Inverse Problems **18** (2002) 441–453 PII: S0266-5611(02)27157-9

Iterative oblique projection onto convex sets and the split feasibility problem

Charles Byrne

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

E-mail: Charles_Byrne@uml.edu

Received 13 July 2001, in final form 25 January 2002
Published 7 March 2002
Online at stacks.iop.org/IP/18/441

Abstract

Let C and Q be nonempty closed convex sets in R^N and R^M, respectively, and A an M by N real matrix. The *split feasibility problem* (SFP) is to find $x \in C$ with $Ax \in Q$, if such x exist. An iterative method for solving the SFP, called the *CQ* algorithm, has the following iterative step:

$$x^{k+1} = P_C(x^k + \gamma A^T(P_Q - I)Ax^k),$$

where $\gamma \in (0, 2/L)$ with L the largest eigenvalue of the matrix A^TA and P_C and P_Q denote the orthogonal projections onto C and Q, respectively; that is, P_Cx minimizes $\|c - x\|$, over all $c \in C$. The *CQ* algorithm converges to a solution of the SFP, or, more generally, to a minimizer of $\|P_QAc - Ac\|$ over c in C, whenever such exist.

The *CQ* algorithm involves only the orthogonal projections onto C and Q, which we shall assume are easily calculated, and involves no matrix inverses. If A is normalized so that each row has length one, then L does not exceed the maximum number of nonzero entries in any column of A, which provides a helpful estimate of L for sparse matrices.

Particular cases of the *CQ* algorithm are the Landweber and projected Landweber methods for obtaining exact or approximate solutions of the linear equations $Ax = b$; the *algebraic reconstruction technique* of Gordon, Bender and Herman is a particular case of a block-iterative version of the *CQ* algorithm.

One application of the *CQ* algorithm that is the subject of ongoing work is dynamic emission tomographic image reconstruction, in which the vector x is the concatenation of several images corresponding to successive discrete times. The matrix A and the set Q can then be selected to impose constraints on the behaviour over time of the intensities at fixed voxels, as well as to require consistency (or near consistency) with measured data.

0266-5611/02/020441+13$30.00

1. Introduction

Let C and Q be nonempty closed convex sets in R^N and R^M, respectively, and A an M by N real matrix. Let $A^{-1}(Q) = \{x|Ax \in Q\}$ and $F = C \cap A^{-1}(Q)$. The problem, to find $x \in C$ with $Ax \in Q$, if such x exist, was called the *split feasibility problem* (SFP) by Censor and Elfving [9], where they used their multidistance method to obtain iterative algorithms for solving the SFP. Their algorithms, as well as others obtained later (see [8]) involve matrix inverses at each step. In this paper we present a new iterative method for solving the SFP, called the *CQ* algorithm, that does not involve matrix inverses. A block-iterative *CQ* algorithm, the BI*CQ* method, is also given here. Particular cases of the *CQ* algorithm are the Landweber and projected Landweber methods for obtaining exact or approximate solutions of the linear equations $Ax = b$; the *algebraic reconstruction technique* (ART) of Gordon *et al* [14] is a particular case of the BI*CQ*.

A number of image reconstruction problems can be formulated as split feasibility problems. The vector x represents a vectorized image, with the entries of x the intensity levels at each voxel or pixel. The set C can be selected to incorporate such features as non-negativity of the entries of x, while the matrix A can describe linear functional or projection measurements we have made, as well as other linear combinations of entries of x on which we wish to impose constraints. The set Q then can be the product of the vector of measured data with other convex sets, such as non-negative cones, that serve to describe the constraints to be imposed.

One particular application that is the subject of ongoing work is dynamic emission tomographic image reconstruction, in which the vector x is the concatenation of several images corresponding to successive discrete times. The matrix A and the set Q can then be selected to impose constraints on the behaviour over time of the intensities at fixed voxels, as well as to require consistency (or near consistency) with measured data.

Denote by P_C and P_Q the *orthogonal projection* (sometimes called the *proximity map*) onto C and Q, respectively; that is, $P_C x$ minimizes $\|c - x\|$, over all $c \in C$. The *CQ* algorithm to solve the SFP is the following.

Algorithm 1.1. *Let x^0 be arbitrary. For $k = 0, 1, \ldots,$ let*

The CQ algorithm

$$x^{k+1} = P_C(x^k + \gamma A^T(P_Q - I)Ax^k), \tag{1.1}$$

where $\gamma \in (0, 2/L)$ and L denotes the largest eigenvalue of the matrix $A^T A$.

Note that the *CQ* algorithm involves only the orthogonal projections onto C and Q, which we shall assume are easily calculated, and involves no matrix inverses. Later we show that if A is normalized so that each row has length one, then L does not exceed the maximum number of nonzero entries in any column of A.

Let K_j, $j = 1, \ldots, J$, be nonempty closed convex subsets of R^M, with nonempty intersection K. The *convex feasibility problem* (CFP) is to find an element of K. Solving the SFP is equivalent to finding a member of the intersection of the two sets Q and $A(C) = \{Ac|c \in C\}$, or of the intersection of the two sets $A^{-1}(Q)$ and C, and so the SFP can be viewed as a particular case of the CFP.

