65

INHERENTLY PARALLEL ALGORITHMS IN FEASIBILITY AND OPTIMIZATION AND THEIR APPLICATIONS

Edited by

Dan BUTNARIU
Department of Mathematics
University of Haifa, Israel

Yair CENSOR
Department of Mathematics
University of Haifa, Israel

Simeon REICH
Department of Mathematics
The Technion – Israel Institute of Technology, Haifa, Israel

2001

ELSEVIER
Amsterdam – London – New York – Oxford – Paris – Shannon – Tokyo

CONTENTS

Inherently Parallel Algorithms in Feasibility and Optimization and their Applications
D. Butnariu, Y. Censor and S. Reich (Editors)

BREGMAN-LEGENDRE MULTIDISTANCE PROJECTION ALGORITHMS FOR CONVEX FEASIBILITY AND OPTIMIZATION

Charles Byrne[a]

[a]Department of Mathematical Sciences,
University of Massachusetts Lowell,
1 University Ave., Lowell, MA 01854, USA

The *convex feasibility problem* (CFP) is to find a member of the nonempty set $C = \bigcap_{i=1}^{I} C_i$, where the C_i are closed convex subsets of R^J. The *multidistance successive generalized projection* (MSGP) algorithm extends Bregman's *successive generalized projection* (SGP) algorithm for solving the CFP to permit the use of generalized projections onto the C_i associated with Bregman-Legendre functions f_i that may vary with the index i. The MSGP method depends on the selection of a super-coercive Bregman-Legendre function h whose Bregman distance D_h satisfies the inequality $D_h(x,z) \geq D_{f_i}(x,z)$ for all $x \in \operatorname{dom} h \subseteq \bigcap_{i=1}^{I} \operatorname{dom} f_i$ and all $z \in \operatorname{int} \operatorname{dom} h$, where $\operatorname{dom} h = \{x | h(x) < +\infty\}$.

The MSGP method is used to obtain an iterative solution procedure for the *split feasibility problem* (SFP): given the M by N matrix A and closed convex sets K and Q in R^N and R^M, respectively, find x in K with Ax in Q.

If $I = 1$ and $f := f_1$ has a unique minimizer $\hat{x}$ in $\operatorname{int} \operatorname{dom} h$, then the MSGP iteration using $C_1 = \{\hat{x}\}$ is

$$\nabla h(x^{k+1}) = \nabla h(x^k) - \nabla f(x^k).$$

This suggests an interior point algorithm that could be applied more broadly to minimize a convex function f over the closure of $\operatorname{dom} h$.

1. INTRODUCTION

The *convex feasibility problem* (CFP) is to find a member of the nonempty set $C = \bigcap_{i=1}^{I} C_i$, where the C_i are closed convex subsets of R^J. In most applications the sets C_i are more easily described than the set C and algorithms are sought whereby a member of C is obtained as the limit of an iterative procedure involving (exact or approximate) orthogonal or generalized projections onto the individual sets C_i. Such algorithms are the topic of this paper.

In his often cited paper [3] Bregman introduced a class of functions that have come to be called *Bregman functions* and used the associated *Bregman distances* to define generalized projections onto closed convex sets (see the book by Censor and Zenios [9] for details concerning Bregman functions).

In [2] Bauschke and Borwein introduce the related class of *Bregman-Legendre functions* and show that these functions provide an appropriate setting in which to study Bregman distances and generalized projections associated with such distances.

Bregman's *successive generalized projection* (SGP) method uses projections with respect to Bregman distances to solve the convex feasibility problem. Let $f : R^J \to (-\infty, +\infty]$ be a closed, proper convex function, with essential domain $D = \operatorname{dom} f = \{x | f(x) < +\infty\}$ and $\emptyset \neq \operatorname{int} D$. Denote by $D_f(\cdot, \cdot) : D \times \operatorname{int} D \to [0, +\infty)$ the Bregman distance, given by

$$D_f(x, z) = f(x) - f(z) - \langle \nabla f(z), x - z \rangle \tag{1}$$

and by $P^f_{C_i}$ the Bregman projection operator associated with the convex function f and the convex set C_i; that is

$$P^f_{C_i} z = \arg\min_{x \in C_i \cap D} D_f(x, z). \tag{2}$$

Bregman considers the following iterative procedure:

Algorithm 1.1 Bregman's method of Successive Generalized Projections (SGP): *Beginning with* $x^0 \in \operatorname{int} \operatorname{dom} f$, *for* $k = 0, 1, ...,$ *let* $i = i(k) := k(\operatorname{mod} \mathrm{I}) + 1$ *and*

$$x^{k+1} = P^f_{C_{i(k)}}(x^k). \tag{3}$$

He proves that the sequence $\{x^k\}$ given by (3) converges to a member of $C \cap \operatorname{dom} f$, whenever this set is nonempty and the function f is what came to be called a Bregman function ([3]).

