales and some applications, in: J. Rustagi, ed., *Optimizing Methods in Statistics* (Academic Press, New York, 1971) 235–257.

[27] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

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

[29] J.B. Rosen, Existence and uniqueness of equilibrium points for concave N-person games, *Econometrica* 33 (1965) 520–534.

[30] A. Ruszczynski, A partial regularization method for saddle point seeking, Working Paper IIASA 20, 1994.

[31] M. Teboulle, Entropic proximal mappings with applications to nonlinear programming, *Mathematics of Operations Research* 17 (1992) 670–690.

[32] J. Tirole, *The Theory of Industrial Organization* (MIT Press, Cambridge, MA, 1988).