Computational Optimization and Applications 8, 21–39 (1997)
© 1997 Kluwer Academic Publishers. Manufactured in The Netherlands.

Iterative Averaging of Entropic Projections for Solving Stochastic Convex Feasibility Problems

DAN BUTNARIU
Department of Mathematics and Computer Science, University of Haifa, 31905 Haifa, Israel

YAIR CENSOR
Department of Mathematics and Computer Science, University of Haifa, 31905 Haifa, Israel

SIMEON REICH
Department of Mathematics, The Technion—Israel Institute of Technology, 32000 Haifa, Israel

Received June 7, 1995; Revised December 7, 1995; Accepted January 2, 1996

Abstract. The problem considered in this paper is that of finding a point which is common to almost all the members of a measurable family of closed convex subsets of $\mathbf{R}^n_{++}$, provided that such a point exists. The main results show that this problem can be solved by an iterative method essentially based on averaging at each step the Bregman projections with respect to $f(x) = \sum_{i=1}^n x_i \cdot \ln x_i$ of the current iterate onto the given sets.

Keywords: stochastic convex feasibility problem, Bregman projection, entropic projection, modulus of local convexity, very convex function

1. Introduction

1.1. In this paper we approach a particular version of the *stochastic convex feasibility problem* (*SCFP*) formulated by Butnariu and Flåm [8]: *Given a complete probability space* $(\Omega, \mathcal{A}, \mu)$ *and a measurable point-to-set mapping* $Q : \Omega \to \mathbf{R}^n_{++}$ *having nonempty, closed and convex values, find an almost common point* x^* *of the sets* Q_ω, *i.e., find a point* $x^* \in \mathbf{R}^n_{++}$ *such that*

$$\mu(\{\omega \in \Omega \mid x^* \in Q_\omega\}) = 1, \tag{1}$$

provided that such a point exists. Recall that the point-to-set mapping Q is called *measurable* if, for every closed subset $F \subseteq \mathbf{R}^n$, the set $\{\omega \in \Omega \mid Q_\omega \cap F \neq \emptyset\}$ is measurable (in the sense that it belongs to $\mathcal{A}$). Note that if Q is measurable, then, for each $x \in \mathbf{R}^n$, the set $\{\omega \in \Omega \mid x \in Q_\omega\}$ is measurable and, therefore, formula (1) makes sense.

The SCFP stated above is a natural generalization of the well-known *consistent convex feasibility problem* (*CFP*). A CFP is a particular SCFP in which Ω is a finite set, $\mathcal{A} = 2^\Omega$ is the collection of all subsets of Ω and the measure $\mu : \mathcal{A} \to [0, 1]$ is defined by

$$\mu(A) = \sum_{\omega \in A} \mu_\omega, \tag{2}$$

for some given positive real numbers $\mu_\omega, \omega \in \Omega$, such that $\sum_{\omega\in\Omega} \mu_\omega = 1$. In this particular case, the set of almost common points of the sets Q_ω is exactly $\bigcap_{\omega\in\Omega} Q_\omega$, that is, the problem of finding an almost common point of the given sets is exactly the problem of determining an element of their intersection.

1.2. We study the SCFP, formulated above, under the assumption that

$$S := cl\,\text{conv}\left(\bigcup_{\omega\in\Omega} Q_\omega\right) \subseteq \mathbf{R}^n_{++}. \tag{3}$$

The question we pose is whether, and under what additional conditions on the data of the SCFP, sequences generated in $\mathbf{R}^n_{++}$ according to the following algorithmic scheme, which we call the *averaged entropic projection method* (*AEPM*), are well-defined and converge to almost common points of the sets $Q_\omega, \omega \in \Omega$: Choose an arbitrary *initial point* $x^0 \in \mathbf{R}^n_{++}$ and, for each integer $k \geq 0$, let[1]

$$x^{k+1} = \int_\Omega \Pi_\omega(x^k) d\mu(\omega), \tag{4}$$

where $\Pi_\omega(x) := \Pi_{Q_\omega}(x)$ denotes the Bregman projection of x onto the set Q_ω with respect to the function $f : \mathbf{R}^n_{++} \to \mathbf{R}$ given by

$$f(x) = \sum_{i=1}^n x_i \cdot \ln x_i. \tag{5}$$

Recall that, according to Censor and Lent [9] , the *Bregman projection with respect to f* of a point x onto the closed convex set $K \subseteq \mathbf{R}^n_{++}$ is the (necessarily unique) minimizer over K, denoted $\Pi_K(x)$, of the functional $D_f(\cdot, x) : \mathbf{R}^n_{++} \to \mathbf{R}_+$ defined by

$$D_f(y, x) = f(y) - f(x) - \langle \nabla f(x), y - x\rangle, \tag{6}$$

i.e.,

$$\Pi_K(x) = \arg\min\{D_f(y, x) \mid y \in K\}. \tag{7}$$

Lemma 2.2 in [9] ensures that $\Pi_K(x)$ exists. Since the function f defined by (5) is the negative of Shannon's *entropy function*, we call it *negentropy* and the Bregman projections with respect to it *entropic projections*.

According to Theorem 8.2.11 of Aubin and Frankowska [2], for each $x \in \mathbf{R}^n_{++}$, the function $\omega \to \Pi_\omega(x) : \Omega \to \mathbf{R}^n_{++}$ is measurable, and thus the average

$$\Pi(x) := \int_\Omega \Pi_\omega(x) d\mu(\omega) \tag{8}$$

exists and belongs to $[0, \infty]^n$. Obviously, if Ω is finite, $\mathcal{A} = 2^\Omega$ and μ is defined by (2), then

$$\Pi(x) = \sum_{\omega\in\Omega} \mu_\omega \cdot \Pi_\omega(x) \in S. \tag{9}$$

We prove in Section 2 that, if (3) holds, then $\Pi(x)$ exists and belongs to S even if Ω is infinite. This implies that the sequence $\{x^k\}_{k\in\mathbf{N}}$, recursively defined by (4), exists and is included in $\mathbf{R}^n_{++}$ no matter how the initial point x^0 is chosen in $\mathbf{R}^n_{++}$.

1.3. In this work we show that the answer to the question posed above is affirmative, i.e., sequences $\{x^k\}_{k\in\mathbf{N}}$ generated by the AEPM converge to almost common points of the sets Q_ω. The convergence result is presented in Section 3. It gives an algorithm to solve SCFPs which, in its specific framework, can be used as an alternative to the stochastic projection method studied by Butnariu and Flåm [8]. Applied to CFPs, the AEPM represents a new addition to a long list of existing techniques for finding common points of finite collections of closed convex subsets of $\mathbf{R}^n$ which have nonempty intersection. The classical methods of finding solutions of CFPs, based on computing metric projections onto the closed convex sets Q_ω, have recently been surveyed by Bauschke and Borwein [3] and Combettes [13]. The AEPM we propose falls in the category of iterative algorithms introduced by Bregman [5] and further developed by Censor and Lent [9] and others. They differ from the classical methods by the essential feature that, at each iterative step, the new iterate is determined by combining not metric projections but Bregman projections of the current iterate onto the sets Q_ω. The known algorithms in this category are either "sequential" in the sense that each new iterate is computed by using the Bregman projection of the previous iterate on a single set Q_ω (see [1, 4, 5, 9–11, 14]) or "simultaneous" in the sense that the computation of the new iterate requires determining the Bregman projection of the current iterate on all sets Q_ω (see [10, 11]). The AEPM, seen as a method for solving CFPs, is simultaneous. In other simultaneous entropic projection based algorithms (see [10, 11]) x^{k+1} is defined as the *geometric mean* of the individual entropic projections $\Pi_\omega(x^k)$. In the AEPM, x^{k+1} is determined as the *weighted average* (see (4) and (9)) of the individual entropic projections $\Pi_\omega(x^k)$, $\omega \in \Omega$. This fact represents an advantage of the AEPM because computation of weighted averages is more efficient and less error prone than computation of geometric means.