Investigations in [4] into several well known iterative algorithms, including the 'expectation maximization maximum likelihood' (EMML) method, the 'multiplicative algebraic reconstruction technique' (MART) as well as block-iterative and simultaneous versions of MART revealed that the iterative step of each algorithm involved weighted arithmetic or geometric means of Bregman projections onto hyperplanes; interestingly, the projections involved were associated with Bregman distances that differed from one hyperplane to the next. This representation of the EMML algorithm as a weighted arithmetic mean of Bregman projections provided the key step in obtaining block-iterative and row-action versions of EMML. Because it is well known that convergence is not guaranteed if one simply extends Bregman's algorithm to multiple distances by replacing the single distance D_f in (3) with multiple distances D_{f_i}, the appearance of distinct distances in these algorithms suggested that a somewhat more sophisticated algorithm employing multiple Bregman distances might be possible.

In [5] such an iterative multiprojection method for solving the CFP, called the *multi-distance successive generalized projection* (MSGP) method, was presented in the context of Bregman functions. In this paper we prove convergence of the MSGP method in the framework of Bregman-Legendre functions. The MSGP extends Bregman's SGP method by allowing the Bregman projection onto each set C_i to be performed with respect to a Bregman distance D_{f_i} derived from a Bregman-Legendre function f_i. The MSGP method depends on the selection of a super-coercive Bregman-Legendre function h whose Bregman distance D_h satisfies the inequality $D_h(x, z) \geq D_{f_i}(x, z)$ for all $x \in \operatorname{dom} h \subseteq \bigcap_{i=1}^I \operatorname{dom} f_i$ and all $z \in \operatorname{int} \operatorname{dom} h$, where $\operatorname{dom} h = \{x | h(x) < +\infty\}$. By using different Bregman distances for different convex sets, we found that we can sometimes calculate the desired Bregman projections in closed form, thereby obtaining computationally tractable iterative algorithms (see [4]).

To illustrate the MSGP method we use it to obtain an iterative solution procedure for the *split feasibility problem* (SFP): given the M by N matrix A and closed convex sets K and Q in R^N and R^M, respectively, find x in K with Ax in Q.

Consideration of a special case of the MSGP, involving only a single convex set C_1, leads us to an interior point optimization method. If $I = 1$ and $f := f_1$ has a unique minimizer $\hat{x}$ in $\operatorname{int}\operatorname{dom} h$, then the MSGP iteration using $C_1 = \{\hat{x}\}$ is

$$\nabla h(x^{k+1}) = \nabla h(x^k) - \nabla f(x^k). \tag{4}$$

This suggests an *interior point algorithm* (IPA) that could be applied more broadly to minimize a convex function f over the closure of $\operatorname{dom} h$.

We begin with a brief discussion of Bregman-Legendre functions, closely following the treatment in Bauschke and Borwein [2]. Then we present the MSGP method and prove convergence, in the context of Bregman-Legendre functions. In the section following we apply the MSGP to the SFP. Finally, we investigate the IPA suggested by the MSGP algorithm.

2. BREGMAN-LEGENDRE FUNCTIONS

In [2] Bauschke and Borwein show convincingly that the Bregman-Legendre functions provide the proper context for the discussion of Bregman projections onto closed convex sets. The summary here follows closely the discussion given in [2].

A convex function $f : R^J \to [-\infty, +\infty]$ is *proper* if there is no x with $f(x) = -\infty$ and some x with $f(x) < +\infty$. The *essential domain* of f is $\operatorname{dom} f = D = \{x | f(x) < +\infty\}$. A proper convex function f is *closed* if it is lower semi-continuous. The *conjugate function* associated with f is the function f^*, with $f^*(x^*) = \sup_z(\langle x^*, z\rangle - f(z))$.

Following [14] we say that a closed proper convex function f is *essentially smooth* if $\operatorname{int} D$ is not empty, f is differentiable on $\operatorname{int} D$ and $x^n \in \operatorname{int} D$, with $x^n \to x \in \operatorname{bdry} D$, implies that $||\nabla f(x^n)|| \to +\infty$. Here $\operatorname{int} D$ and $\operatorname{bdry} D$ denote the interior and boundary of the set D.

An essentially smooth function f is a *Legendre function* if f is strictly convex on the set $\operatorname{int} D$; the gradient operator ∇f is then a topological isomorphism with ∇f^* as its inverse. The gradient operator ∇f maps $\operatorname{int}\operatorname{dom} f$ onto $\operatorname{int}\operatorname{dom} f^*$. If $\operatorname{int}\operatorname{dom} f^* = R^J$ then the range of ∇f is R^J and the equation $\nabla f(x) = y$ can be solved for every $y \in R^J$. In order for $\operatorname{int}\operatorname{dom} f^* = R^J$ it is necessary and sufficient that the Legendre function f be *super-coercive*, that is,

$$\lim_{||x|| \to +\infty} \frac{f(x)}{||x||} = +\infty.$$

If f is Legendre and the essential domain of f is bounded, then f is necessarily super-coercive and its gradient operator is a mapping onto the space R^J.

Let K be a nonempty closed convex set with $K \cap \operatorname{int} D \neq \emptyset$. If f is Legendre, then the Bregman projection $P_K^f(z)$ exists, for all $z \in \operatorname{int} D$, is uniquely defined and is in $\operatorname{int} D$;

$P_K^f(z)$ is the unique member of $K \cap \text{int}D$ satisfying the inequality

$$\langle \nabla f(P_K^f(z)) - \nabla f(z), P_K^f(z) - c \rangle \geq 0, \tag{5}$$

for all $c \in K$. From this one obtains the *Bregman inequality*:

$$D_f(c, z) \geq D_f(c, P_K^f(z)) + D_f(P_K^f(z), z), \tag{6}$$

for all $c \in K \cap D$.

