Inverse Problems **16** (2000) 1405–1419. Printed in the UK PII: S0266-5611(00)11419-4

Block-iterative interior point optimization methods for image reconstruction from limited data

Charles Byrne

Department of Mathematical Sciences, University of Massachusetts at Lowell, Lowell, MA 01854, USA

E-mail: Charles_Byrne@uml.edu

Received 27 January 2000, in final form 31 August 2000

Abstract. Iterative algorithms for image reconstruction often involve minimizing some cost function $h(x)$ that measures the degree of agreement between the measured data and a theoretical parametrized model. In addition, one may wish to have x satisfy certain constraints. It is usually the case that the cost function is the sum of simpler functions:

$$h(x) = \sum_{i=1}^{I} h_i(x).$$

Partitioning the set $\{i = 1, \ldots, I\}$ as the union of the disjoint sets B_n, $n = 1, \ldots, N$, we let

$$h^n(x) = \sum_{i \in B_n} h_i(x).$$

The method presented here is block iterative, in the sense that at each step only the gradient of a single $h^n(x)$ is employed. Convergence can be significantly accelerated, compared to that of the single-block ($N = 1$) method, through the use of appropriately chosen scaling factors.

The algorithm is an interior point method, in the sense that the images x^{k+1} obtained at each step of the iteration satisfy the desired constraints. Here the constraints are imposed by having the next iterate x^{k+1} satisfy the gradient equation

$$\nabla F(x^{k+1}) = \nabla F(x^k) - t_n \nabla h^n(x^k),$$

for appropriate scalars t_n, where the convex function F is defined and differentiable only on vectors satisfying the constraints.

Special cases of the algorithm that apply to tomographic image reconstruction, and permit inclusion of upper and lower bounds on individual pixels, are presented. The focus here is on the development of the underlying convergence theory of the algorithm. Behaviour of special cases has been considered elsewhere.

1. Introduction

Inverse problems commonly involve the estimation or reconstruction of a mathematical object x from partial information about that object. The object x is usually a vector or a function that can be considered as an 'image' and the problem is to reconstruct that image from both limited measurement data and prior information. In this paper x will be a vector in the real J-dimensional space R^J. If we can formulate the problem as a *convex feasibility problem* (CFP), the constraints on x are interpreted as saying that x is a member of certain convex subsets of R^J; the objective is then to find a member of the intersection of these sets. The 'projection onto convex sets' (POCS) method [18] is an iterative procedure for solving the CFP. A variant of POCS, called the 'multiple-distance successive generalized projection' (MSGP) method [7, 8], involves generalized projections based on a family of distances.

0266-5611/00/051405+15$30.00

In the typical case in which the measured data are noisy, there may be no x satisfying both the data constraints and those based on prior information about acceptable reconstructions. The convex sets may, therefore, have empty intersection and the CFP no solution. In these circumstances, we may seek an x that minimizes some cost function that measures the average distance from x to each of the convex sets; such functions are often called *proximity functions*. More generally, we may seek an x that optimizes a certain function constructed from the data and the prior information. In [6] we presented the iterative 'AB methods' for reconstructing images that optimized certain functions and imposed prior upper and lower bounds on the values of the individual pixels.

In this paper we extend the AB methods in [6] to permit greater choice in the selection of the functions to be optimized and to remove limitations on the data values that limited the usefulness of the AB methods.

The algorithm we present here minimizes a convex function $h(x)$, subject to the restriction that x lie within the set on which a second convex function $F(x)$ takes finite values. The function h will typically incorporate the measured data in some way, while the function F will be selected so that its *essential domain* $\mathrm{dom} F$, the set of all x at which $F(x)$ is finite, contains only those x satisfying the prior constraints. Our method is an *interior point algorithm* (IPA) in that at each step of the iteration the current vector, x^k, will be within $\mathrm{int}\,\mathrm{dom} F$, the interior of the essential domain of F.

Image reconstruction methods based on the minimization of some cost function typically seek to minimize functions of the form

$$h(x) = \sum_{i=1}^{I} h_i(x), \tag{1.1}$$

where $h_i(x)$ measures the extent to which the (vectorized) image x is consistent with one of the measured data values. For example, suppose that the data pertaining to x are the linear functional measurements $\langle a^i, x\rangle = b_i$, for $i = 1, \ldots, I$. Let A be the I by J matrix whose ith row is the transpose of the column vector a^i; then $Ax_i = (Ax)_i = \langle a^i, x\rangle$, for $i = 1, \ldots, I$. We may then want to minimize $h(x) = \|Ax - b\|^2$, over those vectors x satisfying certain constraints. With $h_i(x) = |\langle a^i, x\rangle - b_i|^2$ we have $h(x) = \sum_{i=1}^{I} h_i(x)$. Because I, the number of measurements, is usually quite large, minimization methods that use the gradient of h at each step must calculate a large sum,

$$\nabla h(x^k) = \sum_{i=1}^{I} \nabla h_i(x^k) \tag{1.2}$$

for each $k = 0, 1, 2, \ldots$. The conjugate gradient method can be somewhat faster [17], but is still too slow for the sort of problem involving large data sets that one typically encounters in image reconstruction; unless the matrices involved are highly structured, determining suitable preconditioners is difficult. To reduce the computational load one may employ *block-iterative* algorithms [5] (also called *incremental* algorithms in the optimization literature [2]), in which the sum over $i = 1, \ldots, I$ in (1.2) is replaced by a sum over i in some predetermined subset that varies with each step. Recent examination of such methods has found that they can often produce useful reconstructed images in an order of magnitude fewer iterations. We shall consider algorithm acceleration by means of such block-iterative methods.

We assume that the set $\{i = 1, \ldots, I\}$ is partitioned into N disjoint subsets, $\{B_n, n = 1, \ldots, N\}$. Denote by $h^n(x)$ the sum

$$h^n(x) = \sum_{i \in B_n} h_i(x), \tag{1.3}$$

so that $h(x) = \sum_{n=1}^{N} h^n(x)$. Denote by $\sum^n$ the sum over those i in B_n.

Our IPA is the following.

Algorithm 1.1 (The IPA). *For $k = 0, 1, \ldots,$ and $n = k(\mathrm{mod}N) + 1$ and having determined x^k, let x^{k+1} be the unique solution of the gradient equation*

$$\nabla F(x^{k+1}) = \nabla F(x^k) - t_n \sum^n \nabla h_i(x^k), \tag{1.4}$$

where $t_n > 0$ is chosen so that $f^n(x) = F(x) - t_n \sum^n h_i(x)$ is convex.

