INSTITUTE OF PHYSICS PUBLISHING INVERSE PROBLEMS

Inverse Problems **20** (2004) 103–120 PII: S0266-5611(04)68668-0

A unified treatment of some iterative algorithms in signal processing and image reconstruction

Charles Byrne

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

E-mail: Charles_Byrne@uml.edu

Received 9 September 2003
Published 14 November 2003
Online at stacks.iop.org/IP/20/103 (DOI: 10.1088/0266-5611/20/1/006)

Abstract

Let T be a (possibly nonlinear) continuous operator on Hilbert space $\mathcal{H}$. If, for some starting vector x, the orbit sequence $\{T^k x, k = 0, 1, \ldots\}$ converges, then the limit z is a fixed point of T; that is, $Tz = z$. An operator N on a Hilbert space $\mathcal{H}$ is nonexpansive (ne) if, for each x and y in $\mathcal{H}$,

$$\|Nx - Ny\| \leqslant \|x - y\|.$$

Even when N has fixed points the orbit sequence $\{N^k x\}$ need not converge; consider the example $N = -I$, where I denotes the identity operator. However, for any $\alpha \in (0, 1)$ the iterative procedure defined by

$$x^{k+1} = (1 - \alpha)x^k + \alpha N x^k$$

converges (weakly) to a fixed point of N whenever such points exist. This is the Krasnoselskii–Mann (KM) approach to finding fixed points of ne operators.

A wide variety of iterative procedures used in signal processing and image reconstruction and elsewhere are special cases of the KM iterative procedure, for particular choices of the ne operator N. These include the Gerchberg–Papoulis method for bandlimited extrapolation, the SART algorithm of Anderson and Kak, the Landweber and projected Landweber algorithms, simultaneous and sequential methods for solving the convex feasibility problem, the ART and Cimmino methods for solving linear systems of equations, the CQ algorithm for solving the split feasibility problem and Dolidze's procedure for the variational inequality problem for monotone operators.

1. Introduction and overview

Many well-known algorithms in signal processing and image reconstruction are iterative in nature. The *projection onto convex sets* (POCS) methods and iterative optimization procedures, such as entropy or likelihood maximization, are the primary examples. The editorial [49] provides a brief introduction to many of the recent efforts in medical imaging. The purpose of this paper is to give a unified treatment of several of these algorithms.

The iterative methods we shall consider have the form

$$x^{k+1} = Tx^k, \tag{1.1}$$

for $k = 0, 1, \ldots$, where T is a linear or nonlinear continuous operator on a real (possibly infinite dimensional) Hilbert space $\mathcal{H}$ and x^0 is an arbitrary starting vector. For any operator T on $\mathcal{H}$ the *fixed point set* of T is

$$\text{Fix}(T) = \{z | Tz = z\}.$$

If the iterative sequence defined by equation (1.1) converges then the limit is a member of $\text{Fix}(T)$.

A wide variety of problems can be solved by finding a fixed point of a particular operator, and algorithms for finding such points play a prominent role in a number of applications. The paper [61] is an excellent source of background on these topics, particularly as they apply to signal and image processing. The present paper can perhaps be viewed as a sequel to [61]. The more recent article by Bauschke and Borwein [4] is also quite helpful. The book by Borwein and Lewis [9] is also an important reference.

In the algorithms of interest here the operator T is selected so that the set $\text{Fix}(T)$ contains those vectors z that possess the properties we desire in a solution to the original signal processing or image reconstruction problem; finding a fixed point of the iteration leads to a solution of our problem.

To illustrate, suppose that C is a closed convex set in $\mathcal{H}$, such as the nonnegative vectors in R^N. The metric projection operator P_C associates with every x in $\mathcal{H}$ the point $P_C x$ in C that is nearest to x. If C_1 and C_2 are two such sets the fixed points of the operator $T = P_{C_2} P_{C_1}$ are the vectors in the intersection $C = C_1 \cap C_2$, if C is nonempty; then the sequence $\{T^k x^0\}$ converges to a member of C. The convergence is generally in the weak sense, for infinite dimensional spaces. Finding points in the intersection of convex sets is called the *convex feasibility problem* (CFP).