Following [2], we say that a Legendre function f is a *Bregman-Legendre* function if the following properties hold; the specific citations below refer to [2].

B1: For x in D and any $a > 0$ the set $\{z | D_f(x, z) \leq a\}$ is bounded (BL0 and BL1 of Def. 5.2).
B2: If x is in D but not in $\text{int}\, D$, for each positive integer n, y^n is in $\text{int}\, D$ with $y^n \to y \in \text{bdry}\, D$ and if $\{D_f(x, y^n)\}$ remains bounded, then $D_f(y, y^n) \to 0$, so that $y \in D$ (BL2 of Def. 5.2).
B3: If x^n and y^n are in $\text{int}\, D$, with $x^n \to x$ and $y^n \to y$, where x and y are in D but not in $\text{int}\, D$, and if $D_f(x^n, y^n) \to 0$ then $x = y$ (BL3 of Def. 5.2) .

The following results are proved in somewhat more generality in [2]; all references below are to that paper. We assume throughout that f is a Bregman-Legendre function with essential domain D.
R1: If $y^n \in \text{int}\, D$ and $y^n \to y \in \text{int}\, D$, then $D_f(y, y^n) \to 0$. If, in addition, $x^n \in \text{int}\, D$, $x^n \to x \in \text{int}\, D$ and $D_f(x^n, y^n) \to 0$, then $x = y$.
Proof: Both f and ∇f are continuous on $\text{int}\, D$.
R2: If x and $y^n \in \text{int}\, D$ and $y^n \to y \in \text{bdry}\, D$, then $D_f(x, y^n) \to +\infty$ (Theorem 3.8 (i)) .
R3: If $x^n \in D$, $x^n \to x \in D$, $y^n \in \text{int}\, D$, $y^n \to y \in D$, $\{x, y\} \cap \text{int}\, D \neq \emptyset$ and $D_f(x^n, y^n) \to 0$, then $x = y$ and $y \in \text{int}\, D$ (Theorem 3.9(iii)).
R4: If x and y are in D, but are not in $\text{int}\, D$, $y^n \in \text{int}\, D$, $y^n \to y$ and $D_f(x, y^n) \to 0$, then $x = y$ (Proposition 5.5).
As a consequence of these results we have the following.
R5: If $\{D_f(x, y^n)\} \to 0$, for $y^n \in \text{int}\, D$ and $x \in D$, then $\{y^n\} \to x$.
Proof of R5: By Property B1 above it follows that the sequence $\{y^n\}$ is bounded; without loss of generality, we assume that $\{y^n\} \to y$, for some $y \in \overline{D}$.

If y is in $\text{int}\, D$, then, by the continuity of both f and ∇f on $\text{int}\, D$, we have $\{D_f(x, y^n)\} \to D_f(x, y)$. Therefore, $D_f(x, y) = 0$ and $x = y$.

If y is in $\overline{D}$ but not in $\text{int}\, D$, then, by R2, x is not in $\text{int}\, D$. Then, using B2, we have that y is in D; by R4, it follows that $x = y$. This completes the proof of R5. ∎

We turn now to the MSGP algorithm.

3. THE MSGP ALGORITHM

We begin by setting out the assumptions we shall make and the notation we shall use in this section.

Assumptions and notation:

We make the following assumptions throughout this section. Let $C = \cap_{i=1}^{I} C_i$ be the nonempty intersection of closed convex sets C_i. The function h is super-coercive and Bregman-Legendre with essential domain $D = \operatorname{dom} h$ and $C \cap \operatorname{dom} h \neq \emptyset$. For $i = 1, 2, ..., I$ the function f_i is also Bregman-Legendre, with $D \subseteq \operatorname{dom} f_i$, so that $\operatorname{int} D \subseteq \operatorname{int} \operatorname{dom} f_i$; also $C_i \cap \operatorname{int} \operatorname{dom} f_i \neq \emptyset$. For all $x \in \operatorname{dom} h$ and $z \in \operatorname{int} \operatorname{dom} h$ we have $D_h(x, z) \geq D_{f_i}(x, z)$, for each i.

The MSGP algorithm:

Algorithm 3.1 The MSGP algorithm: *Let* $x^0 \in \operatorname{int} \operatorname{dom} h$ *be arbitrary. For* $k = 0, 1, ...$ *and* $i(k) := k (\operatorname{mod} I) + 1$ *let*

$$x^{k+1} = \nabla h^* \Big(\nabla h(x^k) - \nabla f_{i(k)}(x^k) + \nabla f_{i(k)}(P_{C_{i(k)}}^{f_{i(k)}}(x^k)) \Big). \tag{7}$$

A preliminary result:

For each $k = 0, 1, ...$ define the function $G^k(\cdot) : \operatorname{dom} h \rightarrow [0, +\infty)$ by

$$G^k(x) = D_h(x, x^k) - D_{f_{i(k)}}(x, x^k) + D_{f_{i(k)}}(x, P_{C_{i(k)}}^{f_{i(k)}}(x^k)). \tag{8}$$

The next proposition provides a useful identity, which can be viewed as an analogue of Pythagoras' theorem. The proof is not difficult and we omit it.