To provide a solid theoretical framework within which to prove convergence theorems, we shall require, at certain points of the discussion, that a function be a *Bregman–Legendre function*, in the sense of Bauschke and Borwein [1]. Details concerning Bregman–Legendre functions are to be found in the appendix.

Throughout this paper we shall employ the notational conventions that $Ax = b$ denote a general system of (real) linear equations, normalized so that the rows of the I by J matrix A have unit norm, and that $Px = y$ denote a system of linear equations for which the entries of the vector y are positive, the entries of the I by J matrix P are non-negative and the columns of P each sum to one.

Before we present our theorem concerning the IPA, let us consider two examples.

2. Two examples

In this section we consider two examples of the IPA.

2.1. The algebraic reconstruction technique

As an example, let us consider the *algebraic reconstruction technique* (ART), due to Gordon *et al* [11]. In this case the function we wish to minimize is $h(x) = \frac{1}{2}\|Ax - b\|^2$. Wishing to place no constraints on the acceptable x, we select as the second function $F(x) = \frac{1}{2}\|x\|^2$. We assume that the matrix A has been normalized so that the Euclidean norm of each of its rows is unity. To obtain the ART we take $N = I$ and $h_i(x) = \frac{1}{2}((Ax)_i - b_i)^2$. Note that the functions $f_i(x) = F(x) - h_i(x)$ are convex since $\|x - z\|^2 \geqslant |(Ax)_i - (Az)_i|^2$, for all vectors x and z.

The ART. For $k = 0, 1, \ldots$ and $i = k(\mathrm{mod}\, I) + 1$ set

$$x_j^{k+1} = x_j^k + A_{ij}(b_i - (Ax^k)_i). \tag{2.1}$$

A simultaneous version of the ART (SART) is the following.

The SART. For $k = 0, 1, \ldots$ let

$$x_j^{k+1} = x_j^k + t \sum_{i=1}^{I} A_{ij}(b_i - (Ax^k)_i), \tag{2.2}$$

where $t > 0$ is chosen so that $I - tA^TA$ is a positive-definite matrix. Block-iterative variations are also possible, in which the sum in the iterative step is taken only over a subset of the set $\{i = 1, \ldots, I\}$.

When there are solutions of the system $Ax = b$ then both ART and SART converge to the solution of $Ax = b$ closest to the initial vector x^0, according to the Euclidean distance. When there are no solutions of $Ax = b$ SART converges to the least-squares solution closest to x^0, while ART fails to converge. Instead, for each fixed i, as $m \to +\infty$, the ART subsequences $\{x^{mI+i}\}$ converge to distinct vectors $x^{\infty,i}$; we call this set of vectors the *limit cycle* (LC). The greater the minimum value of $\|Ax - b\|^2$ the more the vectors of the LC are distinct from one another.

There are several ways to avoid the LC in ART and to obtain the least-squares solution. One way, which does not easily generalize to other algorithms, is what we call *double ART* (DART).

The DART. In step 1 apply the ART algorithm to the consistent system of linear equations $A^T w = 0$, beginning with $w^0 = b$. The limit is w^∞, the member of the null space of A^T closest to b. In step 2, apply ART to the consistent system of linear equations $Ax = b - w^\infty$. The limit is then the least-squares solution of $Ax = b$. Another method for avoiding the LC is *strong underrelaxation* [9].

Strongly underrelaxed ART. Let $t > 0$. Replace the iterative step in ART with

$$x_j^{k+1} = x_j^k + tA_{ij}(b_i - (Ax^k)_i). \tag{2.3}$$

As $t \to 0$, the vectors of the LC approach the least-squares solution.

In practical situations, one may use only a few iterations of an algorithm and be less concerned with the limiting vector (or vectors) than with the behaviour of the iterates for small values of k. When the minimum value of $\|Ax - b\|^2$ is not too large (that is, the measured data are not too noisy), the ART has been shown to provide usable reconstructions with very few iterations, particularly when the equations are carefully ordered and some underrelaxation is used [13]. In contrast, the SART is typically quite slow to converge.

2.2. The multiplicative algebraic reconstruction technique (MART)

As a second example, we consider the MART, also due to Gordon *et al* [11].

The function to be minimized is now $h(x) = KL(Px, y)$; here $KL(x, z)$ is the Kullback–Leibler (or cross-entropy) distance, defined for non-negative vectors x and z by

$$KL(x, z) = \sum_{j=1}^{J} KL(x_j, z_j), \tag{2.4}$$

where $KL(a, b) = a \log \frac{a}{b} + b - a$, $KL(0, b) = b$ and $KL(a, 0) = +\infty$ for positive scalars a and b.

To impose non-negativity of the entries of the vector x we use the function $F(x) = E(x)$ where $E(x)$ is defined for all x with non-negative entries as

$$E(x) = \sum_{j=1}^{J} x_j \log x_j - x_j. \tag{2.5}$$

If, for each j, we have $\alpha_j < \beta_j$, then we can enforce the constraints that x be such that $\alpha_j \leqslant x_j \leqslant \beta_j$ by taking $F(x) = F_{\alpha\beta}(x)$, defined by

$$F_{\alpha\beta}(x) = \sum_{j=1}^{J} (x_j - \alpha_j)\log(x_j - \alpha_j) + (\beta_j - x_j)\log(\beta_j - x_j). \tag{2.6}$$

The MART. The MART [11] begins with a strictly positive vector x^0 and has the iterative step

$$x_j^{k+1} = x_j^k \left(\frac{y_i}{(Px^k)_i} \right)^{P_{ij}}, \tag{2.7}$$

for $j = 1, 2, \ldots, J$ and $i = k(\text{mod } I) + 1$.

The SMART. The *simultaneous* MART (SMART) [4] begins with a strictly positive vector x^0 and has the iterative step

$$x_j^{k+1} = x_j^k \prod_{i=1}^{I} \left(\frac{y_i}{(Px^k)_i} \right)^{P_{ij}}, \tag{2.8}$$

for $j = 1, 2, \ldots, J$.

In the consistent case, that is, when there are vectors $x \geqslant 0$ with $y = Px$ then both MART and SMART converge to the non-negative solution that minimizes $KL(x, x^0)$. When there are no such non-negative vectors, the SMART converges to the unique non-negative minimizer of $KL(Px, y)$ for which $KL(x, x^0)$ is minimized. The MART, however, fails to converge. What is observed always (but for which no proof exists) is that, for each fixed $i = 1, 2, \ldots, I$, as $m \to +\infty$, the subsequences $\{x^{mI+i}\}$ converge to separate limit vectors, say $x^{\infty,i}$. This LC $= \{x^{\infty,i} | i = 1, \ldots, I\}$ reduces to a single vector whenever there is a non-negative solution of $y = Px$. The greater the minimum value of $KL(Px, y)$ the more distinct from one another the vectors of the LC are.