Some applications involve constrained optimization, in which we seek a vector x in a given convex set C that minimizes a certain function f. For suitable $\gamma > 0$ the fixed points of the operator $T = P_C(I - \gamma \nabla f)$ will solve this problem; under conditions to be discussed below the sequence $\{T^k x\}$ will converge to a solution.

Our concern here is with properties of the operator T sufficient to guarantee convergence of the sequence $\{T^k x\}$ whenever fixed points of T exist. Most studies of iterative fixed point algorithms begin with the class of *nonexpansive* (ne) maps and we shall do the same.

A (possibly nonlinear) operator N on $\mathcal{H}$ is called ne if, for all x and y in $\mathcal{H}$,

$$\|Nx - Ny\| \leqslant \|x - y\|.$$

The identity map $Ix = x$ for all x is clearly ne; more generally, for any fixed vector w in $\mathcal{H}$ the maps $Nx = x + w$ and $Nx = -x + w$ are ne. As the example $Nx = -x$ shows, convergence of the sequence $\{N^k x\}$ is not guaranteed for ne operators, even when $\text{Fix}(N)$ is nonempty.

The Krasnoselskii–Mann (KM) [51] approach to finding fixed points of a ne operator N is quite simple, yet remarkably useful. Given a ne operator N, let

$$T = (1 - \alpha)I + \alpha N$$

for some $\alpha \in (0, 1)$. The operator T is then said to be *averaged* (av); note that T is then also ne. The KM theorem discussed below tells us that the sequence defined by equation (1.1) then converges (weakly) to a fixed point of N whenever such points exist. As will be discussed and shown below, the operators P_C are av, as are the operators $(I - \gamma \nabla f)$ if ∇f is Lipschitz continuous and the parameter γ is appropriately chosen; the product of finitely many av

operators is av, so the operators $P_{C_2}P_{C_1}$ and $P_C(I - \gamma\nabla f)$ are also av. Consequently, fixed points of such operators are limits of the sequence defined by equation (1.1).

We begin, in the next section, with a detailed discussion of av operators, followed by an examination of the proof of the KM theorem. We then consider constrained optimization and convex feasibility generally, followed by examples from signal processing and image reconstruction. In the final sections we broaden the discussion to include projecting onto convex sets using more general notions of distance, such as cross-entropy, as well as operators that are somewhat weaker than av.

2. Averaged nonexpansive operators

As we have seen, the fact that a ne operator N has fixed points is not sufficient to guarantee convergence of the orbit sequence $\{N^k x\}$; additional conditions are needed. An operator S on $\mathcal{H}$ is said to be a *strict contraction* (sc) if there is $\sigma \in (0, 1)$ such that, for all x and y in $\mathcal{H}$,

$$\|Sx - Sy\| \leqslant \sigma\|x - y\|.$$

The well known Banach–Picard theorem [35] assures us that the operator S has a unique fixed point, to which the orbit sequence $\{S^k x\}$ converges, for any starting point x. Requiring the operator to be a sc is quite restrictive; most of the operators we are interested in here have multiple fixed points, so are not sc. The KM theorem suggests strongly that we should concentrate on av operators. We have the following result.

Theorem 2.1. *Let T be an av operator on $\mathcal{H}$ and let* Fix(T) *be nonempty. Then the orbit sequence $\{T^k x\}$ converges weakly to a member of* Fix(T)*, for any x.*

We include a proof of this theorem, for the finite dimensional case, in a later section. Many of the iterative methods used in signal and image processing are special cases of the KM approach. A somewhat more general result is the following [31].

Theorem 2.2. *Let N be a ne operator on $\mathcal{H}$. For $k = 0, 1, \ldots$ let $\alpha_k \in (0, 1)$. Then the sequence $\{x^k\}$ defined by the iterative step*

$$x^{k+1} = (1 - \alpha_k)x^k + \alpha_k N x^k$$

converges weakly to a fixed point of N, provided $\sum_{k=0}^{\infty} \alpha_k(1-\alpha_k) = +\infty$, whenever such fixed points exist.