Proposition 3.1 *For each* $x \in \operatorname{dom} h$, *each* $k = 0, 1, ...,$ *and* x^{k+1} *given by (7) we have*

$$G^k(x) = G^k(x^{k+1}) + D_h(x, x^{k+1}). \tag{9}$$

Consequently, x^{k+1} *is the unique minimizer of the function* $G^k(\cdot)$.

This identity (9) is the key ingredient in the convergence proof for the MSGP algorithm.

The MSGP convergence theorem:

We shall prove the following convergence theorem:

Theorem 3.1 *Let* $x^0 \in \operatorname{int} \operatorname{dom} h$ *be arbitrary. Any sequence* x^k *obtained from the iterative scheme given by Algorithm 3.1 converges to* $x^\infty \in C \cap \operatorname{dom} h$. *If the sets* C_i *are hyperplanes, then* x^∞ *minimizes the function* $D_h(x, x^0)$ *over all* $x \in C \cap \operatorname{dom} h$; *if, in addition,* x^0 *is the global minimizer of* h, *then* x^∞ *minimizes* $h(x)$ *over all* $x \in C \cap \operatorname{dom} h$.

Proof: Let c be a member of $C \cap \operatorname{dom} h$. From the Pythagorean identity (9) it follows that

$$G^k(c) = G^k(x^{k+1}) + D_h(c, x^{k+1}). \tag{10}$$

Using the definition of $G^k(\cdot)$, we write

$$G^k(c) = D_h(c, x^k) - D_{f_{i(k)}}(c, x^k) + D_{f_{i(k)}}(c, P_{C_{i(k)}}^{f_{i(k)}}(x^k)). \tag{11}$$

From Bregman's inequality (6) we have that

$$D_{f_{i(k)}}(c, x^k) - D_{f_{i(k)}}(c, P_{C_{i(k)}}^{f_{i(k)}}(x^k)) \geq D_{f_{i(k)}}(P_{C_{i(k)}}^{f_{i(k)}}(x^k), x^k). \tag{12}$$

Consequently, we know that

$$D_h(c, x^k) - D_h(c, x^{k+1}) \geq G^k(x^{k+1}) + D_{f_{i(k)}}(P_{C_{i(k)}}^{f_{i(k)}}(x^k), x^k) \geq 0. \tag{13}$$

It follows that $\{D_h(c, x^k)\}$ is decreasing and finite and the sequence $\{x^k\}$ is bounded. Therefore, $\{D_{f_{i(k)}}(P_{C_{i(k)}}^{f_{i(k)}}(x^k), x^k)\} \to 0$ and $\{G^k(x^{k+1})\} \to 0$; from the definition of $G^k(x)$ it follows that $\{D_{f_{i(k)}}(x^{k+1}, P_{C_{i(k)}}^{f_{i(k)}}(x^k))\} \to 0$ as well. Using the Bregman inequality we obtain the inequality

$$D_h(c, x^k) \geq D_{f_{i(k)}}(c, x^k) \geq D_{f_{i(k)}}(c, P_{C_{i(k)}}^{f_{i(k)}}(x^k)), \tag{14}$$

which tells us that the sequence $\{P_{C_{i(k)}}^{f_{i(k)}}(x^k)\}$ is also bounded. Let x^* be an arbitrary cluster point of the sequence $\{x^k\}$ and let $\{x^{k_n}\}$ be a subsequence of the sequence $\{x^k\}$ converging to x^*.

We first show that $x^* \in \operatorname{dom} h$ and $\{D_h(x^*, x^k)\} \to 0$. If x^* is in $\operatorname{int}\operatorname{dom} h$ then our claim is verified, so suppose that x^* is in $\operatorname{bdry}\operatorname{dom} h$. If c is in $\operatorname{dom} h$ but not in $\operatorname{int}\operatorname{dom} h$, then, applying B2 above, we conclude that $x^* \in \operatorname{dom} h$ and $\{D_h(x^*, x^k)\} \to 0$. If, on the other hand, c is in $\operatorname{int}\operatorname{dom} h$ then by R2 x^* would have to be in $\operatorname{int}\operatorname{dom} h$ also. It follows that $x^* \in \operatorname{dom} h$ and $\{D_h(x^*, x^k)\} \to 0$. Now we show that x^* is in C.

Label $x^* = x_0^*$. Since there must be at least one index i that occurs infinitely often as $i(k)$, we assume, without loss of generality, that the subsequence $\{x^{k_n}\}$ has been selected so that $i(k) = 1$ for all $n = 1, 2, \ldots$. Passing to subsequences as needed, we assume that, for each $m = 0, 1, 2, \ldots, I-1$, the subsequence $\{x^{k_n+m}\}$ converges to a cluster point x_m^*, which is in $\operatorname{dom} h$, according to the same argument we used in the previous paragraph. For each m the sequence $\{D_{f_m}(c, P_{C_m}^{f_m}(x^{k_n+m-1}))\}$ is bounded, so, again, by passing to subsequences as needed, we assume that the subsequence $\{P_{C_m}^{f_m}(x^{k_n+m-1})\}$ converges to $c_m^* \in C_m \cap \overline{\operatorname{dom} f_m}$.