The MART will converge, in the consistent case, provided that $0 \leqslant P_{ij} \leqslant 1$, for all i and j; this condition holds here since we have assumed that the columns of P sum to one. Since I is typically quite large, the P_{ij} are likely to be a great deal smaller than one. We can accelerate the convergence of MART by rescaling the equations, obtaining what we have called the REMART.

The REMART. The *rescaled multiplicative algebraic reconstruction technique* (REMART) [5] begins with a strictly positive vector x^0 and has the iterative step

$$x_j^{k+1} = x_j^k \left(\frac{y_i}{(Px^k)_i} \right)^{m_i^{-1} P_{ij}}, \tag{2.9}$$

for $j = 1, 2, \ldots, J$ and $i = k(\text{mod}\, I) + 1$, with $m_i = \max\{P_{ij} | j = 1, \ldots, J\}$.

With $h(x) = KL(Px, y)$, $h_i(x) = KL((Px)_i, y_i)$ and $F(x) = E(x)$ it is clear that x^{k+1} in (2.9) is the solution of the gradient equation

$$\nabla F(x^{k+1}) = \nabla F(x^k) - m_i^{-1} \nabla h_i(x^k). \tag{2.10}$$

Although the importance of the rescaling for accelerating MART was not remarked upon in earlier papers on MART, the rescaling was often a part of actual implementations [12]. We see that MART and REMART converge whenever there is a common non-negative minimizer of the functions $h_i(x)$, $i = 1, \ldots, I$. When there is no such vector, we obtain an LC.

For each i the function $f_i(x) = F(x) = h_i(x) = E(x) - m_i^{-1} KL((Px)_i - y_i)$ is convex; this follows from the inequality $KL(x, z) \geqslant KL(x_+, z_+)$, where $x_+ = \sum_{j=1}^{J} x_j$; in fact, the value $t_i = m_i^{-1}$ is the largest for which the function $E(x) - t_i KL((Px)_i - y_i)$ is convex. As we shall see in the proof of convergence of the IPA, larger values of t_i (or of t_n) lead to faster convergence.

When applied in the case of a single block, the convergence theorem for the IPA is somewhat stronger than that for the general case of multiple blocks. We therefore consider the two cases separately.

3. Convergence of the IPA: the case of $N = 1$

Let D be a convex set in R^J with nonempty interior, int$D$. Let $h(x)$ be differentiable on int$D$ and convex and continuous on the closure of D, $\overline{D}$. We want to minimize h over $\overline{D}$, if such minimizers exist. Since we are only interested in the behaviour of h on the set $\overline{D}$, we assume that h takes the value $+\infty$ outside this set. Then h is a closed, proper convex function on R^J, as defined by Rockafellar [16].

Let f be a function that is differentiable on $\operatorname{int}D$ and convex and continuous on D. Let $F(x) = f(x) + h(x)$. The IPA for the case of a single block is the following.

Algorithm 3.1 (The IPA). *Let x^0 be arbitrary in* $\operatorname{int}D$. *Having calculated x^k in* $\operatorname{int}D$ *we solve*

$$\nabla F(x^{k+1}) = \nabla F(x^k) - \nabla h(x^k) \tag{3.1}$$

for x^{k+1} in $\operatorname{int}D$.

In what follows we shall assume that the gradient equation (3.1) can be solved for x^{k+1} in $\operatorname{int}D$ at every step. Let $\overline{h} = \inf\{h(x)|x \in D\} \geqslant -\infty$, which is then the infimum of h on all of R^J.

The convexity of the functions h and F–h forces the sequence $\{h(x^k)\}$ to be decreasing. Rewriting (3.1) using $f(x) = F(x) - h(x)$ we have

$$\nabla f(x^k) - \nabla f(x^{k+1}) = \nabla h(x^{k+1}). \tag{3.2}$$

Therefore

$$0 \leqslant \langle \nabla f(x^k) - \nabla f(x^{k+1}), x^k - x^{k+1}\rangle = \langle \nabla h(x^{k+1}), x^k - x^{k+1}\rangle \leqslant h(x^k) - h(x^{k+1}). \tag{3.3}$$

Therefore, we have $h(x^k) \to \hat{h}$, for some $\hat{h} \geqslant \overline{h}$. In fact, we can show the following.

Proposition 3.1. $\hat{h} = \overline{h}$.

Proof. Suppose not; let $\hat{h} = \overline{h} + \delta$, for some $\delta > 0$. Select $z^n \in D$ such that $h(z^n) \to \overline{h}$, as $n \to +\infty$; without loss of generality, we assume that $h(z^n) \leqslant \overline{h} + \delta/2$, for all n. Then $h(x^k) \geqslant h(z^n) + \delta/2$, for all k and all n. Let n be fixed. Then we have

$$D_f(z^n, x^k) - D_f(z^n, x^{k+1}) = D_f(x^{k+1}, x^k) + \langle \nabla h(x^{k+1}), x^{k+1} - z^n\rangle,$$

where D_f denotes the Bregman distance, as discussed in the appendix. Since, by the convexity of the function h, we have

$$\langle \nabla h(x^{k+1}), x^{k+1} - z^n\rangle \geqslant h(x^{k+1}) - h(z^n),$$

we know

$$D_f(z^n, x^k) - D_f(z^n, x^{k+1}) \geqslant D_f(x^{k+1}, x^k) + (h(x^{k+1}) - h(z^n)) > 0.$$

It follows that the sequence $\{D_f(z^n, x^k)\}$ is decreasing and, therefore, that the sequence $\{h(x^{k+1}) - h(z^n)\} \to 0$. But this is a contradiction, since $h(x^{k+1}) - h(z^n) \geqslant \delta/2$. We conclude that $\hat{h} = \overline{h}$. □

Whether or not the sequence $\{x^k\}$ converges will depend on other factors. We have the following useful result, which appears as corollary 8.7.1 in [16].

Proposition 3.2. *Let G be a closed proper convex function on R^J. If the level set $L_\alpha = \{x|G(x) \leqslant \alpha\}$ is nonempty and bounded for a single value of α, then it is bounded for every α.*

Our theorem is the following.