In [11] Cheney and Goldstein considered the case of two nonempty closed convex sets K_1 and K_2 in Hilbert space. With P_j denoting the orthogonal projection (proximity map) onto K_j, for $j = 1, 2$, respectively, and $T = P_1 P_2$, they show that the sequence $x^k = T^k x^0$ obtained by alternating distance minimization converges to a fixed point of T if either (a) one of the two sets is compact, or (b) one set is finite dimensional and the distance between the two sets is attained. In particular, when the intersection of the two convex sets is nonempty, the sequence $\{x^k\}$ converges to a member of that intersection, in either of the two cases above. This result can be extended in several directions, to include finitely many convex sets with nonempty intersection and projections involving generalized distances; for details see the survey papers by Deutsch [12] and Bauschke and Borwein [2] and the recent books by Censor and Zenios [10] and Stark and Yang [21].

If we try to use the Cheney–Goldstein approach to solve the SFP, we encounter the difficulty of calculating the orthogonal projection onto the set $A(C)$ or onto $A^{-1}(Q)$. In this paper we shall assume that the orthogonal projections P_C and P_Q are easily calculated; the main virtue of the *CQ* algorithm is that it involves only the maps P_C and P_Q at each step.

For the case in which $M = N$ and A is invertible, the map $P^{A^TA}_{A^{-1}(Q)} = A^{-1}P_QA$ is an oblique projection onto the set $A^{-1}(Q)$; that is, $A^{-1}P_Q(Ax)$ minimizes the function $f(z) = (z-x)^TA^TA(z-x)$ over all z in $A^{-1}(Q)$. This suggests the possibility of modifying the Cheney–Goldstein method to include one orthogonal and one oblique projection. But the following counterexample shows us that the iterative sequence

$$x^{k+1} = P_C P^{A^TA}_{A^{-1}(Q)}(x^k) \tag{1.2}$$

need not converge to a solution. Let $M = N = 2$, with C the horizontal axis, Q the vertical axis and A the matrix

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix}.$$

Let $x^0 = (1,0)^T$. Then $Ax^0 = (1,-1)^T$, $P_Q(Ax^0) = (0,-1)^T$, $A^{-1}P_Q(Ax^0) = (1,1)^T$ and $P_CA^{-1}P_Q(Ax^0) = (1,0)^T = x^0$. The solution to the SFP in this case is $(0,0)$, to which the iterative sequence given by (1.2) fails to converge.

It is possible, however, to formulate convergent iterative algorithms using projections based on distinct generalized distances, provided a form of generalized relaxation is included at each step. Such an algorithm, called the *multidistance successive generalized projection* (MSGP) method, is the main topic of the papers [7] and [8]. The *CQ* algorithm involves a relaxation parameter γ in the set $(0, 2/L)$, where L is the largest eigenvalue of the matrix A^TA; for $\gamma \in (0, 1/L]$ the *CQ* algorithm is a particular case of the MSGP and earlier versions of this paper derived the *CQ* algorithm and established its convergence by relating it to the MSGP method. Here, however, we prove convergence of the *CQ* algorithm directly and for the wider bounds on the parameter γ.

In a later section we shall consider the CFP in more detail and describe the connection between the *CQ* algorithm and the MSGP method for solving the CFP.

2. Convergence of the *CQ* algorithm

Let F be the (possibly empty) set of all $c \in C$ at which the function $\|P_QAc - Ac\|$ attains its minimum value over C. We have the following theorem concerning the *CQ* algorithm.

Theorem 2.1. *Let F be nonempty. Then the sequence $\{x^k\}$ defined by equation (1.1) converges to a member of F, for any starting vector x^0.*

Corollary 2.1. *The sequence $\{x^k\}$ defined by equation (1.1) converges to a solution of the SFP, whenever such solutions exist.*

Unlike previously published iterative algorithms for the SFP in [8] and [9], this algorithm involves no nested matrix inverses. Each step is easily performed, assuming, as we do here, that the maps P_C and P_Q are themselves easy to implement.

The orthogonal projection $P_C x$ of x onto C is characterized by the following useful inequality: for all $c \in C$ and all x we have

$$\langle c - P_C x, P_C x - x\rangle \geqslant 0. \tag{2.1}$$

2.1. Some preliminary results

In this section we establish some facts concerning the *CQ* algorithm that we shall need in the proof of convergence. We show that $\hat{c}$ is a fixed point of the *CQ* iteration step if and only if $\hat{c}$ minimizes the function $\|P_Q(Ac) - Ac\|$ over $c \in C$. Such fixed points need not exist, as the following example shows. Let C be the subset of R^2 consisting of all points $(x, y)^T$ with $x > 0$ and $y \geqslant 1/x$. Then C is closed. Let $A(x, y)^T = x$. Then $A(C)$ is not closed; the origin is in the closure of $A(C)$. Let $Q = \{0\}$; then, for each $c = (x, y)^T$ we have $\|P_Q(Ac) - Ac\| = \|x\|$, which can be made arbitrarily close to zero, but not equal to zero. Our main theorem is that the *CQ* algorithm converges to a fixed point of the *CQ* iteration, for all starting vectors, whenever such fixed points exist.

For each x let

$$Sx = x + \gamma A^T(P_Q - I)Ax \tag{2.2}$$

and $Tx = P_C(Sx)$; then the *CQ* iterative step is $x^{k+1} = Tx^k$. We begin with the following proposition.