An operator $G : \mathcal{H} \to \mathcal{H}$ is *monotone* [56, 62, 38] if, for all x and y,

$$\langle Gx - Gy, x - y\rangle \geqslant 0. \tag{2.1}$$

To illustrate, suppose that $g(\cdot)$ is a convex, differentiable real-valued function on $\mathcal{H}$. Then

$$\langle \nabla g(y), x - y\rangle \leqslant g(x) - g(y)$$

and

$$\langle \nabla g(x), y - x\rangle \leqslant g(y) - g(x).$$

Adding, we obtain

$$\langle \nabla g(x) - \nabla g(y), x - y\rangle \geqslant 0.$$

Therefore the derivative of convex function $g(\cdot)$ is a monotone operator. If $\hat{c}$ minimizes the function $g(\cdot)$ over the closed convex set C, then

$$\langle \nabla g(\hat{c}), c - \hat{c}\rangle \geqslant 0$$

for all $c \in C$. For general monotone operator G the *variational inequality problem* (VIP) with respect to G and C, denoted VIP(G, C), is to find $\hat{c}$ in C with

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

for all $c \in C$ [62, 30, 60]. Subject to certain restrictions on G and γ, the sequence defined by the iterative step

$$x^{k+1} = P_C(I - \gamma G)x^k \tag{2.2}$$

will converge to a solution of the VIP(G, C), if solutions exist.

For each $x \in \mathcal{H}$ the metric projection $P_C x$ is that member of C closest to x and is characterized as the unique element of C for which

$$\langle c - P_C x, P_C x - x\rangle \geqslant 0, \tag{2.3}$$

for all $c \in C$ (see [61, p 33], or [57, p 43]). Therefore $\hat{c} = P_C(I - \gamma G)\hat{c}$ if and only if

$$\langle c - \hat{c}, \hat{c} - (\hat{c} - \gamma G\hat{c})\rangle = \gamma\langle c - \hat{c}, G\hat{c}\rangle \geqslant 0,$$

for all $c \in C$. Consequently, the vector $\hat{c}$ solves the VIP(G, C) if and only if $\hat{c}$ is a fixed point of the operator $P_C(I - \gamma G)$. This is the motivation for considering the iteration in (2.2).

As we shall see now, in seeking fixed points for an operator T it is helpful to consider properties of its complement, $I - T$. The following identity relates an operator T to its complement $G = I - T$:

$$\|x - y\|^2 - \|Tx - Ty\|^2 = 2\langle Gx - Gy, x - y\rangle - \|Gx - Gy\|^2. \tag{2.4}$$

An operator G on $\mathcal{H}$ is called *ν-inverse strongly monotone* (ν-ism) [38, 60] (also called *co-coercive* in [31]) if there is $\nu > 0$ such that

$$\langle Gx - Gy, x - y\rangle \geqslant \nu\|Gx - Gy\|^2.$$

From equation (2.4) we see immediately that N is ne if and only if its complement $G = I - N$ is $\frac{1}{2}$-ism. If G is ν-ism and $\gamma > 0$ then the operator γG is $\frac{\nu}{\gamma}$-ism.

Lemma 2.1. *An operator A is av if and only if its complement $G = I - A$ is ν-ism for some $\nu > \frac{1}{2}$.*

Proof. We assume first that A is av. Then there is $\alpha \in (0, 1)$ and ne operator N such that $A = (1 - \alpha)I + \alpha N$, and so $G = I - A = \alpha(I - N)$. Since N is ne, $I - N$ is $\frac{1}{2}$-ism and $G = \alpha(I - N)$ is $\frac{1}{2\alpha}$-ism. Conversely, assume that G is ν-ism for some $\nu > \frac{1}{2}$. Let $\alpha = \frac{1}{2\nu}$ and write $A = (1 - \alpha)I + \alpha N$ for $N = I - \frac{1}{\alpha}G$. Since $I - N = \frac{1}{\alpha}G$, $I - N$ is $\alpha\nu$-ism. Consequently $I - N$ is $\frac{1}{2}$-ism and N is ne. Therefore, A is av. □

Lemma 2.2. *Let $T = (1-\alpha)A + \alpha N$ for some $\alpha \in (0, 1)$. If A is av and N is ne then T is av.*

Proof. Let $A = (1 - \beta)I + \beta M$ for some $\beta \in (0, 1)$ and ne operator M. Let $1 - \gamma = (1 - \alpha)(1 - \beta)$. Then we have

$$T = (1 - \gamma)I + \gamma[(1 - \alpha)\beta\gamma^{-1}M + \alpha\gamma^{-1}N].$$

Since the operator $K = (1 - \alpha)\beta\gamma^{-1}M + \alpha\gamma^{-1}N$ is easily shown to be ne and the convex combination of two ne operators is again ne, T is av. □

An operator F on $\mathcal{H}$ is called *firmly nonexpansive* (fne) [61, 4] if it is 1-ism.