Theorem 3.1. *Let there be a point $\hat{x} \in \overline{D}$, with $h(\hat{x}) = \overline{h}$. If $\hat{x}$ is uniquely defined by these properties, then the sequence $\{x^k\}$ converges to $\hat{x}$. If $\hat{x}$ is not necessarily unique, but can be chosen in D, then the sequences $\{D_f(\hat{x}, x^k)\}$ and $\{D_F(\hat{x}, x^k)\}$ are decreasing. If, in addition, D_F has bounded level sets, that is, if property (B1) holds, then the sequence $\{x^k\}$ is bounded and $h(x^*) = h(\hat{x})$ for every cluster point $x^* \in \overline{D}$. If either f or F is a Bregman–Legendre function with essential domain D then x^* is in D and the sequence $\{x^k\}$ converges to x^*.*

Proof of the theorem. Let $\hat{x} \in \overline{D}$ satisfy $h(\hat{x}) = \overline{h}$. If $\hat{x}$ is uniquely defined by these properties, then, applying proposition 3.2 and using the fact that the sequence $\{h(x^k)\}$ is decreasing, we conclude that the sequence $\{x^k\}$ is bounded; let $x^* \in \overline{D}$ be any cluster point. By the continuity of h we have $h(x^*) = h(\hat{x})$ for all cluster points. Therefore $x^* = \hat{x}$ and the sequence $\{x^k\}$ converges to $\hat{x}$. Now consider the case in which $\hat{x}$ is not necessarily unique, but can be chosen in D.

Because h is convex, we know that, for any $z \in \operatorname{int}D$,

$$\langle \nabla h(z), z - \hat{x}\rangle \geqslant h(z) - h(\hat{x}) \geqslant 0. \tag{3.4}$$

Taking $\hat{x}$ in D, we have

$$D_f(\hat{x}, x^k) - D_f(\hat{x}, x^{k+1}) = D_f(x^{k+1}, x^k) + \langle \nabla h(x^{k+1}), x^{k+1} - \hat{x}\rangle \geqslant 0 \tag{3.5}$$

and

$$\begin{aligned} D_F(\hat{x}, x^k) - D_F(\hat{x}, x^{k+1}) &= D_f(x^{k+1}, x^k) + [D_h(x^{k+1}, x^k) + \langle \nabla h(x^k), x^{k+1} - \hat{x}\rangle] \\ &= D_f(x^{k+1}, x^k) + [h(x^{k+1}) - h(x^k) + \langle \nabla h(x^k), x^k - \hat{x}\rangle] \\ &\geqslant D_f(x^{k+1}, x^k) + [h(x^{k+1}) - h(\hat{x})] \geqslant 0. \end{aligned} \tag{3.6}$$

Consequently, we can conclude that the sequences $\{D_f(\hat{x}, x^k)\}$ and $\{D_F(\hat{x}, x^k)\}$ are decreasing and that the sequences $\{D_f(x^{k+1}, x^k)\}$ and $\{h(x^{k+1}) - h(\hat{x})\}$ are converging to zero. Then, assuming property (B1) of the appendix applied to D_F, the sequence $\{x^k\}$ is bounded and, by the continuity of h, we have $h(x^*) = \overline{h} = h(\hat{x})$ for each cluster point x^* of the sequence $\{x^k\}$. Now we assume that one of f or F is a Bregman–Legendre function.

For simplicity, we assume in what follows that f is Bregman–Legendre; the same argument holds if F is, instead. If $\hat{x}$ is not in the interior of the set D, we apply property (B2) of Bregman–Legendre functions to conclude that x^* is in D and that a sub-sequence of $\{D_f(x^*, x^k)\}$ converges to zero. It follows that the entire sequence $\{D_f(x^*, x^k)\}$ converges to zero, for every cluster point of the sequence $\{x^k\}$. On the other hand, if $\hat{x} \in \operatorname{int}D$, then by result (R2) of the appendix we know that $x^* \in \operatorname{int}D$, for all cluster points x^* of the sequence $\{x^k\}$. Since $h(x^*) = h(\hat{x})$ and $x^* \in D$, we can replace $\hat{x}$ with x^*, to obtain that the sequence $\{D_f(x^*, x^k)\}$ is decreasing. By result (R1), we know that a subsequence of $\{D_f(x^*, x^k)\}$ converges to zero; therefore $\{D_f(x^*, x^k)\} \to 0$. Using result (R5), we conclude that $\{x^k\} \to x^*$. The proof of the theorem is complete. □

4. Convergence of the IPA: the case of $N > 1$

In this section we consider the convergence of the IPA for the more general case of multiple blocks.

From our earlier examination of the two examples we expect that the IPA will converge to a single limiting vector only if the functions $h^n(x)$ have a common minimizer. In general, this is not likely, although, as we have seen, for functions such as $\|Ax - b\|^2$ and $KL(Px, y)$ it simply means that the systems of equations have feasible solutions.

Write $h(x) = \sum_{i=1}^{I} h_i(x)$ for convex functions $h_i(x)$ and $h^n(x) = \sum_{i \in B_n} h_i(x)$. Assume that $f^n = F(x) - t_n h^n(x)$ is differentiable on $\operatorname{int}D$ and convex and continuous on D.

Our theorem is the following.

Theorem 4.1. *Let there be a point $\hat{x} \in D$, that is a common minimizer of each of the functions $h^n(x)$. Then the sequence $\{D_F(\hat{x}, x^k)\}$ is decreasing. If, in addition, D_F has bounded level sets, that is, if property (B1) holds, then the sequence $\{x^k\}$ is bounded. If f^n is a Bregman–Legendre function, then $h^n(x^*) = h^n(\hat{x})$ for every cluster point $x^* \in \overline{D}$. If F is a Bregman–Legendre function with essential domain D then x^* is in D and the sequence $\{x^k\}$ converges to x^*.*

Proof of the theorem. Let $\hat{x} \in D$ be a common minimizer of the functions $h^n(x)$. Let $n = k(\text{mod}\, N) + 1$. We have

$$\begin{aligned} D_F(\hat{x}, x^k) - D_F(\hat{x}, x^{k+1}) &= D_F(x^{k+1}, x^k) + t_n\langle \nabla h^n(x^k), x^{k+1} - x^k\rangle \\ &= D_{f^n}(x^{k+1}, x^k) + t_n[h^n(x^{k+1}) - h^n(x^k) - \langle \nabla h^n(x^k), \hat{x} - x^k\rangle] \\ &\geqslant D_{f^n}(x^{k+1}, x^k) + t_n[h^n(x^{k+1}) - h^n(\hat{x})] \geqslant 0, \end{aligned}$$

since $\langle \nabla h^n(x^k), \hat{x} - x^k\rangle \leqslant h^n(\hat{x}) - h^n(x^k)$ by the convexity of the functions $h^n(x)$.

Consequently, we can conclude that the sequence $\{D_F(\hat{x}, x^k)\}$ is decreasing and that the sequences $\{D_{f^n}(x^{k+1}, x^k)\}$ and $\{h^n(x^{k+1}) - h^n(\hat{x})\}$ are converging to zero. Then, assuming property (B1) of the appendix, applied to D_F, the sequence $\{x^k\}$ is bounded.