Proposition 2.1. *The vector $\hat{c}$ in C is a fixed point of the map T, that is, $T\hat{c} = \hat{c}$, if and only if $\hat{c}$ minimizes the function $\|P_Q(Ac) - Ac\|$ over $c \in C$.*

Proof of the proposition. Assume that $\hat{c}$ minimizes the function $\|P_Q(Ac) - Ac\|$ over $c \in C$. Then

$$\|P_Q(A\hat{c}) - A\hat{c}\| \leqslant \|P_Q(Ac) - Ac\| \leqslant \|q - Ac\|$$

for all $c \in C$ and $q \in Q$. Choosing $q = P_Q(A\hat{c})$ we find that

$$\|P_Q(A\hat{c}) - A\hat{c}\| \leqslant \|Ac - P_Q(A\hat{c})\|$$

for all $c \in C$, which tells us that $A\hat{c} = P_{\overline{A(C)}}(P_Q(A\hat{c}))$. The inequality (2.1) then gives us that

$$\langle Ac - A\hat{c}, A\hat{c} - P_Q(A\hat{c})\rangle \geqslant 0,$$

for all $c \in C$. From

$$\|c - S\hat{c}\|^2 = \|c - \hat{c}\|^2 + 2\gamma\langle Ac - A\hat{c}, A\hat{c} - P_Q(A\hat{c})\rangle + \text{terms without } c$$

it follows that $\hat{c}$ minimizes the function $\|c - S\hat{c}\|$ over $c \in C$, or that $\hat{c} = P_C(S\hat{c}) = T\hat{c}$.

Now assume that $T\hat{c} = \hat{c}$. Then, $\hat{c} = P_C(S\hat{c})$, so that, by inequality (2.1), we have

$$\langle c - \hat{c}, \hat{c} - S\hat{c}\rangle \geqslant 0,$$

for all $c \in C$. Therefore,

$$\langle Ac - A\hat{c}, A\hat{c} - P_Q(A\hat{c})\rangle \geqslant 0,$$

for all $c \in C$. We also have

$$\langle P_Q(A\hat{c}) - P_Q(Ac), A\hat{c} - P_Q(A\hat{c})\rangle \geqslant 0.$$

Adding, we obtain

$$\langle P_Q(Ac) - Ac, P_Q(A\hat{c}) - A\hat{c}\rangle \geqslant \|P_Q(A\hat{c}) - A\hat{c}\|^2.$$

Applying the Cauchy inequality, we have

$$\|P_Q(Ac) - Ac\| \geqslant \|P_Q(A\hat{c}) - A\hat{c}\|.$$

This completes the proof. □

The inequality (2.1) tells us that, for any C, x and z, we have

$$\langle P_C z - P_C x, P_C x - x\rangle \geqslant 0 \qquad \text{and} \qquad \langle P_C z - P_C x, z - P_C z\rangle \geqslant 0.$$

Adding, we obtain

$$\langle P_C z - P_C x, P_C x - P_C z - x + z\rangle \geqslant 0,$$

or

$$\langle P_C z - P_C x, z - x\rangle \geqslant \|P_C z - P_C x\|^2; \tag{2.3}$$

therefore the operator P_C is *firmly nonexpansive* [22]. From the Cauchy inequality we conclude that

$$\|P_C x - P_C z\| \leqslant \|x - z\|; \tag{2.4}$$

that is, the operator P_C is *non-expansive*. In fact, we can say somewhat more (see [11]).

Lemma 2.1. *For any closed nonempty convex set C in R^N the inequality (2.4) holds, with equality only if $\|P_C x - x\| = \|P_C z - z\|$.*

2.2. Proof of the convergence theorem for the CQ algorithm

Let F be nonempty and $\hat{c}$ a member of F. Then $\hat{c} = P_C(S\hat{c})$ and

$$\|\hat{c} - x^{k+1}\| = \|P_C(S\hat{c}) - P_C(Sx^k)\| \leqslant \|S\hat{c} - Sx^k\|.$$

We shall show that

$$\|S\hat{c} - Sx^k\| \leqslant \|\hat{c} - x^k\|.$$

From the definition of Sx we have

$$\|S\hat{c} - Sx^k\|^2 = \|\hat{c} - x^k + \gamma A^T(P_Q - I)A\hat{c} - \gamma A^T(P_Q - I)Ax^k\|^2.$$

Expanding the term on the right side we get

$$\begin{aligned}\|S\hat{c} - Sx^k\|^2 = \|\hat{c} - x^k\|^2 &+ 2\gamma\langle A\hat{c} - Ax^k, P_Q A\hat{c} - P_Q Ax^k + Ax^k - A\hat{c}\rangle \\ &+ \gamma^2\|A^T(P_Q - I)A\hat{c} - A^T(P_Q - I)Ax^k\|^2 \leqslant \|\hat{c} - x^k\|^2 - 2\gamma\|A\hat{c} - Ax^k\|^2 \\ &+ 2\gamma\langle A\hat{c} - Ax^k, P_Q A\hat{c} - P_Q Ax^k\rangle + \gamma^2 L\|(P_Q - I)A\hat{c} - (P_Q - I)Ax^k\|^2.\end{aligned}$$

Using

$$\begin{aligned}\|(P_Q - I)A\hat{c} - (P_Q - I)Ax^k\|^2 &= \|P_Q A\hat{c} - P_Q Ax^k\|^2 \\ &- 2\langle A\hat{c} - Ax^k, P_Q A\hat{c} - P_Q Ax^k\rangle + \|A\hat{c} - Ax^k\|^2\end{aligned}$$

in the line above, we find that

$$\begin{aligned}\|S\hat{c} - Sx^k\|^2 \leqslant \|\hat{c} - x^k\|^2 &- (2\gamma - \gamma^2 L)\|A\hat{c} - Ax^k\|^2 + \gamma^2 L(\|P_Q A\hat{c} - P_Q Ax^k\|^2 \\ &- \langle A\hat{c} - Ax^k, P_Q A\hat{c} - P_Q Ax^k\rangle) \\ &+ (2\gamma - \gamma^2 L)\langle A\hat{c} - Ax^k, P_Q A\hat{c} - P_Q Ax^k\rangle.\end{aligned}$$