Since the sequence $\{D_{f_m}(c, P_{C_m}^{f_m}(x^{k_n+m-1}))\}$ is bounded and $c \in \operatorname{dom} f_m$, it follows, from either B2 or R2, that $c_m^* \in \operatorname{dom} f_m$. We know that

$$\{D_{f_m}(P_{C_m}^{f_m}(x^{k_n+m-1}), x^{k_n+m-1})\} \to 0 \tag{15}$$

and both $P_{C_m}^{f_m}(x^{k_n+m-1})$ and x^{k_n+m-1} are in $\operatorname{int}\operatorname{dom} f_m$. Applying R1, B3 or R3, depending on the assumed locations of c_m^* and x_{m-1}^*, we conclude that $c_m^* = x_{m-1}^*$.

We also know that

$$\{D_{f_m}(x^{k_n+m}, P_{C_m}^{f_m}(x^{k_n+m-1}))\} \to 0, \tag{16}$$

from which it follows, using the same arguments, that $x_m^* = c_m^*$. Therefore, we have $x^* = x_m^* = c_m^*$ for all m; so $x^* \in C$.

Since $x^* \in C \cap \operatorname{dom} h$, we may now use x^* in place of the generic c, to obtain that the sequence $\{D_h(x^*, x^k)\}$ is decreasing. However, we also know that the sequence $\{D_h(x^*, x^{k_n})\} \to 0$. So we have $\{D_h(x^*, x^k)\} \to 0$. Applying R5, we conclude that $\{x^k\} \to x^*$.

If the sets C_i are hyperplanes, then we get equality in Bregman's inequality (6)and so

$$D_h(c, x^k) - D_h(c, x^{k+1}) = G^k(x^{k+1}) + D_{f_{i(k)}}(P_{C_{i(k)}}^{f_{i(k)}}(x^k), x^k). \tag{17}$$

Since the right side of this equation is independent of which c we have chosen in the set $C \cap \operatorname{dom} h$, the left side is also independent of this choice. This implies that

$$D_h(c, x^0) - D_h(c, x^M) = D_h(x^*, x^0) - D_h(x^*, x^M), \tag{18}$$

for any positive integer M and any $c \in C \cap \operatorname{dom} h$. Therefore

$$D_h(c, x^0) - D_h(x^*, x^0) = D_h(c, x^M) - D_h(x^*, x^M). \tag{19}$$

Since $\{D_h(x^*, x^M)\} \to 0$ as $M \to +\infty$ and $\{D_h(c, x^M)\} \to \alpha \geq 0$, we have that $D_h(c, x^0) - D_h(x^*, x^0) \geq 0$. This completes the proof. ∎

4. AN INTERIOR POINT ALGORITHM FOR ITERATIVE OPTIMIZATION

We consider now an interior point algorithm (IPA) for iterative optimization. This algorithm was first presented in [6] and applied to transmission tomography in [13]. The IPA is suggested by a special case of the MSGP, involving functions h and $f := f_1$.

Assumptions:
We assume, for the remainder of this section, that h is a super-coercive Legendre function with essential domain $D = \operatorname{dom} h$. We also assume that f is continuous on the set $\overline{D}$, takes the value $+\infty$ outside this set and is differentiable in $\operatorname{int} \operatorname{dom} D$. Thus, f is a closed, proper convex function on R^J. We assume also that $\hat{x} = \operatorname{argmin}_{x \in \overline{D}} f(x)$ exists, but not that it is unique. As in the previous section, we assume that $D_h(x, z) \geq D_f(x, z)$ for all $x \in \operatorname{dom} h$ and $z \in \operatorname{int} \operatorname{dom} h$. As before, we denote by h^* the function conjugate to h.

The IPA:
The IPA is an iterative procedure that, under conditions to be described shortly, minimizes the function f over the closure of the essential domain of h, provided that such a minimizer exists.

Algorithm 4.1 *Let x^0 be chosen arbitrarily in* $\operatorname{int} D$. *For $k = 0, 1, \ldots$ let x^{k+1} be the unique solution of the equation*

$$\nabla h(x^{k+1}) = \nabla h(x^k) - \nabla f(x^k). \tag{20}$$

Note that equation (20) can also be written as

$$x^{k+1} = \nabla h^*(\nabla h(x^k) - \nabla f(x^k)). \tag{21}$$

Motivating the IPA:
As already noted, the IPA was originally suggested by consideration of a special case of the MSGP. Suppose that $\overline{x} \in \operatorname{dom} h$ is the unique global minimizer of the function f, and that $\nabla f(\overline{x}) = 0$. Take $I = 1$ and $C = C_1 = \{\overline{x}\}$. Then $P^f_{C_1}(x^k) = \overline{x}$ always and the iterative MSGP step becomes that of the IPA. Since we are assuming that $\overline{x}$ is in dom h, the convergence theorem for the MSGP tells us that the iterative sequence $\{x^k\}$ converges to $\overline{x}$.