Consider the bounded sequence $\{x^{mN+1} | m = 1, 2, \ldots\}$. Let $\{x^{m_r N+1} | r = 1, 2, \ldots\}$ be a subsequence converging to $x^{*,1}$. Then extract a subsequence of the sequence $\{x^{m_r N+2} | r = 1, 2, \ldots\}$ converging to $x^{*,2}$. Continuing in this manner, we obtain N subsequences $\{x^{m_t N+n} | t = 1, 2, \ldots\}$ of the original sequence with $\{x^{m_t N+n}\} \to x^{*,n}$ for each n. It follows that $h^{n+1}(x^{*,n}) = h^{n+1}(\hat{x})$ for each n. Since $\{D_{f^n}(x^{k+1}, x^k)\} \to 0$ for each n, it follows that $x^{*,n} = x^{*,n+1} = x^*$ for all n.

If $\hat{x}$ is not in $\text{int}\,\text{dom}F$, then, applying property (B2) of the appendix to the Bregman–Legendre function F, we find that x^* is in $\text{dom}F$. Using x^* in place of $\hat{x}$ now, we find that $D_F(x^*, x^k)$ is decreasing, but, because a subsequence converges to zero, we conclude that $D_F(x^*, x^k) \to 0$, which implies that $x^k \to x^*$.

If $\hat{x}$ is in $\text{int}\,\text{dom}F$ then x^* is also in $\text{int}\,\text{dom}F$, by result (R2) of the appendix, applied to the function F. By result (R1) $D_F(x^*, x^k) \to 0$ so that $x^k \to x^*$ by result (R5). □

We emphasize that we expect the simultaneous version of the IPA, for which $N = 1$, to converge slowly; this is what is always observed in special cases that have been considered so far. It is for this reason that we employ the block-iterative approach in the IPA. When $N > 1$ the inequality

$$D_F(\hat{x}, x^k) - D_F(\hat{x}, x^{k+1}) \geqslant D_{f^n}(x^{k+1}, x^k) + t_n[h^n(x^{k+1}) - h^n(\hat{x})] \geqslant 0,$$

derived in the proof just presented, suggests that larger values of t_n might lead to faster convergence of the algorithm; indeed this is what we observe in practice. This is precisely the difference between the MART and REMART algorithms, as discussed earlier. We select the scalars t_n as large as possible, subject to the constraints that the functions $f^n(x) = F(x) - t_n h^n(x)$ be Bregman–Legendre functions; in particular, they must be strictly convex on the interior of the domain of F.

5. An application of the IPA to transmission tomography image reconstruction

We consider now the application of the IPA to transmission tomography image reconstruction (see [14]). In this case the function $h(x)$ we wish to minimize is usually taken to be the regularized negative log-likelihood function associated with a Poisson model, given, to within an additive constant, by the Kullback–Leibler distance (see (2.4))

$$h(x) = \sum_i KL(y_i, c_i \exp(-(Lx)_i)). \tag{5.1}$$

Here y_i denotes the count associated with the ith line segment through the object, the entry L_{ij} of the matrix L is the length of the intersection of the ith line segment with the jth pixel and constant $c_i > 0$ denotes the input intensity of the radiation along the ith line segment. Let $h_i(x) = KL(y_i, c_i \exp(-(Lx)_i))$ and $h^n(x) = \sum^n h_i(x)$.

With the selection of the function $F(x)$ we impose desired constraints on the reconstructed image. In order to obtain reconstructed images $x = (x_1, \dots, x_J)^T$ with $0 \leqslant \alpha_j \leqslant x_j \leqslant \beta_j$, for $j = 1, \dots, J$, we shall use the Bregman–Legendre function $F(x) = F_{\alpha\beta}(x)$ defined in (2.6). The essential domain of F is the set $D = \{x | \alpha_j \leqslant x_j \leqslant \beta_j, j = 1, \dots, J\}$. Then, at the kth step, we solve

$$\nabla F(x^{k+1}) = \nabla F(x^k) - t_n \nabla h^n(x^k), \tag{5.2}$$

where $t_n > 0$ is to be chosen so that the function $f^n(x) = F(x) - t_n h^n(x)$ is convex. The algorithm we obtain is the following: having calculated x^k, let x^{k+1} be determined by (5.2). Then we have the following.

Algorithm 5.1. *Let $\alpha < x^0 < \beta$ be chosen. For $k = 0, 1, 2, \dots$ and $n = k(\mathrm{mod} N) + 1$, let*

$$x_j^{k+1} = \omega_j^k \alpha_j + (1 - \omega_j^k)\beta_j, \tag{5.3}$$

where

$$\omega_j^k = (\beta_j - x_j^k) \bigg/ \left[(\beta_j - x_j^k) + (x_j^k - \alpha_j) \exp\left(-t_n \frac{\partial h^n}{\partial x_j}(x^k) \right) \right] \tag{5.4}$$

Using as h the function defined in (5.1) we obtain the transmission AB algorithm (TAB) with

$$\omega_j^k = (\beta_j - x_j^k) \bigg/ \left[(\beta_j - x_j^k) + (x_j^k - \alpha_j) \exp\left(-t_n \sum^n (L_{ij}(y_i - c_i \exp(-(Lx^k)_i))) \right) \right]; \tag{5.5}$$

the iterative step is

$$x_j^{k+1} = \omega_j^k \alpha_j + (1 - \omega_j^k)\beta_j. \tag{5.6}$$

The theory of the IPA tells us that we will have convergence to a constrained minimizer of the function $h(x)$ provided we choose t_n so that the function $F(x) - t_n h^n(x)$ is a convex function. We shall obtain an upper limit on the acceptable values of t_n by considering the Hessian matrix of the function f^n.

With $F(x) = F_{\alpha\beta}(x)$ and $h^n(x) = \sum^n KL(y_i, c_i \exp(-(Lx)_i))$ we find that the Hessian matrix of $F(x)$ is

$$\nabla^2 F(x) = \mathrm{diag}\{(x_j - \alpha_j)^{-1} + (\beta_j - x_j)^{-1}\}. \tag{5.7}$$

The Hessian matrix of $h^n(x)$ is

$$\nabla^2 h^n(x) = L^T W L, \tag{5.8}$$

where W is the diagonal matrix with $W_{ii} = c_i \exp(-(Lx)_i)$, for $i \in B_n$ and $W_{ii} = 0$ otherwise. The smallest eigenvalue of $\nabla^2 F(x)$ is the minimum, over all j, of $(x_j - \alpha_j)^{-1} + (\beta_j - x_j)^{-1}$. Since

$$(x_j - \alpha_j)^{-1} + (\beta_j - x_j)^{-1} \geqslant \frac{4}{\beta_j - \alpha_j}, \tag{5.9}$$

we know that the smallest eigenvalue of $\nabla^2 F(x)$ is not smaller than the minimum, over all j, of $\frac{4}{\beta_j - \alpha_j}$.