1.4. The AEPM, whenever applicable, is an alternative to the expected (metric) projection method of solving SCFPs presented in [8]. It was observed in [8] and [6, Section 3] that SCFPs for finding numerical solutions of Fredholm type integral equations or of best approximation problems in L^∞ as well as minimization problems for differentiable convex functionals over compact subsets of $\mathbf{R}^n$ involve sets Q_ω which are hyperplanes or half spaces. Such SCFPs can be efficiently solved by the expected projection method because the metric projections onto hyperplanes and half spaces are easily computable using explicit formulae (see, for instance, [1, Section 5]). However, if the geometry of the sets Q_ω is more complicated (i.e., nonpolyhedral), then computing metric projections onto them is rather difficult. In specific situations, determining entropic projections may be easier and, then, application of the AEPM proposed here provides an alternative device. An example illustrating this fact is given in Section 4. In general, computing the entropic projections required by the AEPM can be demanding. This leads to the question whether, and under what conditions, one can guarantee convergence of algorithms for solving SCFPs similar to the AEPM in which the negentropy f is replaced by a function other than the square of the norm. Replacing the negentropy in the AEPM by the square of the norm one obtains the expected metric projection method mentioned above. Having a large pool of such functions would allow users to fit the method of solving particular SCFPs to the nature of the sets Q_ω involved in such a manner that the effort of computing the corresponding Bregman projections is reduced (see Section 4).

1.5. The question asked above points to another interesting problem: Can the proofs of the main results in this work (namely, Proposition 2.5(ii), Proposition 3.2 and Theorem 3.3) be extended to functions other than the negentropy and the square of the norm? In this respect, a careful analysis of our arguments reveals that these results are based on two features of the negentropy which are common to a large class of convex functions: The property of being very convex and the convexity of D_f on $\text{int}(\text{dom}(f)) \times \text{int}(\text{dom}(f))$. The notion of a very convex function is a generalization of the notion of a uniformly convex function discussed in [19]. Very convex functions are defined in Section 2 via the newly introduced concept of modulus of local convexity (see Definition 2.2) whose basic properties are emphasized in general terms by Proposition 2.4. The proof of Proposition 2.5(ii) can be adapted to show that, if f is any very convex function with solid closed domain and if the function f is differentiable on $\text{int}(\text{dom}(f))$, then the integral $\Pi(x)$ in which entropic projections are replaced by the general Bregman projections with respect to f exists and belongs to $\text{dom}(f)$ for each $x \in \text{int}(\text{dom}(f))$. Thus, the existence of the sequences $\{x^k\}_{k\in\mathbf{N}}$ defined by (4) can be ensured in a more general framework provided that $S \subset \text{int}(\text{dom}(f))$. The convergence analysis of $\{x^k\}_{k\in\mathbf{N}}$ (i.e., the proof of Theorem 3.2) follows the pattern of proof devised by Iusem and De Pierro [16] and Butnariu and Censor [7]. However, in the convergence proof we use particular properties of the negentropy, namely the convexity of $D_f(\cdot,\cdot)$ on $\mathbf{R}^n_{++} \times \mathbf{R}^n_{++}$ and the relationship between the modulus of local convexity of the negentropy and the modulus of local convexity of the function $g(t) = t \cdot \ln t$ (see (14)). The properties of the modulus of local convexity of the negentropy used in the proof of Theorem 3.2 are shared with all uniformly convex functions (in spite of the fact that the negentropy f is not uniformly convex). Thus, we conjecture that, by following the basic ideas of our analysis, new algorithms for solving SCFPs can be developed.

1.6. One can approach SCFPs in $\mathbf{R}^n$ with other techniques in addition to the expected projection method and the averaged entropic projection method discussed above. One such technique, pointed out to us by a referee, consists of reformulating the SCFP as a stochastic programming problem in the sense of Wets [20]. More precisely, the SCFP can be seen as a minimization problem for the average of the indicator functions I_{Q_ω} to which the Progressive Hedging Algorithm of Rockafellar and Wets [18] can be applied. In practice, solving the SCFP in this way amounts to approximating solutions of a sequence of nonlinear programming problems. The relative advantage of the methods of solving SCFPs via iterative averaging of Bregman projections onto the sets Q_ω is that, in specific circumstances as those discussed in Section 1.4, they allow us to avoid computationally costly nonlinear optimization procedures.

2. Modulus of local convexity: existence of $\Pi(x)$

2.1. In this section we show that the averaged entropic projection $\Pi(x)$ defined by (8) exists and belongs to S for any $x \in \mathbf{R}^n_{++}$. As noted in Section 1.2, the measurability of the mapping $\omega \to \Pi_\omega(x) : \Omega \to \mathbf{R}^n_{++}$ is guaranteed. Thus it remains to prove that this mapping is integrable (in the sense that the integral of each one of its coordinates is finite). To do this we need the following notions and results.

2.2. Let U be a nonempty, convex and open subset of $\mathbf{R}^n$ and let $g : U \to \mathbf{R}$ be a convex function. Recall (see, for instance, [17, Lemma 1.2] or [12, Proposition 2.1.2]) that the right hand side derivative of g at a point $x \in U$ in a direction $d \in \mathbf{R}^n$, defined by

$$g^\circ(x, d) = \lim_{\tau \searrow 0} \frac{g(x + \tau d) - g(x)}{\tau}, \tag{10}$$

exists, is finite and satisfies

$$g^\circ(x, y - x) \leq g(y) - g(x). \tag{11}$$

This allows us to introduce the following notions.

Definition. The *modulus of local convexity of the function g at the point $x \in U$* is the function $\nu_g(x, \cdot) : [0, \infty) \to [0, \infty]$ given by[2]

$$\nu_g(x, t) \ = \ \inf\{D_g(y, x) \mid y \in U, \|y - x\| = t\}, \tag{12}$$

where

$$D_g(y, x) := g(y) - g(x) - g^\circ(x, y - x). \tag{13}$$

The function g is called *very convex* if, for each $x \in U$, $\nu_g(x, t) > 0$ whenever $t > 0$.

2.3. The notions of modulus of local convexity and of very convex functions introduced here are not equivalent to the analogous notions of modulus of (uniform) convexity and, respectively, uniformly convex function studied by Vladimirov et al. in [19]. Recall that the *modulus of convexity* of g is the function $\delta_g : [0, \infty) \to [0, \infty]$ defined by

$$\delta_g(t) \ = \ \inf\left\{\frac{\tau \cdot g(y) + (1 - \tau) \cdot g(x) - g(\tau y + (1 - \tau)x)}{\tau \cdot (1 - \tau)} \,\middle|\, \tau \in (0, 1), \|y - x\| = t\right\},$$

and that the function g is called uniformly convex if $\delta_g(t) > 0$ whenever $t > 0$. It follows from Remark 4 in [19] that $0 \leq \delta_g(t) \leq \nu_g(x, t)$ for all $t \in [0, \infty)$ and $x \in U$. However, in general, δ_g and ν_g are not equal. For instance, if $g : (0, \infty) \to \mathbf{R}$ is the function defined by $g(x) = x \cdot \ln x$, then

$$\begin{aligned} \nu_g(x, t) &= \min\{D_g(x + t, x), D_g(x - t, x)\} \\ &= x \cdot \left[\left(1 + \frac{t}{x}\right) \cdot \ln\left(1 + \frac{t}{x}\right) - \frac{t}{x}\right]. \end{aligned} \tag{14}$$

This shows that the function g is very convex because $\nu_g(x, 0) = 0$ and the function $\nu_g(x, \cdot)$ is strictly increasing on $(0, \infty)$ (its derivative is positive). Nevertheless, since for any $t > 0$, we have

$$0 \leq \delta_g(t) \leq \lim_{x \to \infty} \nu_g(x, t) = 0,$$

it follows that $\delta_g(t) = 0$ and that the function g is not uniformly convex.