From equation (2.3) we have

$$\|P_Q A\hat{c} - P_Q Ax^k\|^2 - \langle A\hat{c} - Ax^k, P_Q A\hat{c} - P_Q Ax^k\rangle \leqslant 0$$

and from Cauchy's inequality and the nonexpansiveness of the projection P_Q we obtain

$$\langle A\hat{c} - Ax^k, P_Q A\hat{c} - P_Q Ax^k \rangle \leqslant \|A\hat{c} - Ax^k\|^2.$$

Since $2\gamma - \gamma^2 L \geqslant 0$, it follows that

$$\|S\hat{c} - Sx^k\|^2 \leqslant \|\hat{c} - x^k\|^2.$$

More precisely, we have

$$\begin{aligned}\|\hat{c} - x^k\|^2 - \|\hat{c} - x^{k+1}\|^2 \geqslant \gamma^2 L(\langle A\hat{c} - Ax^k, P_Q A\hat{c} - P_Q Ax^k\rangle - \|P_Q A\hat{c} - P_Q Ax^k\|^2) \\ + (2\gamma - \gamma^2 L)(\|A\hat{c} - Ax^k\|^2 - \langle A\hat{c} - Ax^k, P_Q A\hat{c} - P_Q Ax^k\rangle).\end{aligned} \tag{2.5}$$

Therefore, the sequence $\{\|\hat{c} - x^k\|^2\}$ is decreasing (so the sequence $\{x^k\}$ is bounded). Also

$$\{\langle A\hat{c} - Ax^k, P_Q A\hat{c} - P_Q Ax^k\rangle - \|P_Q A\hat{c} - P_Q Ax^k\|^2\} \to 0$$

and

$$\{\|A\hat{c} - Ax^k\|^2 - \langle A\hat{c} - Ax^k, P_Q A\hat{c} - P_Q Ax^k\rangle\} \to 0,$$

since both sequences are non-negative.

Let x^* be an arbitrary cluster point of the sequence $\{x^k\}$. Then we have

$$\langle A\hat{c} - Ax^*, P_Q A\hat{c} - P_Q Ax^*\rangle = \|P_Q A\hat{c} - P_Q Ax^*\|^2$$

and

$$\langle A\hat{c} - Ax^*, P_Q A\hat{c} - P_Q Ax^*\rangle = \|A\hat{c} - Ax^*\|^2,$$

so that

$$\|A\hat{c} - Ax^*\| = \|P_Q A\hat{c} - P_Q Ax^*\|.$$

From lemma 2.1 it follows that

$$\|P_Q A\hat{c} - A\hat{c}\| = \|P_Q Ax^* - Ax^*\|,$$

so that x^* is in the set F. Replacing the generic $\hat{c} \in F$ with x^*, we see that the sequence $\{\|x^* - x^k\|\}$ is decreasing; but a subsequence converges to zero, so the entire sequence converges to zero. This completes the proof of the theorem.

We turn now to two particular cases of the *CQ* algorithm, the Landweber and projected Landweber methods for solving $Ax = b$.

3. The Landweber methods

As particular cases of the *CQ* algorithm we obtain the Landweber [17] and projected Landweber methods. These algorithms are discussed in some detail in the book by Bertero and Boccacci [3], primarily in the context of image restoration within infinite-dimensional spaces of functions (see also [18]). With $C = R^N$ and $Q = \{b\}$ the *CQ* algorithm becomes the Landweber iterative method for solving the linear equations $Ax = b$.

The Landweber algorithm

With x^0 arbitrary and $k = 0, 1, \ldots,$ let

$$x^{k+1} = x^k + \gamma A^T(b - Ax^k). \tag{3.1}$$

For general nonempty closed convex C we obtain the projected Landweber method for finding a solution of $Ax = b$ in C.

The projected Landweber algorithm

For x^0 arbitrary and $k = 0, 1, \ldots,$ let

$$x^{k+1} = P_C(x^k + \gamma A^T(b - Ax^k)). \tag{3.2}$$

From the proof of convergence of the *CQ* algorithm it follows that the Landweber algorithm converges to a least-squares solution of $Ax = b$ and the projected Landweber algorithm converges to a minimizer of $\|Ac - b\|$ over all $c \in C$, whenever such solutions exist.

3.1. An example: the SART

The *simultaneous algebraic reconstruction* technique (SART) of Anderson and Kak [1] is an iterative method for solving $Ax = b$ for the case in which A is a non-negative matrix. We prove convergence by showing SART to be a particular case of the Landweber method and determining the largest eigenvalue of $A^T A$.

Let A be an M by N matrix with non-negative entries. Let $A_{i+} > 0$ be the sum of the entries in the ith row of A and $A_{+j} > 0$ be the sum of the entries in the jth column of A. Consider the (possibly inconsistent) system $Ax = b$. The SART algorithm has the following iterative step:

$$x_j^{k+1} = x_j^k + \frac{1}{A_{+j}} \sum_{i=1}^{M} (b_i - (Ax^k)_i)/A_{i+}.$$

We make the following changes of variables:

$$B_{ij} = A_{ij}/(A_{i+})^{1/2}(A_{+j})^{1/2}, \qquad z_j = x_j(A_{+j})^{1/2}, \qquad \text{and} \qquad c_i = b_i/(A_{i+})^{1/2}.$$

Then the SART iterative step can be written as

$$z^{k+1} = z^k + B^T(c - Bz^k).$$

This is a particular case of the Landweber algorithm, with $\gamma = 1$. The convergence of SART will follow from theorem 2.1, once we have shown that the largest eigenvalue of $B^T B$ is less than two; in fact, we show it is one.