The trace of the matrix $\nabla^2 h^n(x)$, which is the sum of its eigenvalues, is

$$\mathrm{trace} \nabla^2 h^n(x) = \sum^n \left(\sum_j L_{ij}^2 \right) c_i \exp(-(Lx)_i). \tag{5.10}$$

Therefore, the Hessian of $f^n(x) = F(x) - \gamma_n h^n(x)$ will be positive definite, for all $x \in \mathrm{int} D$, if the constant $t_n > 0$ satisfies the inequality

$$t_n \leqslant 4 \min_j \left\{ \frac{1}{\beta_j - \alpha_j} \right\} \bigg/ \sum^n \left(\sum_j L_{ij}^2 \right) c_i \exp(-(L\alpha)_i), \tag{5.11}$$

where $\alpha = (\alpha_1, \dots, \alpha_J)^T$.

This algorithm and related methods are discussed in the context of medical imaging in [15].

6. Minimizing $KL(Px, y)$ for $\alpha \leqslant x \leqslant \beta$

Suppose now that we wish to minimize the function $KL(Px, y)$, where the entries of the matrix P and the vector x are non-negative and the entries of y are positive. Suppose, in addition, that we wish to impose the constraints $\alpha_j \leqslant x_j \leqslant \beta_j$, or, in vector notation, $\alpha \leqslant x \leqslant \beta$ or $x \in [\alpha, \beta]$, where $0 < \alpha_j < \beta_j$, for $j = 1, \ldots, J$. Clearly, if there is a solution $x \in [\alpha, \beta]$ with $y = Px$, then, from the positivity of the entries of P, we know that $y \in [P\alpha, P\beta]$. Unlike the algorithms given in [6], the method we present in this section does not restrict the entries of y in this way, thereby permitting its application to the important case in which the vector y consists of noisy measurement data. A related optimization algorithm is applied to medical imaging in [15].

We let $h(x) = KL(Px, y)$ and $F(x) = F_{\alpha\beta}(x)$. The following proposition will be helpful.

Proposition 6.1. *For all vectors x and z with $\alpha \leqslant x, z \leqslant \beta$ we have*

$$D_F(x, z) \geqslant KL(x, z) = D_E(x, z) \geqslant D_h(x, z) = KL(Px, Pz). \tag{6.1}$$

Proof. Since the functions $F(x)$ and $E(x)$ are separable, it suffices to prove the first inequality for the case of $J = 1$. For real numbers x in (α, β) consider the function

$$g(x) = (x - \alpha)\log(x - \alpha) + (\beta - x)\log(\beta - x) - x\log x + x.$$

We show that this function is strictly convex. The second derivative of g is

$$g''(x) = \frac{1}{x - \alpha} + \frac{1}{\beta - x} - \frac{1}{x},$$

which is easily shown to be positive for $\alpha < x < \beta$. It follows that

$$D_F(x, z) - KL(x, z) \geqslant g(x) - g(z) - \langle \nabla g(z), x - z\rangle \geqslant 0,$$

for all vectors x and z with $\alpha < x, z < \beta$. The second inequality in (6.1) follows from the inequality $KL(x, z) \geqslant KL(x_+, z_+)$ and the fact that the columns of P sum to unity. This completes the proof. □

The convexity of $f^n(x) = F(x) - t_n^{-1}h^n(x)$ follows immediately, where $h^n(x) = \sum^n KL(Px_i, y_i)$ and $t_n = \max\{\sum^n P_{ij} | j = 1, \ldots, J\}$. Applying the IPA, we find that the iterative step of the algorithm involves solving

$$\log \frac{x_j^{k+1} - \alpha_j}{\beta_j - x_j^{k+1}} = \log \frac{x_j^k - \alpha_j}{\beta_j - x_j^k} + t_n^{-1} \sum^n P_{ij} \log \frac{y_i}{(Px^k)_i} \tag{6.2}$$

for x_j^{k+1}. Our algorithm then is the following.

Algorithm 6.1. *Let $\alpha < x^0 < \beta$ be chosen. For $k = 0, 1, 2, \ldots$ let*

$$x_j^{k+1} = \omega_j^k \alpha_j + (1 - \omega_j^k)\beta_j, \tag{6.3}$$

where

$$\omega_j^k = (\beta_j - x_j^k) \Bigg/ \left[(\beta_j - x_j^k) + (x_j^k - \alpha_j)\left(t_n^{-1} \sum^n P_{ij} \log \frac{y_i}{(Px^k)_i}\right)\right]. \tag{6.4}$$

We have the following convergence result.

Theorem 6.1. *For $N = 1$ the sequence $\{x^k\}$ defined by (6.3) converges to x^∞, minimizing the function $KL(Px, y)$ over all x in $[\alpha, \beta]$. For $N \geqslant 1$, if there is a vector x in $[\alpha, \beta]$ with $y = Px$, then, for any initial vector x^0 with $\alpha < x^0 < \beta$, the sequence $\{x^k\}$ converges to x^∞; we have $y = Px^\infty$ and $\alpha \leqslant x^\infty \leqslant \beta$. Also, x^∞ is the unique such solution for which the function $D_F(x, x^0)$ is minimized.*

In the next section we obtain a similar theorem for the general linear system $Ax = b$.

7. Minimizing $\|Ax - b\|^2$ subject to $\alpha \leqslant x \leqslant \beta$

In this section we consider the minimization of $\|Ax - b\|^2$, subject to upper and lower bounds on the values of the entries of the vector x; specifically, we seek $x \in [\alpha, \beta]$. As before, we employ the Bregman–Legendre function $F(x) = F_{\alpha\beta}$, given by equation (2.6). The essential domain of this function is a compact set. Let $B = \max\{\frac{1}{4}(\beta_j - \alpha_j)\}$, with the maximum taken over the indices j. We let $h(x) = \|Ax - b\|^2$. Our algorithm is the following.

Algorithm 7.1. *Let x^0 be chosen so that, for all $j = 1, \ldots, J$, we have $\alpha_j < x_j^0 < \beta_j$. For $k = 0, 1, \ldots$ set*

$$x_j^{k+1} = \omega_j^k \alpha_j + (1 - \omega_j^k)\beta_j, \tag{7.1}$$

where

$$\omega_j^k = (\beta_j - x_j^k) \Big/ \left((\beta_j - x_j^k) + (x_j^k - \alpha_j) \exp\left(\frac{1}{2BI_n} \left(\sum^n A_{ij}(b_i - (Ax^k)_i) \right)_j \right) \right) \tag{7.2}$$

and $I_n = \sum^n (AA^T)_{ii}$.