2.4. Our proof that the function $\omega \to \Pi_\omega(x)$ is integrable for each $x \in \mathbf{R}^n_{++}$ is based on the properties of the modulus of local convexity summarized in the next result.

Proposition. *If U is an open, convex and unbounded subset of $\mathbf{R}^n$ and if $g : U \to \mathbf{R}$ is a convex function, then, for each $x \in U$, the function $\nu_g(x, \cdot)$ is everywhere finite and has the following properties*:

(i) *If $c \in [1, \infty)$, then, for any $t \in [0, \infty)$, $\nu_g(x, c \cdot t) \geq c \cdot \nu_g(x, t)$;*

(ii) *$\nu_g(x, \cdot)$ is nondecreasing and it is strictly increasing iff g is very convex*;

(iii) *$\nu_g(x, \cdot)$ is continuous from the right on $[0, \infty)$;*

(iv) *If $\bar{g} : clU \to \mathbf{R}$ is a convex continuous function whose restriction to U is g and if $\nu_g(x, \cdot)$ is continuous, then, for each $t \in [0, \infty)$,*

$$\nu_g(x, t) = \inf\{D_{\bar{g}}(y, x) \mid y \in clU, \|y - x\| = t\}. \tag{15}$$

Proof: Since U is unbounded and convex one can find a point $y \in U$ at any distance t from x. Thus, $\nu_g(x, \cdot)$ is everywhere finite. We prove the other statements point by point.

(i) It is sufficient to prove this statement for $c > 1$ and $t > 0$. In this case, let ε be a positive real number. According to (12), there exists a point $u \in U$ such that $\|u - x\| = c \cdot t$ and

$$\nu_g(x, c \cdot t) + \varepsilon > D_g(u, x) = g(u) - g(x) - g^\circ(x, u - x). \tag{16}$$

For every $\alpha \in (0, 1)$, denote $u_\alpha = \alpha u + (1 - \alpha)x$. Let $\beta = c^{-1}$ and observe that

$$\|u_\beta - x\| = \beta \cdot \|u - x\| = t. \tag{17}$$

Note that, for any $\alpha \in (0, 1)$,

$$\frac{\alpha}{\beta} u_\beta + \left(1 - \frac{\alpha}{\beta}\right)x = \frac{\alpha}{\beta}[\beta u + (1 - \beta)x] + \left(1 - \frac{\alpha}{\beta}\right)x = u_\alpha. \tag{18}$$

The function $\alpha \to \frac{g(x+\alpha(u-x))-g(x)}{\alpha}$ is nondecreasing on $(0, 1)$ because g is convex (see, for instance, [17, p. 2]). Therefore, by combining (10) and (16), we obtain

$$\nu_g(x, c \cdot t) + \varepsilon > g(u) - g(x) - \frac{g(x + \alpha(u - x)) - g(x)}{\alpha},$$

for all $\alpha \in (0, 1)$. As a consequence,

$$\begin{aligned}\nu_g(x, c\cdot t)+\varepsilon &> \frac{\alpha\cdot g(u)+(1-\alpha)\cdot g(x)-g(x+\alpha(u-x))}{\alpha}\\ &= \frac{\alpha\cdot g(u)+(1-\alpha)\cdot g(x)-\frac{\alpha}{\beta}\cdot g(u_\beta)-\left(1-\frac{\alpha}{\beta}\right)\cdot g(x)}{\alpha}\\ &\quad+\frac{\frac{\alpha}{\beta}\cdot g(u_\beta)+\left(1-\frac{\alpha}{\beta}\right)\cdot g(x)-g(u_\alpha)}{\alpha}\\ &= \frac{\beta\cdot g(u)+(1-\beta)\cdot g(x)-g(u_\beta)}{\beta}\\ &\quad+\frac{\frac{\alpha}{\beta}\cdot g(u_\beta)+\left(1-\frac{\alpha}{\beta}\right)\cdot g(x)-g\left(\frac{\alpha}{\beta}u_\beta+\left(1-\frac{\alpha}{\beta}\right)x\right)}{\alpha},\end{aligned}$$

where the last equality results from (18). The first term of the last sum is nonnegative because g is convex. Thus,

$$\begin{aligned}\nu_g(x, c\cdot t)+\varepsilon &> \frac{\frac{\alpha}{\beta}\cdot g(u_\beta)+\left(1-\frac{\alpha}{\beta}\right)\cdot g(x)-g\left(\frac{\alpha}{\beta}u_\beta+\left(1-\frac{\alpha}{\beta}\right)x\right)}{\alpha}\\ &= \frac{1}{\beta}\cdot\left[g(u_\beta)-g(x)-\frac{g\left(x+\frac{\alpha}{\beta}(u_\beta-x)\right)-g(x)}{\frac{\alpha}{\beta}}\right].\end{aligned}$$

Letting $\alpha\searrow 0$ and taking into account (16) and (10) we deduce that

$$\nu_g(x, c\cdot t)+\varepsilon > c\cdot D_g(u_\beta, x)\geq c\cdot \nu_g(x, t).$$

Since ε is an arbitrary positive real number, this proves (i).

(ii) Suppose that s and t are real numbers such that $0 < s < t$. Then,

$$\nu_g(x, t)=\nu_g\left(x, \frac{t}{s}\cdot s\right)\geq \frac{t}{s}\cdot \nu_g(x, s)\geq \nu_g(x, s), \tag{19}$$

where the first inequality follows from (i). Thus, $\nu_g(x, \cdot)$ is nondecreasing. If g is very convex, then the last inequality in (19) is strict and this shows that the function $\nu_g(x, \cdot)$ is strictly increasing on $(0, \infty)$. The converse is obvious. Hence, (ii) is proven.

(iii) First we show that $\nu_g(x, \cdot)$ is continuous from the right at 0. To this end, let $\{t_k\}_{k\in\mathbf{N}}$ be a sequence in $(0, 1)$ converging nonincreasingly to 0. Applying (ii) and (i) we deduce

$$\nu_g(x, 1)\geq \nu_g(x, \sqrt{t_k})=\nu_g\left(x, \frac{t_k}{\sqrt{t_k}}\right)\geq \frac{1}{\sqrt{t_k}}\cdot \nu_g(x, t_k).$$

Hence,

$$\nu_g(x, 0)=0\leq \lim_{k\to\infty}\nu_g(x, t_k)\leq \nu_g(x, 1)\cdot \lim_{k\to\infty}\sqrt{t_k}=0.$$

This shows that $\nu_g(x, \cdot)$ is continuous from the right at 0. Now suppose that $0 < s < t < \infty$. Fix an arbitrary real number $\varepsilon > 0$. According to (12), there exists a point $y_\varepsilon \in U$ such that $\|y_\varepsilon - x\| = s$ and

$$\nu_g(x, s) + \frac{\varepsilon}{4} > D_g(y_\varepsilon, x).$$

Applying (ii), we deduce that

$$0 \le |\nu_g(x,t) - \nu_g(x,s)| = \nu_g(x,t) - \nu_g(x,s) < \nu_g(x,t) - D_g(y_\varepsilon, x) + \frac{\varepsilon}{4}.$$

The function $D_g(\cdot, x)$ is continuous on U (see (13)) because g and $g^\circ(x, \cdot)$ are continuous (cf. [17, Proposition 1.19 and Corollary 1.7]). Therefore, there exists a number $\delta(\varepsilon) > 0$ such that, for any $z \in \mathbf{R}^n$ with $\|z - y_\varepsilon\| < \delta(\varepsilon)$, we have $z \in U$ and

$$|D_g(z, x) - D_g(y_\varepsilon, x)| < \frac{\varepsilon}{4}.$$

If $0 < t - s < \delta(\varepsilon)$, then the vector

$$y'_\varepsilon = \frac{t}{s} y_\varepsilon + \left(1 - \frac{t}{s}\right) x$$

satisfies $\|y'_\varepsilon - y_\varepsilon\| = t - s < \delta(\varepsilon)$ and $\|y'_\varepsilon - x\| = t$. Hence $y'_\varepsilon \in U$ and

$$0 \le \nu_g(x,t) - \nu_g(x,s) < D_g(y'_\varepsilon, x) - D_g(y_\varepsilon, x) + \frac{\varepsilon}{4} < \frac{\varepsilon}{4} + \frac{\varepsilon}{4} < \varepsilon.$$

This proves (iii).