Lemma 2.3. *An operator F is fne if and only if its complement $I - F$ is fne. If F is fne then F is av.*

Proof. For any operator F with $G = I - F$ we have

$$\langle Fx - Fy, x - y\rangle - \|Fx - Fy\|^2 = \langle Gx - Gy, x - y\rangle - \|Gx - Gy\|^2.$$

The left-hand side is nonnegative if and only if the right-hand side is. Finally, if F is fne then $I - F$ is fne, so $I - F$ is ν-ism for $\nu = 1$. Therefore F is av by lemma 2.1. □

Corollary 2.1. *Let $T = (1-\alpha)F + \alpha N$ for some $\alpha \in (0, 1)$. If F is fne and N is ne then T is av.*

Since the metric projection of x onto C is characterized by the inequalities

$$\langle c - P_C x,\, P_C x - x\rangle \geqslant 0$$

for all $c \in C$, we have

$$\langle P_C y - P_C x,\, P_C x - x\rangle \geqslant 0$$

and

$$\langle P_C x - P_C y,\, P_C y - y\rangle \geqslant 0.$$

Adding, we find that

$$\langle P_C x - P_C y, x - y\rangle \geqslant \|P_C x - P_C y\|^2;$$

the operator P_C is fne, and therefore also av.

The product of finitely many ne operators is again ne, while the product of finitely many fne operators, even metric projections, need not be fne. It is a helpful fact that the product of finitely many av operators is again av.

If $A = (1-\alpha)I + \alpha N$ is av and B is av then $T = AB$ has the form $T = (1-\alpha)B + \alpha NB$. Since B is av and NB is ne, it follows from lemma 2.1 that T is av. Summarizing, we have

Proposition 2.1. *If A and B are av, then $T = AB$ is av.*

Combining this proposition with theorem 2.1 we obtain Dolidze's theorem [34, 38]:

Theorem 2.3. *Let G be ν-ism and $\gamma \in (0, 2\nu)$. Then, for any x, the sequence $\{(P_C(I - \gamma G))^k x\}$ converges weakly to a solution of* VIP(G,C)*, whenever solutions exist.*

Proof. The operator γG is $\frac{1}{2\alpha}$-ism, so $I - \gamma G$ and $P_C(I - \gamma G)$ are av. □

It is possible for $\mathrm{Fix}(AB)$ to be nonempty while $\mathrm{Fix}(A) \cap \mathrm{Fix}(B)$ is empty; however, if the latter is nonempty, it must coincide with $\mathrm{Fix}(AB)$ [4]:

Proposition 2.2. *Let A and B be av operators and suppose that* $\mathrm{Fix}(A) \cap \mathrm{Fix}(B)$ *is nonempty. Then* $\mathrm{Fix}(A) \cap \mathrm{Fix}(B) = \mathrm{Fix}(AB) = \mathrm{Fix}(BA)$.

Proof. Let $I - A$ be ν_A-ism and $I - B$ be ν_B-ism, where both ν_A and ν_B are taken greater than $\frac{1}{2}$. Let z be in $\mathrm{Fix}(A) \cap \mathrm{Fix}(B)$ and x in $\mathrm{Fix}(BA)$. Then

$$\begin{aligned}\|z - x\|^2 &\geqslant \|z - Ax\|^2 + (2\nu_A - 1)\|Ax - x\|^2 \\ &\geqslant \|z - BAx\|^2 + (2\nu_B - 1)\|BAx - Ax\|^2 + (2\nu_A - 1)\|Ax - x\|^2 \\ &= \|z - x\|^2 + (2\nu_B - 1)\|BAx - Ax\|^2 + (2\nu_A - 1)\|Ax - x\|^2.\end{aligned}$$