We have the following result.

Theorem 7.1. *Let x^0 be chosen so that, for all $j = 1, \ldots, J$, we have $\alpha_j < x_j^0 < \beta_j$. Let $B = \max\{\frac{1}{4}(\beta_j - \alpha_j)\}$, with the maximum taken over the indices j. Then, for $N = 1$, the iterative sequence $\{x^k\}$ given by (7.1) converges to a minimizer of $\|Ax - b\|^2$ for $\alpha_j \leqslant x_j \leqslant \beta_j$, $j = 1, \ldots, J$. For $N \geqslant 1$, if there is an exact solution of $Ax = b$ then the sequence $\{x^k\}$ converges to the one for which the distance $D_F(x, x^0)$ is minimized.*

8. Summary and conclusions

We have presented a new iterative interior point algorithm, called the IPA, for minimizing a convex differentiable function over certain convex sets. The IPA applies when the associated convex set C is the essential domain of a Bregman–Legendre function f. The IPA is related to interior point methods recently proposed by Censor *et al* [10]. In the special case of minimizing a convex differentiable function over a convex set, the IPA, for the case of a single block, is a descent method.

The IPA is a block-iterative method, special cases of which have been shown to provide useful reconstructed tomographic images with few iterations. Our focus in this paper has been on the theoretical aspects of the method; simulations involving special cases of the IPA have appeared elsewhere.

As is the case with all block-iterative methods that do not employ strong underrelaxation, in the inconsistent case we obtain convergence not to a single vector, but to an LC. No proof of this has been obtained, except for the case of ART, but this is what has been consistently observed. For moderate levels of noise in the data, the images of the LC will still prove useful. For problems involving large data sets, only a few iterations will be performed, typically, so limiting behaviour is less important than providing reasonable reconstructions early in the iterative sequence.

In several of the examples presented here we use the interior point approach to incorporate upper and lower bounds on the individual pixels of the reconstructed image. It is possible to enforce such constraints merely by clipping the image at each step of an iteration procedure. In applications, however, the constraints used are typically conservatively chosen. Clipping tends to produce images with numerous values equal to the bounds. One benefit of the IPA approach is that the images truly lie within the interior of the constraint set, to the extent permitted by the data.

Acknowledgments

I wish to thank Yair Censor, Gabor Herman, Arnold Lent and Manoj Narayanan for helpful discussions of these topics.

Appendix. Bregman–Legendre functions and Bregman projections

In [1] 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 [1].

A.1. Essential smoothness and essential strict convexity

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 $D = \{x | f(x) < +\infty\}$. A proper convex function f is *closed* if it is lower semi-continuous. The *subdifferential* of f at x is the set $\partial f(x) = \{x^* | \langle x^*, z - x\rangle \leqslant f(z) - f(x), \text{for all } z\}$. The domain of ∂f is the set $\mathrm{dom}\partial f = \{x | \partial f(x) \neq \emptyset\}$. The *conjugate function* associated with f is the function $f^*(x^*) = \sup_z(\langle x^*, z\rangle - f(z))$.

Following [16] we say that a closed proper convex function f is *essentially smooth* if $\mathrm{int}D$ is not empty, f is differentiable on $\mathrm{int}D$ and $x^n \in \mathrm{int}D$, with $x^n \to x \in \mathrm{bd}D$, implies that $\|\nabla f(x^n)\| \to +\infty$. Here $\mathrm{int}D$ and $\mathrm{bd}D$ denote the interior and boundary of the set D.

A closed proper convex function f is *essentially strictly convex* if f is strictly convex on every convex subset of $\mathrm{dom}\partial f$.

The closed proper convex function f is essentially smooth if and only if the subdifferential $\partial f(x)$ is empty for $x \in \mathrm{bd}D$ and is $\{\nabla f(x)\}$ for $x \in \mathrm{int}D$ (so f is differentiable on $\mathrm{int}D$) if and only if the function f^* is essentially strictly convex.

A closed proper convex function f is said to be a *Legendre function* if it is both essentially smooth and essentially strictly convex, so f is Legendre if and only if its conjugate function is Legendre, in which case the gradient operator ∇f is a topological isomorphism with ∇f^* as its inverse. The gradient operator ∇f maps $\mathrm{int\,dom} f$ onto $\mathrm{int\,dom} f^*$. If $\mathrm{int\,dom} f^* = R^J$ 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 $\mathrm{int\,dom} f^* = R^J$ 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 the essential domain of f is bounded, then f is super-coercive and its gradient operator is a mapping onto the space R^J.

A.2. Bregman projections onto closed convex sets

Let f be a closed proper convex function that is differentiable on the nonempty set $\mathrm{int}D$. The corresponding *Bregman distance* $D_f(x, z)$ is defined for $x \in R^J$ and $z \in \mathrm{int}D$ by

$$D_f(x, z) = f(x) - f(z) - \langle\nabla f(z), x - z\rangle.$$

Note that $D_f(x, z) \geqslant 0$ always and that $D_f(x, z) = +\infty$ is possible. If f is essentially strictly convex then $D_f(x, z) = 0$ implies that $x = z$.

Let K be a nonempty closed convex set with $K \cap \mathrm{int}D \neq \emptyset$. Pick $z \in \mathrm{int}D$. The *Bregman projection* of z onto K, with respect to f, is

$$P_K^f(z) = \mathrm{argmin}_{x \in K \cap D} D_f(x, z).$$

If f is essentially strictly convex, then $P_K^f(z)$ exists. If f is strictly convex on D then $P_K^f(z)$ is unique. If f is Legendre, then $P_K^f(z)$ is uniquely defined and is in $\mathrm{int}D$; this last condition is sometimes called *zone consistency*.

Example. Let $J = 2$ and $f(x)$ be the function that is equal to one-half the norm squared on D, the non-negative quadrant, $+\infty$ elsewhere. Let K be the set $K = \{(x_1, x_2) | x_1 + x_2 = 1\}$. The Bregman projection of $(2, 1)$ onto K is $(1, 0)$, which is not in $\mathrm{int}D$. The function f is not essentially smooth, although it is essentially strictly convex. Its conjugate is the function f^* that is equal to one-half the norm squared on D and equal to zero elsewhere; it is essentially smooth, but not essentially strictly convex.

If f is Legendre, then $P_K^f(z)$ is the unique member of $K \cap \mathrm{int}D$ satisfying the inequality

$$\langle \nabla f(P_K^f(z)) - \nabla f(z), P_K^f(z) - c \rangle \geqslant 0,$$

for all $c \in K$. From this we obtain the *Bregman inequality*:

$$D_f(c, z) \geqslant D_f(c, P_K^f(z)) + D_f(P_K^f(z), z),$$

for all $c \in K$.