(iv) Denote by ν_0 the right hand side of (15). Clearly, $\nu_0 \le \nu_g(x,t)$. Let $\varepsilon > 0$ be arbitrarily fixed and observe that there exists $z_\varepsilon \in clU$ such that $\|z_\varepsilon - x\| = t$ and $\nu_0 + \frac{\varepsilon}{4} > D_{\bar{g}}(z_\varepsilon, x)$. If $z_\varepsilon \in U$, then

$$\nu_0 + \frac{\varepsilon}{4} > D_{\bar{g}}(z_\varepsilon, x) = D_g(z_\varepsilon, x) \ge \nu_g(x,t). \tag{20}$$

Otherwise, there exists a sequence $\{z^k\}_{k \in \mathbf{N}} \subset U$ which converges to z_ε. For $t_k := \|z^k - x\|$ we have $\lim_{k\to\infty} t_k = t$ and, therefore, $\lim_{k\to\infty} \nu_g(x, t_k) = \nu_g(x,t)$ because $\nu_g(x, \cdot)$ is continuous. Thus, there exists a positive integer k_ε such that, for any integer $k \ge k_\varepsilon$, we have $\nu_0 + \frac{\varepsilon}{4} > D_g(z^k, x)$ and $\nu_g(x, t_k) \ge \nu_g(x,t) - \frac{\varepsilon}{4}$. By consequence, if $k \ge k_\varepsilon$, then

$$\nu_0 + \frac{\varepsilon}{4} > D_g(z^k, x) \ge \nu_g(x, t_k) \ge \nu_g(x,t) - \frac{\varepsilon}{4}.$$

This and (20) imply that, for any $\varepsilon > 0$, we have $\nu_0 + \frac{\varepsilon}{2} > \nu_g(x,t)$. The proof is complete. □

2.5. The next result emphasizes important features of the negentropy f. It shows that the averaged entropic projection operator $\Pi(x)$ exists and belongs to S no matter how $x \in \mathbf{R}^n_{++}$ is chosen.

Proposition.

(i) *The function* $f : \mathbf{R}^n_{++} \to \mathbf{R}$ *defined by* (5) *is a very convex function.*

(ii) *For each* $x \in \mathbf{R}^n_{++}$ *the averaged entropic projection* $\Pi(x) = \int_\Omega \Pi_\omega(x) d\mu(\omega)$ *exists and belongs to* S.

Proof:

(i) Let $\bar{g} : [0, \infty) \to \mathbf{R}$ be the continuous convex function defined by

$$\bar{g}(x) = \begin{cases} x \cdot \ln x, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

The function $\bar{f} : \mathbf{R}^n_+ \to \mathbf{R}$ given by

$$\bar{f}(x) = \sum_{i=1}^{n} \bar{g}\ (x_i), \tag{22}$$

is convex and continuous and its restriction to $\mathbf{R}^n_{++}$ is exactly f. Let

$$\bar{\nu}_f(x, t) = \inf \left\{ D_{\bar{f}}(y, x) \mid y \in \mathbf{R}^n_+, \|y - x\| = t \right\}.$$

Since the set $\{y \in \mathbf{R}^n_+ \mid \|y - x\| = t\}$ is compact in $\mathbf{R}^n$ and $D_{\bar{f}}(\cdot, x)$ is continuous on this set, there exists $y^* \in \mathbf{R}^n_+$ such that $\|y^* - x\| = t$ and

$$\nu_f(x, t) \geq \bar{\nu}_f(x, t) = D_{\bar{f}}(y^*, x) = \sum_{i=1}^{n} D_{\bar{g}}(y_i^*, x_i).$$

Let g be the restriction of $\bar{g}$ to $(0, \infty)$. The modulus of local convexity of g is given by (14) and is continuous in t. Therefore, we can apply Proposition 2.4(iv) and obtain that, for each $i \in \{1, 2, \ldots, n\}$,

$$D_{\bar{g}}(y_i^*, x_i) \geq \nu_g(x_i, |y_i^* - x_i|).$$

Hence,

$$\nu_f(x, t) \geq \sum_{i=1}^{n} \nu_g(x_i, |y_i^* - x_i|). \tag{23}$$

When $t > 0$, we have $\|y_i^* - x_i\| > 0$ for at least one index i. As noted in Section 2.3 the function g is very convex. Consequently, $\nu_g(x_i, |y_i^* - x_i|) > 0$ for at least one index i. This and (23) show that, if $t > 0$, then $\nu_f(x, t) > 0$, i.e., f is very convex.

(ii) It was noted in Section 1.2 that the function $\omega \to \Pi_\omega(x) : \Omega \to \mathbf{R}^n_{++}$ is measurable. By the definition of the integral, the averaged entropic projection $\Pi(x)$ is the limit of a coordinatewise nondecreasing sequence of elements of $\mathbf{R}^n_{++}$ consisting of points of the form $\sum_{j=1}^{m} \mu(\Omega_j) \cdot \Pi_{\omega_j}(x)$, where $\Omega_1, \ldots, \Omega_m$ is a partition of Ω and, for each

$j \in \{1, \ldots, m\}$, $\omega_j \in \Omega_j$ (see Halmos [15, Theorem B, p. 85]). Each such point is a convex combination of elements of S. Therefore, if we show that the averaged entropic projection $\Pi(x)$ is (coordinatewise) finite, then it will follow that $\Pi(x)$ must be an element of the closed set S. In order to show this, let $z \in \mathbf{R}^n_{++}$ be an arbitrary almost common point of the sets Q_ω, $\omega \in \Omega$. Set

$$I = \int_\Omega \|\Pi_\omega(x) - x\| d\mu(\omega).$$

This integral exists because the function $\omega \to \Pi_\omega(x)$ is measurable. It is sufficient to prove that I is finite because

$$\int_\Omega \|\Pi_\omega(x)\| d\mu(\omega) \leq I + \|x\|.$$

Observe that, for almost all $\omega \in \Omega$,

$$\nu_f(x, \|\Pi_\omega(x) - x\|) \leq D_f(\Pi_\omega(x), x) \leq D_f(z, x),$$

because z is an almost common point of the sets Q_ω, $\omega \in \Omega$. Since f is very convex, we have $\nu_f(x, 1) > 0$. Also, according to Proposition 2.4(i), for each $t \in [1, \infty)$, we have $\nu_f(x, t) \geq t \cdot \nu_f(x, 1)$. Hence, $\lim_{t\to\infty} \nu_f(x, t) = +\infty$. Consequently, there exists a number $t_0 > 0$ such that $D_f(z, x) < \nu_f(x, t_0)$. These show that, for almost all $\omega \in \Omega$,

$$\nu_f(x, \|\Pi_\omega(x) - x\|) \leq D_f(z, x) < \nu_f(x, t_0).$$

According to Proposition 2.4(ii), this can happen only if

$$\|\Pi_\omega(x) - x\| < t_0, \quad a.s.$$

Integrating this inequality we obtain that $I \leq t_0$ and the proof is complete. □

3. A convergence analysis of the AEPM

3.1. Proposition 2.5 ensures that, for any initial point $x^0 \in \mathbf{R}^n_{++}$, the sequence $\{x^k\}_{k\in\mathbf{N}}$ generated by the AEPM exists and is included in S. In this section we show that the sequence $\{x^k\}_{k\in\mathbf{N}}$ converges and its limit is an almost common point of the sets Q_ω, $\omega \in \Omega$. Our convergence proof is based on the following fact.