If $B^T B$ had an eigenvalue greater than one and some of the entries of A are zero, then, replacing these zero entries with very small positive entries, we could obtain a new A whose associated $B^T B$ also had an eigenvalue greater than one. Therefore, we assume, without loss of generality, that A has all positive entries. Since the new $B^T B$ also has only positive entries, this matrix is irreducible and the Perron–Frobenius theorem applies. We shall use this to complete the proof.

Let $u = (u_1, \ldots, u_N)^T$ with $u_j = (A_{+j})^{1/2}$ and $v = (v_1, \ldots, v_M)^T$, with $v_i = (A_{i+})^{1/2}$. Then we have $Bu = v$ and $B^T v = u$; that is, u is an eigenvector of $B^T B$ with associated eigenvalue equal to one, and all the entries of u are positive, by assumption. The Perron–Frobenius theorem applies and tells us that the eigenvector associated with the largest eigenvalue has all positive entries. Since the matrix $B^T B$ is symmetric its eigenvectors are orthogonal; therefore u is an eigenvector associated with the largest eigenvalue of $B^T B$. The convergence of SART follows.

4. A helpful eigenvalue inequality

The *CQ* algorithm employs the relaxation parameter γ in the interval $(0, 2/L)$, where L is the largest eigenvalue of the matrix $A^T A$. Choosing the best relaxation parameter in any algorithm is a nontrivial procedure. The inequality (2.5) in the proof of convergence of the *CQ* algorithm, as well as experience with related algorithms, suggests that, generally speaking, we want to select γ nearer to its upper bound $2/L$ than to zero. In practice, it would be helpful to have a quick method for estimating L. In this section we present such a method, particularly useful for sparse matrices.

Let A be an M by N matrix. For each $n = 1, \ldots, N$, let $s_n > 0$ be the number of nonzero entries in the nth column of A and let s be the maximum of the s_n. Let G be the M by N matrix with entries

$$G_{mn} = A_{mn} \bigg/ \left(\sum_{l=1}^{N} s_l A_{ml}^2 \right)^{1/2}.$$

Lent has shown that the eigenvalues of the matrix $G^T G$ do not exceed one [19]. The following proposition and its proof are straightforward extensions of Lent's work.

Proposition 4.1. *Let A be an M by N matrix. For each $m = 1, \ldots, M$ let $\nu_m = \sum_{n=1}^{N} A_{mn}^2 > 0$. For each $n = 1, \ldots, N$ let $\sigma_n = \sum_{m=1}^{M} e_{mn}\nu_m$, where $e_{mn} = 1$ if $A_{mn} \neq 0$ and $e_{mn} = 0$ otherwise. Let σ denote the maximum of the σ_n. Then the eigenvalues of the matrix $A^T A$ do not exceed σ. If A is normalized so that the Euclidean length of each of its rows is one, then the eigenvalues of $A^T A$ do not exceed s, the maximum number of nonzero elements in any column of A.*

Proof. Let $A^T A v = cv$ for some nonzero vector v. We show that $c \leqslant \sigma$. We have $AA^T A v = cAv$ and so $w^T AA^T w = v^T A^T AA^T Av = cv^T A^T Av = cw^T w$, for $w = Av$. Then

$$\left(\sum_{m=1}^{M} A_{mn} w_m \right)^2 = \left(\sum_{m=1}^{M} A_{mn} e_{mn} w_m \right)^2 \leqslant \left(\sum_{m=1}^{M} A_{mn}^2 w_m^2/\nu_m \right) \left(\sum_{m=1}^{M} \nu_m e_{mn}^2 \right)$$
$$= \left(\sum_{m=1}^{M} A_{mn}^2 w_m^2/\nu_m \right) \sigma_n \leqslant \left(\sum_{m=1}^{M} A_{mn}^2 w_m^2/\nu_m \right) \sigma.$$

Therefore,

$$w^T A^T A w = \sum_{n=1}^{N} \left(\sum_{m=1}^{M} A_{mn} w_m \right)^2 \leqslant \sum_{n=1}^{N} \left(\sum_{m=1}^{M} A_{mn}^2 w_m^2/\nu_m \right) \sigma.$$

We also have

$$w^T A^T A w = c \sum_{m=1}^{M} w_m^2 = c \sum_{m=1}^{M} w_m^2 \left(\sum_{n=1}^{N} A_{mn}^2 \right) \bigg/ \nu_m = c \sum_{m=1}^{M} \sum_{n=1}^{N} w_m^2 A_{mn}^2 \bigg/ \nu_m.$$

The result follows immediately. □

If we normalize A so that its rows have length one, then the trace of the matrix AA^T is $\mathrm{tr}(AA^T) = M$, which is also the sum of the eigenvalues of $A^T A$. Consequently, the maximum eigenvalue of $A^T A$ does not exceed M; the result above improves that considerably, if A is sparse and so $s \ll M$.