In most cases, the global minimizer of f will not lie within the essential domain of the function h and we are interested in the minimum value of f on the set $\overline{D}$, where $D = \operatorname{dom} h$; that is, we want $\hat{x} = \operatorname{argmin}_{x \in \overline{D}} f(x)$, whenever such a minimum exists. As we shall see, the IPA can be used to advantage even when the specific conditions of the MSGP do not hold.

Preliminary results for the IPA:

Two aspects of the IPA suggest strongly that it may converge under more general conditions than those required for convergence of the MSGP. The sequence $\{x^k\}$ defined by (20) is entirely within the interior of $\operatorname{dom} h$. In addition, as we now show, the sequence $\{f(x^k)\}$ is decreasing. Adding both sides of the inequalities $D_h(x^{k+1}, x^k) - D_f(x^{k+1}, x^k) \geq 0$ and $D_h(x^k, x^{k+1}) - D_f(x^k, x^{k+1}) \geq 0$ gives

$$\langle \nabla h(x^k) - \nabla h(x^{k+1}) - \nabla f(x^k) + \nabla f(x^{k+1}), x^k - x^{k+1} \rangle \geq 0. \tag{22}$$

Substituting according to equation (20) and using the convexity of the function f, we obtain

$$f(x^k) - f(x^{k+1}) \geq \langle \nabla f(x^{k+1}), x^k - x^{k+1} \rangle \geq 0. \tag{23}$$

Therefore, the sequence $\{f(x^k)\}$ is decreasing; since it is bounded below by $f(\hat{x})$, it has a limit, $\hat{f} \geq f(\hat{x})$. We have the following result (see [6], Prop. 3.1).

Lemma 4.1 $\hat{f} = f(\hat{x})$.

Proof: Suppose, to the contrary, that $0 < \delta = \hat{f} - f(\hat{x})$. Select $z \in D$ with $f(z) \leq f(\hat{x}) + \delta/2$. Then $f(x^k) - f(z) \geq \delta/2$ for all k. Writing $H_k = D_h(z, x^k) - D_f(z, x^k)$ for each k, we have

$$H_k - H_{k+1} = D_h(x^{k+1}, x^k) - D_f(x^{k+1}, x^k) + \langle \nabla f(x^{k+1}), x^{k+1} - z \rangle . \tag{24}$$

Since $\langle \nabla f(x^{k+1}), x^{k+1} - z \rangle \geq f(x^{k+1}) - f(z) \geq \delta/2 > 0$ and $D_h(x^{k+1}, x^k) - D_f(x^{k+1}, x^k) \geq 0$, it follows that $\{H_k\}$ is a decreasing sequence of positive numbers, so that the successive differences converge to zero. This is a contradiction; we conclude that $\hat{f} = f(\hat{x})$. ∎

Convergence of the IPA:

We prove the following convergence result for the IPA (see also [6]).

Theorem 4.1 *If* $\hat{x} = \operatorname{argmin}_{x \in \overline{D}} f(x)$ *is unique, then the sequence* $\{x^k\}$ *generated by the IPA according to equation (20) converges to* $\hat{x}$. *If* $\hat{x}$ *is not unique, but can be chosen in* D, *then the sequence* $\{D_h(\hat{x}, x^k)\}$ *is decreasing. If, in addition, the function* $D_h(\hat{x}, \cdot)$ *has bounded level sets, then the sequence* $\{x^k\}$ *is bounded and so has cluster points* $x^* \in \overline{D}$ *with* $f(x^*) = f(\hat{x})$. *Finally, if* h *is a Bregman-Legendre function, then* $x^* \in D$ *and the sequence* $\{x^k\}$ *converges to* x^*.

Proof: According to Corollary 8.7.1 of [14], if G is a closed, proper convex function on R^J and if the level set $L_\alpha = \{x | G(x) \leq \alpha\}$ is nonempty and bounded for at least one value of α, then L_α is bounded for all values of α. If the constrained minimizer $\hat{x}$ is unique, then, by the continuity of f on $\overline{D}$ and Rockafellar's corollary, we can conclude that the

If $M = N$ then $G = (AA^T)^{-1}$ is a suitable choice; since A is invertible, we then have $P^G_{A(K)}(z) = AP_K(A^{-1}(z))$ for all z in R^M. Our iterative algorithms then involve P_Q and $P^G_{A(K)}$.

If $M < N$ there is no such matrix G. Moreover, the convex set $A(K)$ need not be closed. To obtain iterative algorithms for this case we augment A using a full-rank $N - M$ by N matrix B to obtain the N by N invertible matrix $R = [A^T\, B^T]^T$. Although it is not necessary, we shall assume that B has been chosen so that $AB^T = 0$.