Lemma. *If $K \subseteq \mathbf{R}^n_{++}$ is nonempty, convex and closed, then the entropic projection $\Pi_K : \mathbf{R}^n_{++} \to K$ is continuous.*

Proof: Let $\{x^k\}_{k\in\mathbf{N}}$ be a convergent sequence in $\mathbf{R}^n_{++}$ such that $x^* := \lim_{k\to\infty} x^k \in \mathbf{R}^n_{++}$. Denote $y^k = \Pi_K(x^k)$, $k \in \mathbf{N}$. The sequence $\{y^k\}_{k\in\mathbf{N}}$ is bounded. Indeed, observe that for an arbitrarily fixed $z \in K$ and for all $k \in \mathbf{N}$ we have

$$D_f(y^k, x^k) \leq D_f(z, x^k). \tag{24}$$

Since the function $D_f(z, \cdot)$ is continuous, the sequence $\{D_f(z, x^k)\}_{k\in\mathbf{N}}$ is bounded. According to (24), this implies that $\{D_f(y^k, x^k)\}_{k\in\mathbf{N}}$ is bounded too. By (6), this cannot happen unless $\{y^k\}_{k\in\mathbf{N}}$ is bounded. Therefore, $\{y^k\}_{k\in\mathbf{N}}$ has an accumulation point and, obviously, all accumulation points of $\{y^k\}_{k\in\mathbf{N}}$ are in K. We show that any accumulation point of $\{y^k\}_{k\in\mathbf{N}}$ coincides with $\Pi_K(x)$. This will imply

$$\lim_{k\to\infty} \Pi_K(x^k) = \lim_{k\to\infty} y^k = \Pi_K(x^*),$$

that is, Π_K is continuous.

In order to prove that any accumulation point y^* of $\{y^k\}_{k\in\mathbf{N}}$ equals $\Pi_K(x^*)$, let $\{y^{k_p}\}_{p\in\mathbf{N}}$ be a subsequence of $\{y^k\}_{k\in\mathbf{N}}$ which converges to y^*. According to [11, Theorem 2.2], for each nonnegative integer p and for all $z \in K$,

$$\sum_{i=1}^{n} \left(\ln x_i^{k_p} - \ln y_i^{k_p}\right) \cdot \left(z_i - y_i^{k_p}\right) \leq 0.$$

Letting $p \to \infty$ we get

$$\sum_{i=1}^{n} (\ln x_i^* - \ln y_i^*) \cdot (z_i - y_i^*) \leq 0,$$

for all $z \in K$. Applying again Theorem 2.2 of [11], this implies $y^* = \Pi_K(x^*)$. □

3.2. Now we apply Lemma 3.1 to obtain another useful result.

Lemma. *If $K \subseteq \mathbf{R}^n_{++}$ is nonempty, convex and closed, then the function $\Phi_K : \mathbf{R}^n_{++} \to \mathbf{R}^n_{+}$ defined by*

$$\Phi_K(x) = D_f(\Pi_K(x), x), \tag{25}$$

is convex, differentiable and, for every $x \in \mathbf{R}^n_{++}$,

$$\frac{\partial \Phi_K}{\partial x_i}(x) = \frac{x_i - \Pi_K^i(x)}{x_i}, \quad 1 \leq i \leq n, \tag{26}$$

where $\Pi_K^i(x)$ denotes the ith coordinate of $\Pi_K(x)$.

Proof: Let $x, y \in \mathbf{R}^n_{++}$ and $\alpha \in (0, 1)$. Then,

$$\begin{aligned}
\Phi_K(\alpha x + (1-\alpha)y) &= D_f[\Pi_K(\alpha x + (1-\alpha)y), \alpha x + (1-\alpha)y] \\
&\leq D_f[\alpha \Pi_K(x) + (1-\alpha)\Pi_K(y), \alpha x + (1-\alpha)y] \\
&\leq \alpha \cdot D_f(\Pi_K(x), x) + (1-\alpha) \cdot D_f(\Pi_K(y), y) \\
&= \alpha \cdot \Phi_K(x) + (1-\alpha) \cdot \Phi_K(y),
\end{aligned}$$

where the first inequality follows from the convexity of K combined with the definition of the entropic projection Π_K (see (7)) and the second inequality results from the convexity of the function $D_f : \mathbf{R}^n_{++} \times \mathbf{R}^n_{++} \to \mathbf{R}$. This shows that Φ_K is convex. Now, in order to prove that Φ_K is differentiable at $x \in \mathbf{R}^n_{++}$, let $u \in \mathbf{R}^n \backslash \{0\}$ and suppose that $\theta \in (0, \infty)$ is sufficiently small so that $x + \theta u \in \mathbf{R}^n_{++}$. Then, applying (7) twice, we get

$$\begin{aligned}
&\theta^{-1} \cdot [D_f(\Pi_K(x+\theta u), x+\theta u) - D_f(\Pi_K(x+\theta u), x)] \\
&\quad \le \theta^{-1} \cdot [D_f(\Pi_K(x+\theta u), x+\theta u) - D_f(\Pi_K(x), x)] \\
&\quad \le \theta^{-1} \cdot [D_f(\Pi_K(x), x+\theta u) - D_f(\Pi_K(x), x)]. \qquad (27)
\end{aligned}$$

Observe that

$$\begin{aligned}
&\lim_{\theta \searrow 0} \frac{D_f(\Pi_K(x), x+\theta u) - D_f(\Pi_K(x), x)}{\theta} \\
&\quad = \lim_{\theta \searrow 0} \sum_{i=1}^n \left[u_i - \Pi_K^i(x) \cdot \frac{\ln(x_i + \theta \cdot u_i) - \ln x_i}{\theta \cdot u_i} \cdot u_i \right] \\
&\quad = \sum_{i=1}^n u_i \cdot \left(1 - \frac{\Pi_K^i(x)}{x_i}\right). \qquad (28)
\end{aligned}$$

According to Lemma 3.1, the entropic projection Π_K is continuous. Therefore,

$$\begin{aligned}
&\lim_{\theta \searrow 0} \frac{D_f(\Pi_K(x+\theta \cdot u), x+\theta u) - D_f(\Pi_K(x+\theta u), x)}{\theta} \\
&\quad = \lim_{\theta \searrow 0} \sum_{i=1}^n \left[u_i - \Pi_K^i(x+\theta u) \cdot \frac{\ln(x_i + \theta \cdot u_i) - \ln x_i}{\theta \cdot u_i} \cdot u_i \right] \\
&\quad = \sum_{i=1}^n u_i \cdot \left(1 - \frac{\Pi_K^i(x)}{x_i}\right). \qquad (29)
\end{aligned}$$

Letting $\theta \searrow 0$ in (27) and taking into account (28) and (29), we obtain that the middle term converges and its limit is the common limit of the left and the right hand sides, i.e., Φ_K is differentiable and (26) holds. □

3.3. Using the results above we are in a position to prove that AEPM generated sequences converge to almost common points of the sets Q_ω, $\omega \in \Omega$.

Theorem. *For any SCFP which satisfies* (3), *and for each* $x^0 \in \mathbf{R}^n_{++}$, *the AEPM generated sequence* $\{x^k\}_{k \in \mathbf{N}}$, *having* x^0 *as initial point, exists and converges to an almost common point of the sets* Q_ω, $\omega \in \Omega$.