5. The CFP and the MSFP algorithm

For $j = 1, \ldots, J$, let P_j denote the orthogonal projection onto the nonempty closed convex set $K_j \subseteq R^N$ and $T = P_J P_{J-1} \cdots P_1$. If the intersection K of the K_j is nonempty then, for any starting vector x^0, the sequence $\{T^k x^0\}$ converges to a member of K. This follows from the results on the method of *successive orthogonal projections* (SOP) in [15]. In [4] Bregman presents the method of *successive generalized projections* (SGP); he shows that convergence holds if the orthogonal projections P_j are replaced by projections defined with respect to a single fixed Bregman generalized distance. Counterexamples demonstrate that convergence may be lost if T is composed of projections associated with different Bregman distances. Little is known about the case of empty K, beyond the Cheney–Goldstein theorem for the case of orthogonal projections and $J = 2$.

The MSGP algorithm extends Bregman's iterative SGP procedure to include Bregman projections corresponding to multiple Bregman distances. We discuss the MSGP briefly, and ask the reader to consult the book by Censor and Zenios [10] for details and definitions regarding Bregman functions, distances and projections.

Let h and f be Bregman functions with zones S_h, S_f, respectively. Assume that their associated Bregman distances, D_h and D_f satisfy $D_h(x, z) \geqslant D_f(x, z)$ for all $x \in \overline{S}_h \cap \overline{S}_f$, $z \in S_h \cap S_f$. For $x \in \overline{S}_h \cap \overline{S}_f$, $y \in S_f$, $z \in S_h \cap S_f$ let $G(x; y, z, f, h)$ be the function of x defined as follows:

$$G(x; y, z, f, h) = D_f(x, y) + D_h(x, z) - D_f(x, z). \tag{5.1}$$

Let f_j, $j = 1, \dots, J$, be a family of Bregman functions and let h be a Bregman function that 'dominates' the family, that is, for which $D_h(x, z) \geqslant D_{f_j}(x, z)$, for all j and all $x \in \overline{S}_h \cap \overline{S}_{f_j}$ and $z \in S_h \cap S_{f_j}$. Let $S = S_h \cap (\cap_{j=1}^{J} S_j)$. Denote by $P_{K_j}^{f_j} x$ the minimizer of $D_{f_j}(z, x)$, over $z \in K_j$. The MSGP algorithm is the following.

Algorithm 5.1 (the MSGP algorithm). *For $k = 0, 1, \dots,$ and having calculated x^k, we obtain x^{k+1} as follows: with $j = k(\mathrm{mod}\, J) + 1$, let $G^k(x) := G(x; P_{K_j}^{f_j}(x^k), x^k, f_j, h)$. We assume that $G^k(x)$ has a unique minimizer, which we take as x^{k+1}. We assume also that $x^{k+1} \in S_h$, so that*

$$\nabla h(x^{k+1}) = \nabla h(x^k) - \nabla f_j(x^k) + \nabla f_j(P_{K_j}^{f_j}(x^k)). \tag{5.2}$$

Finally, we assume that we have cyclic zone consistency; that is, for each k, the vector x^{k+1} defined by (5.2) is in S_{f_m}, $m = (k + 1)(\mathrm{mod}\, J) + 1$.

We have the following convergence theorem.

Theorem 5.1. *Let $K \cap S$ be nonempty. Any sequence x^k obtained from the iterative scheme given by algorithm 5.1 converges to a member of $K \cap S$.*

As we noted earlier, when A is invertible the oblique projection $P_{A^{-1}(Q)}^{A^T A}$ is $A^{-1} P_Q A$, so can be calculated easily, if P_Q can be. Using the MSGP algorithm, with $K_1 = C$ and $K_2 = A^{-1}(Q)$, and with distances and projections defined using the Bregman functions $h(x) = f_1(x) = x^T x$ and $f_2(x) = x^T A^T A x$, we obtain the CQ algorithm, provided that $\gamma \in (0, 1/L]$. As shown above, the CQ algorithm is valid even for $\gamma \in (0, 2/L)$ and for arbitrary matrix A.

6. Regularization and the *CQ* algorithm

Many of the problems in image reconstruction and remote sensing for which we might use the CQ algorithm are ill conditioned; the data are noisy and some form of regularization is required to avoid noisy reconstructions. We wish to point out that regularization can be included within the CQ algorithm, although other methods, such as using Bayesian techniques or penalty functions, or simply stopping the iteration early, may prove superior. For example, if we wish to solve $Ax = b$ we can use the Landweber method, which is the CQ algorithm with $C = R^N$ and $Q = \{b\}$. But, if the data vector b is noisy and $A^T A$ has some small eigenvalues, an exact or least-squares solution may be useless. Instead, we might let $Q = \{y | \, \|y-b\| \leqslant \epsilon\}$, for some small $\epsilon > 0$. Another choice for Q could be $Q = \{y | |y_m - b_m| \leqslant \epsilon|b_m|, \; m = 1, \ldots, M\}$. If we are using the projected Landweber method, with, say, C the non-negative cone in R^N, we could enlarge C to permit slightly negative entries in x. If we are performing Fourier band-limited extrapolation via the Gerchberg method (see [3]), which is another particular case of the Landweber algorithm, we could expand the frequency band to permit out-of-band noise in the data. To illustrate the sort of difficulty that can arise, we present a result concerning the behaviour of the projected Landweber algorithm for the case in which the set C is the non-negative cone.

For concreteness, imagine that the vector x represents a vectorization of a two-dimensional image, with each of the N entries of x equal to the non-negative image intensity at the corresponding pixel; ultimately, these intensities will be quantized to grey-levels, with zero denoting (say) black, but we view them as continuous for now. Suppose that our data consist of linear functional values, so that we wish to solve $Ax = b$, a system of M linear equations in N unknowns, for $x \geqslant 0$.

It is natural to suppose that a coarse image can be made finer by using a greater number of pixels in the image. The result below shows that, counter to our intuition, the use of more pixels can make the image worse.