Therefore $\|Ax - x\| = 0$ and $\|BAx - Ax\| = \|Bx - x\| = 0$. □

If $A_1, \ldots, A_M$ are av operators, then so are the operators $A = \frac{1}{M}\sum_{m=1}^{M} A_m$ and $B = A_M A_{M-1} \cdots A_1$. The orbit sequence $\{A^k x\}$ will converge weakly whenever $\mathrm{Fix}(A)$

is nonempty; such an iterative scheme is sometimes called a *simultaneous* method. If the operators A_m have common fixed points, then the orbit sequence $\{B^k x\}$ will converge weakly to such a common fixed point; such methods are sometimes called *sequential* methods. If $\text{Fix}(B)$ is nonempty, but the A_m have no common fixed point, then the sequence $\{B^k x\}$ converges to a fixed point z^0 such that, with $z^1 = A_1 z, z^2 = A_2 z^1, \ldots, z^{M-1} = A_{M-1} z^{M-2}$, we have $A_M z^{M-1} = z^0$. Such a set of M vectors is called a *limit cycle*.

In the next section we examine the proof of theorem 2.1, in order to better understand the advantages of sequential methods over simultaneous ones.

3. The proof of theorem 2.1

In the previous section we noted that, given av operators $A_m, m = 1, \ldots, M$, we can iterate using either of two av operators, the operator $A = \frac{1}{M}\sum_{m=1}^{M} A_m$ or the operator $B = A_M A_{M-1} \cdots A_1$. Now we examine the proof of theorem 2.1 to see what the advantages of these two choices might be. Although the theorem holds for infinite dimensional $\mathcal{H}$ using weak convergence, we limit the discussion here to the finite dimensional case.

Let z be a fixed point of ne operator N and let $\alpha \in (0, 1)$. Let $T = (1-\alpha)I + \alpha N$, so the iterative step becomes

$$x^{k+1} = Tx^k = (1-\alpha)x^k + \alpha N x^k. \tag{3.1}$$

The identity in equation (2.4) is the key to proving theorem 2.1.

Using $Tz = z$ and $(I - T)z = 0$ and setting $G = I - T$ we have

$$\|z - x^k\|^2 - \|Tz - x^{k+1}\|^2 = 2\langle Gz - Gx^k, z - x^k\rangle - \|Gz - Gx^k\|^2.$$

Since, by lemma 2.1, G is $\frac{1}{2\alpha}$-ism, we have

$$\|z - x^k\|^2 - \|z - x^{k+1}\|^2 \geqslant \left(\frac{1}{\alpha} - 1\right)\|x^k - x^{k+1}\|^2. \tag{3.2}$$

Consequently the sequence $\{x^k\}$ is bounded, the sequence $\{\|z - x^k\|\}$ is decreasing and the sequence $\{\|x^k - x^{k+1}\|\}$ converges to zero. Let x^* be a cluster point of $\{x^k\}$. Then we have $Tx^* = x^*$, so we may use x^* in place of the arbitrary fixed point z. It follows then that the sequence $\{\|x^* - x^k\|\}$ is decreasing; since a subsequence converges to zero, the entire sequence converges to zero. The proof is complete. □

Equation (3.2) provides an estimate of the improvement we make in approaching the set $\text{Fix}(T)$ at each step of the iteration. We now apply this equation to $T = A$ and B.

We begin with $T = A$. By lemma 2.1 the operator $I - A_m$ is $\frac{1}{2\alpha_m}$-ism for each m, where, in most applications, the α_m is determined by the selection of a parameter. Therefore, we assume that each of α_m is equal to α. The improvement after one step is $(\frac{1}{\alpha} - 1)\|x^k - x^{k+1}\|^2$. For $T = B$ we see that we could make roughly the same order of magnitude improvement after applying only a single one of the operators A_m, and after applying all M operators that make up B our improvement could be of the order of M times that obtained by one iteration using A. This order of magnitude acceleration in convergence using sequential methods is commonly seen with the ART, MART and RBI-EMML algorithms discussed below, particularly if the operators A_m are randomly ordered [41]. If the calculation needed to compute Ax is roughly M times that required to calculate $A_m x$ for a single value of m, then the sequential method will converge, when there are common fixed points, about M times as fast as the simultaneous method. Of course, if we begin with an av operator T and define $A_m = \frac{1}{M}T$, for $m = 1, 2, \ldots, M$, then calculating Tx is the same as calculating a single $A_m x$, so no acceleration is achieved.

To illustrate, suppose $A_m = P_{C_m}$ for $m = 1, 2, \ldots, M$, where C_m are closed nonempty convex subsets. If the sets are distinct, then calculating all the M projections $P_{C_m}x$ requires M times the computation needed to calculate a single projection, more or less.