Proof: The existence of the sequence $\{x^k\}_{k \in \mathbf{N}}$ follows from Proposition 2.5. We now prove the convergence of $\{x^k\}_{k \in \mathbf{N}}$. To this end, let z be an arbitrary almost common point of the sets $Q_\omega, \omega \in \Omega$. Denote $\Phi_\omega = \Phi_{Q_\omega}$, where Φ_{Q_ω} is the function given in (25) for $K = Q_\omega$. Define the function $\Phi : \mathbf{R}^n_{++} \to \mathbf{R}$ by

$$\Phi(x) = \int_\Omega \Phi_\omega(x)\, d\mu(\omega). \tag{30}$$

This function is well-defined because the mapping $\omega \to \Phi_\omega(x)$ is measurable (it is the composition of the measurable function $\omega \to \Pi_\omega(x)$ with the continuous function $D_f(\cdot, x)$) and, for each $\omega \in \Omega$ and all $x \in \mathbf{R}^n_{++}$,

$$0 \le \Phi_\omega(x) \le D_f(z, x). \tag{31}$$

Observe that Φ is convex because, for all $\omega \in \Omega$, the functions Φ_ω are convex (cf. Proposition 3.2). Note that, if $x \in \mathbf{R}^n_{++}$ and $u \in \mathbf{R}^n \backslash \{0\}$, then there exists a real number $\theta_0 > 0$ such that, for any $\theta \in (0, \theta_0]$, $x + \theta u \in \mathbf{R}^n_{++}$ and

$$\langle \nabla \Phi_\omega(x), u\rangle \le \frac{\Phi_\omega(x+\theta u) - \Phi_\omega(x)}{\theta} \le \Phi_\omega(x + \theta_0 u) - \Phi_\omega(x). \tag{32}$$

This is because the convexity of Φ_ω implies that the mapping

$$\theta \to \frac{\Phi_\omega(x+\theta u) - \Phi_\omega(x)}{\theta}$$

is nondecreasing on $(0, \theta_0]$. Since the functions $\omega \to \Phi_\omega(x + \theta_0 u) - \Phi_\omega(x)$ and $\omega \to \langle \nabla \Phi_\omega(x), u\rangle$ are integrable (cf. Proposition 2.5(ii), Proposition 3.2 and (31)), the inequality (32) allows us to apply the bounded convergence theorem in order to conclude that

$$\begin{aligned}
\lim_{\theta \searrow 0} \frac{\Phi(x+\theta u) - \Phi(x)}{\theta} &= \lim_{\theta \searrow 0} \int_\Omega \frac{\Phi_\omega(x+\theta u) - \Phi_\omega(x)}{\theta}\, d\mu(\omega) \\
&= \int_\Omega \lim_{\theta \searrow 0} \frac{\Phi_\omega(x+\theta u) - \Phi_\omega(x)}{\theta}\, d\mu(\omega) \\
&= \int_\Omega \langle \nabla \Phi_\omega(x), u\rangle\, d\mu(\omega).
\end{aligned}$$

This and Proposition 3.2 show that Φ is differentiable on $\mathbf{R}^n_{++}$ and

$$\frac{\partial \Phi}{\partial x_i}(x) = \int_\Omega \frac{x_i - \Pi^i_\omega(x)}{x_i} d\mu(\omega), \quad 1 \le i \le n, \tag{33}$$

where $\Pi^i_\omega(x)$ denotes the ith coordinate of $\Pi_\omega(x)$. Taking into account (4) and (33) we deduce that, for each nonnegative integer k,

$$\frac{\partial \Phi}{\partial x_i}(x^k) = \frac{x^k_i - x^{k+1}_i}{x^k_i}, \quad 1 \le i \le n. \tag{34}$$

By a specialization of [5, Lemma 1] we have that, for any $x \in \mathbf{R}^n_{++}$ and for almost all $\omega \in \Omega$,

$$D_f(\Pi_\omega(x), x) + D_f(z, \Pi_\omega(x)) \leq D_f(z, x). \tag{35}$$

Observe that the function $u \to D_f(u, x) + D_f(z, u)$ is convex on $\mathbf{R}^n_{++}$. Therefore, Jensen's inequality applies and gives

$$D_f(\Pi(x), x) + D_f(z, \Pi(x)) \leq \int_\Omega \left[D_f(\Pi_\omega(x), x) + D_f(z, \Pi_\omega(x))\right] d\mu(\omega). \tag{36}$$

Letting $x = x^k$ in (35) and in (36) and, after that, integrating the resulting inequality with respect to $\omega \in \Omega$, we obtain that, for each nonnegative integer k,

$$D_f(x^{k+1}, x^k) + D_f(z, x^{k+1}) \leq D_f(z, x^k). \tag{37}$$

Summing up the inequalities in (37) corresponding to $k = 0, 1, \ldots, m$, we get

$$\sum_{k=0}^{m} D_f(x^{k+1}, x^k) \leq D_f(z, x^0) - D_f(z, x^{k+1}) \leq D_f(z, x^0).$$

This shows that the series $\sum_{k=0}^{\infty} D_f(x^{k+1}, x^k)$ converges and, therefore, that

$$\lim_{k\to\infty} D_f(x^{k+1}, x^k) = 0. \tag{38}$$

Let $g : (0, \infty) \to \mathbf{R}$ be the function defined by $g(t) = t \cdot \ln t$. Then,

$$D_f(x^{k+1}, x^k) = \sum_{i=1}^{n} D_g\left(x_i^{k+1}, x_i^k\right),$$

where, for each i, $D_g(x_i^{k+1}, x_i^k) \geq 0$. This, together with (12) and (38), implies that, for each $i \in \{1, \ldots, n\}$,

$$\begin{aligned} 0 &\leq \varliminf_{k\to\infty} \nu_g\left(x_i^k, \left|x_i^{k+1} - x_i^k\right|\right) \leq \varlimsup_{k\to\infty} \nu_g\left(x_i^k, \left|x_i^{k+1} - x_i^k\right|\right) \\ &\leq \lim_{k\to\infty} D_g\left(x_i^{k+1}, x_i^k\right) \leq \lim_{k\to\infty} D_f(x^{k+1}, x^k) = 0, \end{aligned}$$

i.e.,

$$\lim_{k\to\infty} \nu_g\left(x_i^k, \left|x_i^{k+1} - x_i^k\right|\right) = 0, \quad 1 \leq i \leq n. \tag{39}$$

Observe that, according to (37), the sequence $\{D_f(z, x^k)\}_{k\in\mathbf{N}}$ is nonincreasing and bounded from above by $\gamma := D_f(z, x^0)$. The set $\{u \in \mathbf{R}^n_{++} \mid D_f(z, u) \leq \gamma\}$ is bounded in $\mathbf{R}^n$ and contains the sequence $\{x^k\}_{k\in\mathbf{N}}$. Thus, $\{x^k\}_{k\in\mathbf{N}}$ is bounded. Also, according to Proposition 2.5(ii), $\{x^k\}_{k\in\mathbf{N}} \subset S \subseteq \mathbf{R}^n_{++}$ (cf. (3)) and this shows that $\inf_{k\in\mathbf{N}} x_i^k > 0$, for each $i \in \{1, 2, \ldots, n\}$. These facts, combined with (14), imply that the equalities in (39) are satisfied only if

$$\lim_{k\to\infty} \frac{\left|x_i^{k+1} - x_i^k\right|}{x_i^k} = 0, \quad 1 \le i \le n.$$