If $M > N$ the choice of $G = A(A^TA)^{-2}A^T$ gives $A^TGA = I$. With $A^+ := (A^TA)^{-1}A^T$ and $A(R^N)$ denoting the range of A, we have $A(K) = (A^+)^{-1}(K) \cap A(R^N)$, so the set $A(K)$ is closed. Proceeding as above, we have $P^G_{A(K)}(Ax) = AP_K(x)$; for those z in R^M that are not in the range of A, we have $P^G_{A(K)}(z) = AP_K(A^TA)^{-1}A^Tz$. Because the function $g(z) = z^TGz$ is not a Legendre function on R^M, we do not consider $P^g_{A(K)}(z)$, a Bregman projection onto the convex set $A(K)$. We replace G with $H = G + U$, where U is a nonnegative-definite symmetric matrix such that H is positive-definite and $A^TU = 0$. The function $f(z) = z^THz$ is a Legendre function on R^M and the associated Bregman projection of z onto $A(K)$ is $P^f_{A(K)}(z) = P^H_{A(K)}(z) = AP_K(A^TA)^{-1}A^Tz$ once again.

When we employ P_Q and $P^H_{A(K)}$ in the simultaneous algorithm of Censor and Elfving [8], we find that the algorithm involves the choice of the matrix U. Here we illustrate the use of the MSGP algorithm and find that we obtain an iterative solution to the SFP, which, for the case in which $M > N$, does not depend on the choice of the matrix U.

Case 1: $M = N$

Now A is an N by N matrix with rank equal to N. In order to apply the MSGP algorithm we define $C_1 = A^{-1}(Q) = \{x|\, Ax \in Q\}$, $C_2 = K$, $f_1(x) = \gamma x^TA^TAx$, and $h(x) = f_2(x) = x^Tx$; here γ is a positive constant chosen so that

$$D_h(x,z) - D_{f_1}(x,z) = (x-z)^T(I - \gamma A^TA)(x-z) \geq 0, \tag{25}$$

for all x and z. Therefore, $\gamma \leq 1/\lambda_{max}(A^TA)$, where $\lambda_{max}(A^TA)$ denotes the largest eigenvalue of the matrix A^TA. Let us normalize A so that each of its rows has Euclidean norm equal to one. Then, since trace (A^TA) =trace$(AA^T) = M$, it follows that we can choose $\gamma \leq 1/M$. The corresponding Bregman projections are

$$P^{f_1}_{C_1}(x) = (A^TA)^{-1}A^TP_Q(Ax) \tag{26}$$

and

$$P^{f_2}_{C_2}(x) = P_K(x), \tag{27}$$

where P_Q and P_K denote the orthogonal projections onto Q and K, respectively.

Applying the MSGP algorithm, we obtain the following iteration step. For $k = 0, 1, ...,$ having obtained x^k, let

$$x^{k+1} = P_K(I + \gamma A^T(P_Q - I)A)x^k. \tag{28}$$

Then the sequence $\{x^k\}$ given by (28) converges to an $x \in K$ for which $Ax \in Q$.

Case 2: $M < N$

In this case we augment the matrix A, as discussed above, to obtain the invertible square matrix R. This puts us in Case 1, where the iterative algorithm for that case gives the following. For $k = 0, 1, ...,$ having obtained x^k, we have

$$x^{k+1} = P_K(I + \gamma R^T(P_V - I)R)x^k, \tag{29}$$

where γ is now selected so that the matrix $I - \gamma R^T R$ is nonnegative definite. Recalling that $AB^T = 0$, we rewrite (29) in terms of the matrices A and B:

$$x^{k+1} = P_K(x^k + \gamma(A^T(P_Q - I)A + B^T(I - I)B)x^k) \tag{30}$$

so that

$$x^{k+1} = P_K(x^k + \gamma(A^T(P_Q - I)A)x^k). \tag{31}$$

This is the same algorithm as in the previous case, except that here the γ must be chosen so that $I - \gamma(A^T A + B^T B)$ is nonnegative definite. However, since there is no lower limit on the maximum eigenvalue of $B^T B$, we may conclude that the iteration in (28) applies for all $M \leq N$. This is reasonable, since there is no inversion involved. Notice that we have not needed that $A(K)$ be closed.

For example, suppose we wish to solve the consistent system of linear equations represented by the matrix equation $Ax = b$, with $M \leq N$. If we have no constraints that we wish to impose on the solution, we let $K = R^N$; let $Q = \{b\}$. Then (28) becomes

$$x^{k+1} = x^k + \gamma A^T(b - Ax^k), \tag{32}$$

which is sometimes called the Landweber algorithm [11], known to converge to a solution of $Ax = b$ whenever γ is chosen so that $I - \gamma A^T A$ is nonnegative definite. If we wish to impose the constraint that the solution x be a nonnegative vector, then we take K to be the nonnegative orthant in R^N. The resulting algorithm is a nonnegatively clipped version of the iteration in (32).

Case 3: $M > N$

For simplicity, we formulate the SFP as a CFP involving the three closed convex sets $C_1 = A(R^N)$, $C_2 = A(K)$ and $C_3 = Q$ in R^M. We let $h(x) = f_1(x) = f_3(x) = x^T x/2$ and $f_2(x) = \gamma x^T Hx/2$, where, as above, H is the positive-definite matrix $H = G + U$, built from $G = A(A^T A)^{-2} A^T$ and the nonnegative-definite matrix U and γ chosen so that $I - \gamma H$ is positive-definite. Recall that $A^T U = 0$. Then we have $P_1 := P_{C_1}^{f_1}$ the orthogonal projection onto the range of A in R^M, $P_2 := P_{C_2}^{f_2}$ with $P_2(Ax) = P_{A(K)}^H(Ax) = AP_K(x)$ and $P_3 := P_{C_3}^{f_3} = P_Q$. We now apply the MSGP algorithm.