4. Constrained optimization algorithms

Algorithms for signal and image processing are often iterative constrained optimization procedures designed to minimize a convex differentiable function $f(x)$ over a closed convex set C in $\mathcal{H}$.

If the gradient operator ∇f is λ-Lipschitz continuous, that is, for each x and y in $\mathcal{H}$ we have

$$\|\nabla f(x) - \nabla f(y)\| \leqslant \lambda\|x - y\|,$$

then the operator ∇f is $\frac{1}{\lambda}$-ism [3]. If $\gamma \in (0, \frac{2}{\lambda})$ then the operator $G = \gamma\nabla f$ is $\frac{1}{2\alpha}$-ism and the operators $A = I - \gamma\nabla f$ and $P_C A$ are av. From theorem 2.1 we then conclude the following.

Corollary 4.1. *Let f be convex and differentiable on an open set D containing the closed convex set $C \subseteq \mathcal{H}$. If ∇f is a λ-Lipschitz continuous operator on D and $\gamma \in (0, \frac{2}{\lambda})$, then the sequence defined by $x^{k+1} = P_C(x^k - \gamma\nabla f(x^k))$ converges weakly to a minimizer of f relative to the set C, whenever such minimizers exist, for any starting vector x^0.*

In the sections to follow we consider a number of special cases of orbit sequences of av operators arising in signal and image processing.

5. The convex feasibility problem

Let $C_1, \ldots, C_M$ be closed nonempty convex subsets of Hilbert space $\mathcal{H}$. The CFP is to find a member of their intersection, if such elements exist [4, 29]. Problems in image reconstruction are sometimes formulated in this way, with the elements of the Hilbert space $\mathcal{H} = R^N$ corresponding to vectorized images and the convex sets representing various constraints to be placed on the reconstructed images. For example, we may have measured data in the form of linear functional values associated with the image, say $b_m = \langle a^m, x\rangle, m = 1, 2, \ldots, M$, where the a^m denote fixed vectors. The set C_m might then be the set of all vectors w with $\langle a^m, w\rangle = b_m$. We may wish to impose the condition that the entries of x be nonnegative, in which case we would include the nonnegative cone of R^N as one of the convex sets. Methods involving the metric projection operators are usually termed POCS methods [61, 57].

The *proximity function* associated with $C_1, \ldots, C_M$ is

$$f(x) = \frac{1}{2M}\sum_{m=1}^{M}\|P_{C_m}x - x\|^2. \tag{5.1}$$

The gradient of f is

$$\nabla f(x) = x - \frac{1}{M}\sum\nolimits_{m=1}^{M} P_{C_m}x;$$

see [2]. A minimizer of f is a zero of ∇f, which is a fixed point of the av operator A given by

$$A = \frac{1}{M}\sum\nolimits_{m=1}^{M} P_{C_m}.$$

If the intersection C of the sets $C_1, \ldots, C_M$ is nonempty, then $\mathrm{Fix}(A) = C$ and the orbit sequence $\{A^k x\}$ converges weakly to a member of C; if C is empty, the sequence converges weakly to a minimizer of f. In the latter case the limit need not be a member of any of the C_m.

If we wish to minimize f relative to vectors x in closed nonempty convex set K, then we use the iterative scheme defined by

$$x^{k+1} = P_K(Ax^k).$$

The operator $P_K A$ is also av, so this sequence converges, for any starting vector x^0, provided f has a minimum relative to K.

The methods just described are often called *simultaneous* because we compute the metric projections onto each of the sets C_m at each step of the iteration. We may also proceed sequentially, as follows.

Since each of the operators P_{C_m} is av, so is their product. We can therefore consider the sequence

$$x^{k+1} = P_{C_{m(k)}} x^k,$$

where $k = 0, 1, \ldots$ and $m(k) = k(\text{mod } M) + 1$. If the intersection C is nonempty, then the sequence $\{x^k\}$ converges to a member of C. If C is empty, the sequence $\{x^k\}$ will not converge. Because the product of the operators P_{C_m} is also av, the subsequences $\{x^{jM+m} | j = 1, 2, \ldots\}$ will converge weakly, for each fixed m, but to distinct limit points, provided the product operator has fixed points. Sequential methods are closely related to *incremental* methods [8].