A.3. Bregman's sequential generalized projection algorithm

Let C_i, $i = 1, \ldots, I$, be closed nonempty convex sets in R^J with nonempty intersection K. Assume that $C_i \cap \mathrm{int}D \neq \emptyset$, for all i, and that $K \cap D \neq \emptyset$. In [3] Bregman presents the following sequential generalized projection (SGP) algorithm for finding a member of K.

Bregman's SGP algorithm. For $k = 0, 1, 2, \ldots$ and $i = k(\mathrm{mod}\, I)$ let

$$x^{k+1} = P_{C_i}^f(x^k).$$

In order for this algorithm to converge to a member of K additional restrictions on the function f are needed.

A.4. Bregman–Legendre functions

Following Bauschke and Borwein [1], we say that a Legendre function f is a *Bregman–Legendre* function if the following properties hold.

(B1) For x in D and any $a > 0$ the set $\{z | D_f(x, z) \leqslant a\}$ is bounded.

(B2) If x is in D but not in $\mathrm{int}D$, for each positive integer n, y^n is in $\mathrm{int}D$ with $y^n \to y \in \mathrm{bd}D$, and if $\{D_f(x, y^n)\}$ remains bounded then $D_f(y, y^n) \to 0$, so that $y \in D$.

(B3) If x^n and y^n are in $\mathrm{int}D$, with $x^n \to x$ and $y^n \to y$, where x and y are in D but not in $\mathrm{int}D$, and if $D_f(x^n, y^n) \to 0$ then $x = y$.

Bauschke and Borwein then prove that Bregman's SGP method converges to a member of K provided that one of the following holds: (1) f is Bregman–Legendre; (2) $K \cap \mathrm{int}D \neq \emptyset$ and $\mathrm{dom} f^*$ is open or (3) $\mathrm{dom} f$ and $\mathrm{dom} f^*$ are both open.

A.5. Useful results about Bregman–Legendre functions

The following results are proved in somewhat more generality in [1].

(R1) If $y^n \in \operatorname{int}\operatorname{dom} f$ and $y^n \to y \in \operatorname{int}\operatorname{dom} f$, then $D_f(y, y^n) \to 0$.
(R2) If x and $y^n \in \operatorname{int}\operatorname{dom} f$ and $y^n \to y \in \operatorname{bd}\operatorname{dom} f$, then $D_f(x, y^n) \to +\infty$.
(R3) If $x^n \in D$, $x^n \to x \in D$, $y^n \in \operatorname{int} D$, $y^n \to y \in D$, $\{x, y\} \cap \operatorname{int} D \neq \emptyset$ and $D_f(x^n, y^n) \to 0$, then $x = y$ and $y \in \operatorname{int} D$.
(R4) If x and y are in D, but are not in $\operatorname{int} D$, $y^n \in \operatorname{int} D$, $y^n \to y$ and $D_f(x, y^n) \to 0$, then $x = y$.
As a consequence of these results we have the following.
(R5) If $\{D_f(x, y^n)\} \to 0$, for $y^n \in \operatorname{int} D$ and $x \in R^J$, then $\{y^n\} \to x$.

Proof of (R5). Since $\{D_f(x, y^n)\}$ is eventually finite, we have $x \in D$. 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 x is in $\operatorname{int} D$, then, by result (R2) above, we know that y is also in $\operatorname{int} D$. Applying result (R3), with $x^n = x$, for all n, we conclude that $x = y$. If, on the other hand, x is in D, but not in $\operatorname{int} D$, then y is in D, by result (R2). There are two cases to consider: (1) y is in $\operatorname{int} D$; (2) y is not in $\operatorname{int} D$. In case (1) we have $D_f(x, y^n) \to D_f(x, y) = 0$, from which it follows that $x = y$. In case (2) we apply result (R4) to conclude that $x = y$. □

References

[1] Bauschke H and Borwein J 1997 Legendre functions and the method of random Bregman projections *J. Convex Anal.* **4** 27–67
[2] Bertsekas D P 1997 A new class of incremental gradient methods for least squares problems *SIAM J. Optim.* **7** 913–26
[3] Bregman L M 1967 The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming *USSR Comput. Math. Math. Phys.* **7** 200–17
[4] Byrne C L 1993 Iterative image reconstruction algorithms based on cross-entropy minimization *IEEE Trans. Image Process.* **2** 96–103
[5] Byrne C 1996 Block-iterative methods for image reconstruction from projections *IEEE Trans. Image Process.* **5** 792–4
[6] Byrne C L 1998 Iterative algorithms for deblurring and deconvolution with constraints *Inverse Problems* **14** 1455–67
[7] Byrne C 1999 Iterative projection onto convex sets using multiple Bregman distances *Inverse Problems* **15** 1295–313
[8] Byrne C 2000 Bregman–Legendre multidistance projection algorithms for convex feasibility and optimization *Proc. Workshop on Inherently Parallel Algorithms in Feasibility and Optimization (Haifa, 2000)* at press
[9] Censor Y, Eggermont P P B and Gordon D 1983 Strong underrelaxation in Kaczmarz's method for inconsistent systems *Numer. Math.* **41** 83–92
[10] Censor Y, Iusem A and Zenios S 1998 An interior point method with Bregman functions for the variational inequality problem with paramonotone operators *Math. Prog.* **81** 373–400
[11] Gordon R, Bender R and Herman G T 1970 Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography *J. Theor. Biol.* **29** 471–81
[12] Herman G T 1999 private communication
[13] Herman G T and Meyer L 1993 Algebraic reconstruction techniques can be made computationally efficient *IEEE Trans. Med. Imaging* **12** 600–9
[14] Lange K, Bahn M and Little R 1987 A theoretical study of some maximum likelihood algorithms for emission and transmission tomography *IEEE Trans. Med. Imaging* **6** 106–14
[15] Narayanan M, Byrne C and King M 1999 An interior point iterative reconstruction algorithm imcorporating upper and lower bounds, with application to SPECT transmission imaging in preparation
[16] Rockafellar R 1970 *Convex Analysis* (Princeton, NJ: Princeton University Press)

[17] Tsui B, Zhao X, Frey E and Gullberg G 1991 Comparison between ML-EM and WLS-CG algorithms for SPECT image reconstruction *IEEE Trans. Nucl. Sci.* **38** 1766–72

[18] Youla D C 1987 Mathematical theory of image restoration by the method of convex projections *Image Recovery: Theory and Applications* ed H Stark (Orlando, FL: Academic) pp 29–78