Beginning with an arbitrary z^0 in R^M, we take

$$z^1 = P_1(z^0) = A(A^T A)^{-1} A^T z^0 = Au^1, \tag{33}$$

where

$$u^1 = (A^T A)^{-1} A^T z^0. \tag{34}$$

The next step gives z^2, which minimizes the function

$$(h - f_2)(z - z^1) + f_2(z - P_2(z^1)). \tag{35}$$

Therefore, we have

$$0 = (I - \gamma H)(z - Au^1) + \gamma H(z - AP_K(u^1)), \tag{36}$$

so that

$$z^2 = A((I - \gamma(A^TA)^{-1}(I - P_K))((A^TA)^{-1}A^Tz^0). \tag{37}$$

Finally,

$$z^3 = P_Q(z^2). \tag{38}$$

Writing the iterative algorithm in terms of completed cycles, we have $w^0 = z^0$ and

$$w^{k+1} = P_QA(I + \gamma(A^TA)^{-1}(P_K - I))(A^TA)^{-1}A^Tw^k. \tag{39}$$

The iterative step for $x^k := (A^TA)^{-1}A^Tw^k$ is then

$$x^{k+1} = (A^TA)^{-1}A^TP_QA(I + \gamma(A^TA)^{-1}(P_K - I))x^k. \tag{40}$$

We must select γ so that $I - \gamma H$ is positive-definite. Because there is no lower limit to the maximum eigenvalue of U, it follows that γ must be chosen so that $I - \gamma G$ is positive-definite. Since $G = A(A^TA)^{-2}A^T$ we have $Gz = \lambda z$ implies that $\lambda A^Tz = A^TGz = (A^TA)^{-1}A^Tz$, so that the nonzero eigenvalues of G are those of $(A^TA)^{-1}$. It follows that we must select γ not greater than the smallest eigenvalue of A^TA.

SUMMARY

In this paper we have considered the iterative method of successive generalized projections onto convex sets for solving the convex feasibility problem. The generalized projections are derived from Bregman-Legendre functions. In particular, we have extended Bregman's method to permit the generalized projections used at each step to be taken with respect generalized distances that vary with the convex set. Merely replacing the distance $D_f(x, z)$ in Bregman's method with distances $D_{f_i}(x, z)$ is not enough; counterexamples show that such a simple extension may not converge. We show that a convergent algorithm, the MSGP, can be obtained through the use of a dominating Bregman-Legendre distance, that is $D_h(x, z) \geq D_{f_i}(x, z)$, for all i, and a form of relaxation based on the notion of generalized convex combination.

Particular problems are solved through the selection of appropriate functions h and f_i. The MSGP algorithm can be used to solve the split feasibility problem. Iterative interior point optimization algorithms can also be based on the MSGP approach.

REFERENCES

1. H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Review* **38** (1996) 367–426.
2. H. H. Bauschke and J. M. Borwein, Legendre functions and the method of random Bregman projections, *J. of Convex Analysis* **4** (1997) 27–67.
3. L. M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, *USSR Computational Mathematics and Mathematical Physics* **7** (1967) 200–217.
4. C. L. Byrne, Block-iterative methods for image reconstruction from projections, *IEEE Transactions on Image Processing* **IP-5** (1996) 792–794.
5. C. L. Byrne, Iterative projection onto convex sets using multiple Bregman distances. *Inverse Problems* **15** (1999) 1295–1313.
6. C. L. Byrne, Block-iterative Interior point optimization methods for image reconstruction from limited data, *Inverse Problems*, **16** (2000) 1405–1419.
7. C. L. Byrne and Y. Censor, Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization, *Annals of Operations Research* accepted for publication.
8. Y. Censor and T. Elfving, A multiprojection algorithm using Bregman projections in a product space, *Numerical Algorithms*, **8** (1994) 221–239.
9. Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms and Applications* (Oxford University Press, New York, 1997).
10. T. Kotzer, N. Cohen and J. Shamir, A projection-based algorithm for consistent and inconsistent constraints, *SIAM J. Optim.* **7(2)** (1997) pp. 527–546.
11. L. Landweber, An iterative formula for Fredholm integral equations of the first kind, *Amer. J. of Math.* **73** (1951) pp. 615–624.
12. K. Lange, M. Bahn and R. Little, A theoretical study of some maximum likelihood algorithms for emission and transmission tomography, *IEEE Trans. Med. Imag.* **6** (1987) 106–114.
13. M. Narayanan, C. L. Byrne and M. King, An interior point iterative reconstruction algorithm incorporating upper and lower bounds, with application to SPECT transmission imaging, *IEEE Trans. Medical Imaging*, submitted for publication.
14. R. T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, New Jersey, 1970).
15. D.C. Youla, Mathematical theory of image restoration by the method of convex projections, *in* Stark, H. (Editor) *Image Recovery: Theory and Applications*, (Academic Press, Orlando, Florida, USA, 1987), pp. 29–78.