According to (34), this means

$$\lim_{k\to\infty} \nabla\Phi(x^k) = 0. \tag{40}$$

The bounded sequence $\{x^k\}_{k\in\mathbf{N}}$ has a convergent subsequence $\{x^{k_p}\}_{p\in\mathbf{N}}$. Let $x^* = \lim_{p\to\infty} x^{k_p}$. Note that $x^* \in S \subset \mathbf{R}^n_{++}$ because $\{x^k\}_{k\in\mathbf{N}} \subset S$ and S is closed. The function Φ is convex and, therefore, for any $y \in \mathbf{R}^n_{++}$, we have

$$\begin{aligned}\Phi(y) - \Phi(x^*) &= (\Phi(y) - \Phi(x^{k_p})) + (\Phi(x^{k_p}) - \Phi(x^*))\\ &\ge \langle\nabla\Phi(x^{k_p}), y - x^{k_p}\rangle + \langle\nabla\Phi(x^*), x^{k_p} - x^*\rangle.\end{aligned}$$

Since $\{x^{k_p}\}_{p\in\mathbf{N}}$ is bounded, the right hand side of the last inequality converges to zero as $p \to \infty$ by (40). Thus, letting $p \to \infty$ on both sides of the last inequality, we obtain $\Phi(y) \ge \Phi(x^*)$, for all $y \in \mathbf{R}^n_{++}$. This shows that x^* is a minimizer of Φ over $\mathbf{R}^n_{++}$. Note that Φ is nonnegative and that $\Phi(z) = 0$. Hence, $\Phi(x^*) = \Phi(z) = 0$. According to (30), this means that

$$\int_\Omega D_f(\Pi_\omega(x^*), x^*)\, d\mu(\omega) = 0,$$

i.e., $D_f(\Pi_\omega(x^*), x^*) = 0$ for almost all $\omega \in \Omega$. Observe that $D_f(y, x)$ vanishes if and only if $x = y$. Consequently, we have $\Pi_\omega(x^*) = x^*$ for almost all $\omega \in \Omega$. This proves that x^* is an almost common point of the sets Q_ω, $\omega \in \Omega$.

Observe that (37) was proven above for an arbitrary almost common point z of the sets Q_ω, $\omega \in \Omega$. This means that it still holds for $z = x^*$. Therefore, the sequence $\{D_f(x^*, x^k)\}_{k\in\mathbf{N}}$ is nonincreasing. Hence, this sequence converges and it must have the same limit as its subsequence $\{D_f(x^*, x^{k_p})\}_{p\in\mathbf{N}}$, i.e.,

$$\lim_{k\to\infty} D_f(x^*, x^k) = \lim_{p\to\infty} D_f(x^*, x^{k_p}) = 0. \tag{41}$$

Since, for each $k \in \mathbf{N}$,

$$D_f(x^*, x^k) = \sum_{i=1}^n D_g\left(x_i^*, x_i^k\right),$$

and the terms of the last sum are nonnegative, we deduce from (41) that

$$\lim_{k\to\infty} D_g\left(x^*, x_i^k\right) = 0, \quad 1 \le i \le n,$$

and, as a consequence (see (12)),

$$\lim_{k\to\infty} \nu_g\left(x_i^k, \left|x_i^* - x_i^k\right|\right) = 0, \quad 1 \le i \le n.$$

The boundedness of $\{x^k\}_{k\in\mathbf{N}}$ and the fact that $\inf_{k\in\mathbf{N}} x_i^k > 0$, $1 \le i \le n$, combined with (14), show that this can happen only if

$$\lim_{k\to\infty} \frac{\left|x_i^* - x_i^k\right|}{x_i^k} = 0, \quad 1 \le i \le n,$$

that is, only if

$$\lim_{k\to\infty} \left|x_i^* - x_i^k\right| = 0, \quad 1 \le i \le n.$$

This means that $\{x^k\}_{k\in\mathbf{N}}$ converges to x^* which is an almost common point of the sets Q_ω, $\omega \in \Omega$. The proof is complete. □

4. An example and comments

4.1. In this section we demonstrate how the AEPM works and illustrate the idea mentioned in Sections 1.3 and 1.4 that employing Bregman projection type methods for solving the SCFP can have computational advantages when the function f with respect to which the Bregman projections are computed is fitted to the nature of the sets Q_ω in the given problem. To this purpose we consider the following example.

Let Ω be the interval $[a, b]$ and, for each $i \in \{0, 1, 2, \ldots, n\}$, let $\phi_i : \Omega \to (0, \infty)$ be a continuous function. For each $\omega \in \Omega$, we consider the function $g_\omega : \mathbf{R}_+^n \to \mathbf{R}$ given by

$$g_\omega(y) = \sum_{i=1}^{n} \phi_i(\omega) \cdot y_i \cdot \ln y_i + \phi_0(\omega), \tag{42}$$

with the usual convention that $0 \cdot \ln 0 = 0$. We want to find an element $x^* \in \mathbf{R}_{++}^n$ which, for all $\omega \in \Omega$, satisfies $g_\omega(x^*) \le 0$. To this end, observe that, for each $\omega \in \Omega$, the set

$$Q_\omega = \left\{z \in \mathbf{R}_+^n \mid g_\omega(z) \le 0\right\} \tag{43}$$

is convex and closed and that finding a point x^* as mentioned above, is equivalent to finding a common element of the sets Q_ω, $\omega \in \Omega$. We assume that the sets Q_ω have common points (that is, $\bigcap_{\omega\in\Omega} Q_\omega \ne \emptyset$) and that (3) is satisfied.

4.2. We reduce the problem of finding a point x^* in $\bigcap_{\omega\in\Omega} Q_\omega$ to a SCFP as follows. We provide Ω with the probability structure $(\mathcal{A}, \mu)$, where $\mathcal{A}$ is the family of all Lebesgue measurable subsets of Ω and $\mu = (b-a)^{-1} \cdot \lambda$, and λ denotes the Lebesgue measure on Ω. The probability space $(\Omega, \mathcal{A}, \mu)$ is complete and the point-to-set mapping $Q : \Omega \to \mathbf{R}_+^n$, which assigns to each $\omega \in \Omega$ the closed convex set Q_ω, is measurable (its graph is closed).

Note that, for every $x \in \mathbf{R}^n_+$, the function $\omega \to g_\omega(x) : \Omega \to \mathbf{R}$ is continuous. Therefore, any almost common point of the sets Q_ω, $\omega \in \Omega$, is a point in $\bigcap_{\omega\in\Omega} Q_\omega$. This is so because if z is an almost common point of the given sets and $z \notin \bigcap_{\omega\in\Omega} Q_\omega$, then $g_{\omega_0}(z) > 0$, for some $\omega_0 \in \Omega$ and this implies $g_\omega(z) > 0$ for all $\omega \in N_0 \cap [a, b]$, where N_0 is an open neighborhood of ω_0; this is a contradiction since $\mu(N_0 \cap [a, b]) > 0$. Hence, the problem of finding a point $x^* \in \bigcap_{\omega\in\Omega} Q_\omega$ is exactly the SCFP of finding an almost common point of the sets Q_ω, $\omega \in \Omega$.