Suppose that the system $Ax = b$ has no non-negative solution. Let F denote the set of all $\hat{x} \geqslant 0$ that minimize the function $\|Ax - b\|$ over all non-negative x; although $\hat{x}$ need not be unique, the vector $A\hat{x}$ is unique, by the strict convexity of Euclidean space. We know from the Kuhn–Tucker theorem (see [20]) that $(A^T(b - A\hat{x}))_n = 0$ for all n for which there is $\hat{x} \in F$ whose nth entry is positive.

Let

$$D = \{n | \, n \in \{1, 2, \ldots, N\} \text{ and } \hat{x}_n > 0, \text{ for some } \hat{x} \in F\}.$$

Let B be the matrix formed from A by removing those columns of A whose index is not a member of D. Then we have $B^T(b - A\hat{x}) = 0$. If the number of rows of B does not exceed the number of columns and B has full rank, then the mapping induced by B^T is one-to-one. Consequently $b = A\hat{x}$, which contradicts our assumption. Therefore, if we assume that A and any matrix obtained from A by deleting columns have full rank, then the cardinality of the set D cannot exceed $M - 1$. In that case all the vectors $\hat{x}$ in F are supported on the entries in set D, so $\hat{x}$ is unique and has fewer than M nonzero entries, regardless of how large we make N. A similar result was obtained in [5] for likelihood maximization.

7. Applications of the *CQ* algorithm in tomography

To illustrate how an image reconstruction problem can be formulated as an SFP, we consider briefly *single-photon emission computed tomography* (SPECT) image reconstruction. The objective in SPECT is to reconstruct the internal spatial distribution of intensity of a radionuclide from counts of photons detected outside the patient. In static SPECT the intensity distribution is assumed constant over the scanning time. Our data are photon counts at the detectors, forming the positive vector b and we have a matrix B of detection probabilities; our model is $Bx = b$, for x a non-negative vector. We could then take $A = B$, $Q = \{b\}$ and $C = R^N_+$, the non-negative cone in R^N.

In dynamic SPECT [13] the intensity levels at each voxel may vary with time. The observation time is subdivided into, say, T intervals and one static image, call it x^t, is associated with the time interval denoted by t, for $t = 1, \ldots, T$. The vector x is the concatenation of these T image vectors x^t. The discrete time interval at which each data value is collected is also recorded and the problem is to reconstruct this succession of images. Because the data associated with a single time interval are insufficient, by themselves, to generate a useful image, one often uses prior information concerning the time history at each fixed voxel to devise a model of the behaviour of the intensity levels at each voxel, as functions of time. One may, for example, assume that the radionuclide intensities at a fixed voxel are increasing with time, or are concave (or convex) with time. The problem then is to find $x \geqslant 0$ with $Bx = b$ and $Dx \geqslant 0$, where D is a matrix chosen to describe this additional prior information. For example, we may wish to require that, for each fixed voxel, the intensity is an increasing function of (discrete) time; then we want

$$x_j^{t+1} - x_j^t \geqslant 0,$$

for each t and each voxel index j. Or, we may wish to require that the intensity at each voxel describes a concave function of time, in which case non-negative second differences would be imposed:

$$(x_j^{t+1} - x_j^t) - (x_j^{t+2} - x_j^{t+1}) \geqslant 0.$$

In either case, the matrix D can be selected to include the left sides of these inequalities, while the set Q can include the non-negative cone as one factor.

Application of the *CQ* algorithm to dynamic SPECT is the subject of ongoing work and will not be discussed further here.

8. The SFP and the block-iterative *CQ* algorithm

Experience with algorithms such as the *expectation maximization maximum likelihood* (EMML) method suggests strongly that significant acceleration of convergence of the *CQ* algorithm can be had through the use of block-iterative variants (see [6]). We present such a block-iterative version of the *CQ* algorithm now.

For $j = 1, \ldots, J$ let A_j be an M_j by N matrix and Q_j a nonempty closed convex subset of R^{M_j}. We wish to find $x \in R^N$ with $x \in C$ and $A_j x \in Q_j$, for $j = 1, \ldots, J$. We could, of course, apply the *CQ* algorithm, once we concatenate the matrices A_j to form a single matrix A and take Q to be the product of the sets Q_j. Instead, we consider a block-iterative extension of the *CQ* algorithm.

We operate under the assumption that the orthogonal projections onto the sets Q_j are easily calculated. The iterative algorithm we obtain we call the *block-iterative CQ*, or BI*CQ*, algorithm.

The BICQ algorithm

Let x^0 be arbitrary. For $k = 0, 1, \ldots,$ and $j(k) = k \pmod J + 1$ let

$$x^{k+1} = P_C(x^k + \gamma_{j(k)} A_{j(k)}^T (P_{Q_{j(k)}} - I)(A_{j(k)} x^k)), \tag{8.1}$$

with $\gamma_j \in (0, 2/L_j)$, where L_j denotes the largest eigenvalue of the matrix $A_j^T A_j$.

Essentially the same proof as for the *CQ* algorithm establishes that the BI*CQ* algorithm converges to a vector $\hat{c} \in C$ with $A_j \hat{c} \in Q_j$, for all j, whenever such vectors $\hat{c}$ exist. Experience with similar algorithms suggests that convergence is accelerated most when J, the number of blocks, is large and the relaxation parameters γ_j are chosen near their upper bounds, rather than near to zero.

The BI*CQ* becomes the ART of Gordon *et al* [14] (see also [16]) for solving $Ax = b$ if we select each block to contain exactly one row index and the corresponding matrix to be that row of A.