4.3. Under the assumptions made in Section 4.1, Theorem 3.2 guarantees the convergence of each AEPM generated sequence to a common point of the sets Q_ω, $\omega \in \Omega$. For determining an AEPM generated sequence we would like to have an explicit formula for computing the entropic projections required in (4). To this end we have to solve, for every fixed $x \in \mathbf{R}^n_{++}$, the optimization problem

$$\begin{cases} \min D_{\bar{f}}(y, x) \\ \text{such that} \\ g_\omega(y) \le 0, \quad y \in \mathbf{R}^n_+, \end{cases} \tag{44}$$

where $\bar{f}$ is defined in (22). The Kuhn-Tucker conditions for this optimization problem (note that, according to [4, Theorem 3.12], the optimum of (44) is necessarily a point in $\mathbf{R}^n_{++}$) are

$$\begin{cases} \ln y_i - \ln x_i + \alpha \cdot \phi_i(\omega) \cdot (\ln y_i + 1) = 0, \quad 1 \le i \le n, \\ g_\omega(y) = 0, \end{cases}$$

where $\alpha \ge 0$ represents the Lagrange multiplier. The solution of this system of equations is the vector y with the coordinates

$$y_i = \exp\left(-\frac{\alpha \cdot \phi_i(\omega) - \ln x_i}{\alpha \cdot \phi_i(\omega) + 1}\right), \quad 1 \le i \le n, \tag{45}$$

where α is a positive solution of the equation

$$\sum_{i=1}^n \phi_i(\omega) \cdot \frac{\alpha \cdot \phi_i(\omega) - \ln x_i}{\alpha \cdot \phi_i(\omega) + 1} \cdot \exp\left(-\frac{\alpha \cdot \phi_i(\omega) - \ln x_i}{\alpha \cdot \phi_i(\omega) + 1}\right) - \phi_0(\omega) = 0. \tag{46}$$

This equation can be solved by standard numerical procedures.

4.4. Theorem 4.4 in [8] ensures that the SCFP formulated in Section 4.1 can be also approached with the expected metric projection method of Butnariu and Flåm [8]. This method, applied to the given SCFP, produces sequences which converge to common points of the sets Q_ω, $\omega \in \Omega$, defined in (43). Computing the metric projection $P_\omega(x)$ of the point

x onto the set Q_ω amounts to solving the optimization problem

$$\begin{cases} \min \|y - x\|^2 \\ \text{such that} \\ g_\omega(y) \le 0, \quad y \in \mathbf{R}^n_+. \end{cases} \tag{47}$$

Compared with the relatively simple way of solving (44) which practically consists of computing a positive solution of (46) and of using (45), finding a solution of (47) may be computationally costly. In fact, writing down the Kuhn-Tucker conditions for the optimization problem (47) shows that this requires numerical computation of a solution $\alpha, y_1, \ldots, y_n$ to the system

$$\begin{cases} 2 \cdot (y_i - x_i) + \alpha \cdot \phi_i(\omega)(\ln y_i + 1) = 0; \quad 1 \le i \le n, \\ g_\omega(y) \le 0. \end{cases}$$

4.5. The relative easiness with which the entropic projections $\Pi_\omega(x)$ were determined in Section 4.3 advocates the AEPM in this specific case. In other situations the expected metric projection method may be more friendly than the AEPM. In fact, this happens when one has to solve SCFPs like those required for computing the inverse of the Radon transform in computed tomography (see [8, Section 6]). This is a SCFP similar to that in Section 4.1 but with

$$Q_\omega := \{z \in \mathbf{R}^n \mid \langle a(\omega), z \rangle = b(\omega)\}. \tag{48}$$

In this case,

$$P_\omega(x) = x + \frac{b(\omega) - \langle a(\omega), x \rangle}{\|a(\omega)\|^2} \cdot a(\omega).$$

However, computing the entropic projection $\Pi_\omega(x)$ onto the set Q_ω defined in (48), provided that they exist, is rather complicated. It can be shown that, if $Q_\omega \bigcap \mathbf{R}^n_{++} \neq \emptyset$ for all $\omega \in \Omega$, then

$$\Pi^i_\omega(x) = x_i \cdot [v_\omega(x)]^{a_n(\omega)},$$

where $v_\omega(x)$ is the unique positive solution of the equation

$$\sum_{i=1}^n a_i(\omega) \cdot x_i \cdot v^{a_i(\omega)} = b(\omega).$$

The comparison of the AEPM and of the expected metric projection method emphasizes the practical meaning of the conjecture in Section 1.5 that our results concerning the convergence of the AEPM can be extended to procedures based on functions other than the negentropy and the square of the norm.

Acknowledgments

The work of Y. Censor was partially supported by the grant HL-28438 at the Medical Image Processing Group (MIPG), Department of Radiology, the University of Pennsylvania, Philadelphia, PA, USA. The work of S. Reich was partially supported by the Fund for the Promotion of Research at the Technion and by the M. and M.L. Bank Mathematics Research Fund at the Technion.

The authors are grateful to the referees for their suggestions.

Notes

1. All over this paper the integral of a vector function is the vector of the integrals of its coordinates.
2. We use the convention that the infimum of the empty set is $+\infty$.

References

1. Y. Alber and D. Butnariu, “Convergence of Bregman-projection methods for solving consistent convex feasibility problems in reflexive Banach spaces,” J. Optim. Theory and Appl., vol. 92, pp. 33–61, 1997.
2. J.-P. Aubin and H. Frankowska, Set-Valued Analysis, Birkhäuser: Boston, 1990.
3. H.H. Bauschke and J.M. Borwein, “On projection algorithms for solving convex feasibility problems,” SIAM Review, vol. 38, pp. 367–426, 1996.
4. H.H. Bauschke and J.M. Borwein, “Legendre functions and the method of random Bregman projections,” Convex Analysis (to appear).
5. L.M. Bregman, “The relaxation method for finding the common point of convex sets and its application to the solution of convex programming,” USSR Comp. Math. and Math. Phys., vol. 7, pp. 200–217, 1967.
6. D. Butnariu, “The expected-projection method: Its behavior and applications to linear operator equations and convex optimization,” J. Applied Analysis, vol. 1, pp. 95–108, 1995.
7. D. Butnariu and Y. Censor, “Strong convergence of almost simultaneous projection methods in Hilbert spaces,” J. Comput. Appl. Math., vol. 53, pp. 33–42, 1994.
8. D. Butnariu and S.D. Flåm, “Strong convergence of expected-projection methods in Hilbert spaces,” Numer. Funct. Anal. Optim., vol. 15, pp. 601–636, 1995.
9. Y. Censor and A. Lent, “An iterative row-action method for interval convex programming,” J. Optim. Theory and Appl., vol. 34, no. 3, pp. 321–353, 1981.
10. Y. Censor and T. Elfving, “A multiprojection algorithm using Bregman projections in a product space,” Numerical Algorithms, vol. 8, pp. 221–239, 1994.
11. Y. Censor and S. Reich, “Iterations of paracontractions and firmly nonexpansive operators with applications to feasibility and optimization,” Optimization, vol. 37, pp. 323–339, 1996.
12. F.H. Clarke, Optimization and Nonsmooth Analysis, John Willey & Sons: New York, 1983.
13. P.L. Combettes, “The foundations of set theoretic estimation,” Proc. IEEE, vol. 81, pp. 182–208, 1993.
14. A.R. De Pierro and A.N. Iusem, “A relaxed version of Bregman’s method for convex programming,” J. Optim. Theory Appl., vol. 51, pp. 421–440, 1986.
15. P.R. Halmos, Measure Theory, Springer-Verlag: New York, 1974.
16. A.N. Iusem and A.R. De Pierro, “Convergence results for an accelerated Cimmino algorithm,” Numer. Math., vol. 49, pp. 347–368, 1986.
17. R.R. Phelps, Convex Functions, Monotone Operators and Differentiability, Springer-Verlag: Berlin, 1993.
18. R.T. Rockafellar and R.J.-B. Wets, “Scenarios and policy aggregation in optimization under uncertainty,” Mathematics of Operation Research, vol. 16, pp. 119–147, 1991.
19. A.A. Vladimirov, Y.E. Nesterov, and Y.N. Cekanov, “Uniformly convex functions,” Vestnik Moskovskaya Universiteta, Series Matematika i Kybernetika, vol. 3, pp. 12–23, 1978.
20. R.J.-B. Wets, “Stochastic Programming,” in Handbook of Operation Research and Management Sciences, G.L. Nemhauser, A.H.G. Rinnooy Kan and M.J. Todd (Eds.), vol. 1: Optimization, North-Holland, Amsterdam, pp. 573–629, 1989.