If the system of equation $Ax = b$ is inconsistent, the ART will not converge to a single vector, but to a limit cycle, consisting of (usually) as many separate vectors as there are rows in A. We can obtain the least-squares approximate solution of $Ax = b$ using a modification of ART that we call the *double* ART (DART). The DART involves two applications of ART, both to consistent systems of equations. First, use ART to solve $A^T w = 0$, starting at $b = w^0$. The limit, w^*, is then the orthogonal projection of b onto the null space of A^T. Now apply ART again to the consistent system $Ax = b - w^*$; the limit is the least-squares approximate solution of $Ax = b$ closest to x^0. If we normalize A so that its rows have length one, then x^* is the geometric least-squares approximate solution closest to x^0.

9. Summary

An image reconstruction problem can often be formulated as a CFP, in which we seek a (vectorized) image x in the intersection of finitely many convex sets. The SFP is a special case of the CFP of interest to us here. Let $C \subset R^N$ and $Q \subset R^M$ be convex sets and A an M by N matrix. The SFP is to find $x \in C$ with $Ax \in Q$, if such x exist. We often find that it is easy to calculate the orthogonal projections P_C and P_Q onto C and Q, respectively, but not the orthogonal projections onto $A(C) = \{z | z = Ax, x \in C\}$ or $A^{-1}(Q) = \{x | Ax \in Q\}$. The sequence $\{x^k\}$ generated by our *CQ* algorithm converges to a solution of the SFP, or, more generally, to a minimizer of the function $\|P_Q Ac - Ac\|$ over $c \in C$, whenever such exist, for any scalar γ in the interval $(0, 2/L)$, where L is the largest eigenvalue of $A^T A$.

The main points to note about the *CQ* algorithm are that it involves only easily calculated orthogonal projections and it requires no matrix inversions. When the matrix A is sparse, the upper bound L is easily estimated from the number of nonzero entries of A; this not only makes the *CQ* algorithm easier to implement, but it also accelerates the convergence by permitting γ to take on larger values.

The *CQ* algorithm can be extended to a block-iterative version, the BI*CQ*, which may also provide further acceleration.

The application of the *CQ* algorithm to dynamic tomographic image reconstruction is the subject of ongoing work.

Acknowledgments

The author wishes to thank Professor James Graham-Eagle of the University of Massachusetts Lowell Department of Mathematical Sciences and Dr Troy Farncombe of the University of Massachusetts Medical School for helpful discussions concerning these matters. This work was supported, in part, by the National Institutes of Health (NIH), under grant CA23452. The contents of this paper are solely the responsibility of the author and do not necessarily represent the official views of NIH.

References

[1] Anderson A and Kak A 1984 Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm *Ultrason. Imaging* **6** 81–94

[2] Bauschke H H and Borwein J M 1996 On projection algorithms for solving convex feasibility problems *SIAM Rev.* **38** 367–426

[3] Bertero M and Boccacci P 1998 *Introduction to Inverse Problems in Imaging* (Bristol: Institute of Physics Publishing)

[4] 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

[5] Byrne C L 1993 Iterative image reconstruction algorithms based on cross-entropy minimization *IEEE Trans. Image Process.* **2** 96–103

[6] Byrne C L 1996 Block-iterative methods for image reconstruction from projections *IEEE Trans. Image Process.* **5** 792–4

[7] Byrne C L 1999 Iterative projection onto convex sets using multiple Bregman distances *Inverse Problems* **15** 1295–313

[8] Byrne C L 2001 Bregman–Legendre multidistance projection algorithms for convex feasibility and optimization *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications* ed D Butnariu, Y Censor and S Reich (Amsterdam: Elsevier) pp 87–100

[9] Censor Y and Elfving T 1994 A multiprojection algorithm using Bregman projections in a product space *Numer. Algorithms* **8** 221–39

[10] Censor Y and Zenios S A 1997 *Parallel Optimization: Theory, Algorithms and Applications* (New York: Oxford University Press)

[11] Cheney W and Goldstein A 1959 Proximity maps for convex sets *Proc. Am. Math. Soc.* **10** 448–50

[12] Deutsch F 1992 The method of alternating orthogonal projections *Approximation Theory, Spline Functions and Applications* ed S P Singh (Dordrecht: Kluwer) pp 105–21

[13] Farncombe T 2000 Functional dynamic SPECT imaging using a single slow camera rotation *PhD Thesis* Department of Physics, University of British Columbia

[14] 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

[15] Gubin L G, Polyak B T and Raik E V 1967 The method of projections for finding the common point of convex sets *USSR Comput. Math. Math. Phys.* **7** 1–24

[16] Kaczmarz S 1937 Angenäherte Auflösung von Systemen linearer Gleichungen *Bull. Acad. Pol. Sci. Lett.* A **35** 355–7

[17] Landweber L 1951 An iterative formula for Fredholm integral equations of the first kind *Am. J. Math.* **73** 615–24

[18] Lagendijk R and Biemond J 1991 *Iterative Identification and Restoration of Images* (Dordrecht: Kluwer)

[19] Lent A 1998 private communication

[20] Luenberger D 1969 *Optimization by Vector Space Methods* (New York: Wiley)

[21] Stark H and Yang Y 1998 *Vector Space Projections: a Numerical Approach to Signal and Image Processing, Neural Nets and Optics* (New York: Wiley)

[22] Tseng P 1992 On the convergence of the products of firmly nonexpansive mappings *SIAM J. Optim.* **2** 425–34