6. Cimmino's method and the algebraic reconstruction technique

To illustrate the simultaneous and sequential methods just described, we consider the problem of solving a system of linear equations $Ax = b$, where A is a real M by N matrix. For $m = 1, \ldots, M$ let a^m be the mth column of A^{T}, so that $b_m = \langle a^m, x\rangle$, and let $C_m = \{w | \langle a^m, w\rangle = b_m\}$. Assume for notational convenience that the rows of A have length one. Then we have

$$P_{C_m} x = x + (b_m - \langle a^m, x\rangle)a^m.$$

The simultaneous algorithm now takes the form

$$x^{k+1} = x^k + \frac{1}{M}\sum_{m=1}^{M}(b_m - \langle a^m, x^k\rangle)a^m$$

or

$$x^{k+1} = x^k + \frac{1}{M}A^{\mathrm{T}}(b - Ax^k).$$

This method is sometimes called *Cimmino's method* (see [29]). The sequential method has the iterative step

$$x^{k+1} = x^k + (b_{m(k)} - \langle a^{m(k)}, x^k\rangle)a^{m(k)},$$

for $m(k) = k(\text{mod } M) + 1$. This method, originally due to Kaczmarz [44], is called the *algebraic reconstruction technique* (ART) [39, 41]. It has been shown by Tanabe [58] that the product of the metric projections has fixed points in this case (see also [40]).

When the system $Ax = b$ has multiple solutions both Cimmino's method and the ART converge to that solution closest to the starting vector x^0.

7. Bandlimited extrapolation methods

In this section we consider the bandlimited extrapolation problem as an illustration of alternating POCS. The resulting iterative algorithm is the Gerchberg–Papoulis (GP)

method [37, 55]. As we shall see, the method can be implemented in a noniterative manner, leading to more general linear and nonlinear extrapolation procedures that have been used for image and array processing.

The continuous formulation of *the bandlimited extrapolation problem* is the following: let $f(t)$ and $F(\omega)$ be a Fourier transform pair, where t and ω are real variables and

$$F(\omega) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{\mathrm{i}t\omega}\,\mathrm{d}t, \tag{7.1}$$

$$f(t) = \int_{-\infty}^{\infty} F(\omega)\mathrm{e}^{-\mathrm{i}t\omega}\,\mathrm{d}\omega/2\pi. \tag{7.2}$$

We assume that $F(\omega) = 0$, for $|\omega| > \Omega$, where Ω is a positive quantity. The function $f(t)$ is then said to be Ω-*bandlimited*. If we know $f(t)$ for t in some bounded interval of the real line, then these data determine $F(\omega)$ uniquely, by analyticity; the extension of $f(t)$ to complex z, given by

$$f(z) = \int_{-\infty}^{\infty} F(\omega)\mathrm{e}^{-\mathrm{i}z\omega}\,\mathrm{d}\omega/2\pi, \tag{7.3}$$

can be differentiated under the integral sign, since the limits of integration are finite. Therefore, the known values of $f(t)$ determine $f(z)$ for all other values of z; we can, in theory, extrapolate f outside the data window.

In practice, we have only finitely many values of $f(t)$ and these are typically noisy. We shall not address the noise problem here, except to say that it is usually handled by including regularization in the solving of each of the systems of linear equations we encounter in what follows.

The finitely many values of f, say $f(t_1), \ldots, f(t_N)$, may be obtained at irregularly spaced sample points $\{t_n\}$ but often correspond to uniformly spaced sampling points $\{t_n = a + n\Delta\}$. We consider the latter case here.

For the remainder of this section we assume that the function $F(\omega)$ is supported on the interval $[-\Omega, \Omega]$, for some $\Omega < \pi$. The sequence of Fourier coefficients of F is denoted f. Our data are the Fourier coefficients $f(n)$, for $n \in \{M, M+1, \ldots, N\}$, forming the vector d. For any function $G(\omega)$ let $\Omega G(\omega)$ be the function that equals $G(\omega)$ for $|\omega| \leqslant \Omega$ and equals zero otherwise. For any sequence of Fourier coefficients $g = \{g(n)\}$ let Dg denote the sequence whose terms are $g(n)$ for $n \in \{M, M+1, \ldots, N\}$ and zero otherwise. Let $\mathcal{F}g = G$ be the operator taking a sequence of Fourier coefficients g into the function

$$G(\omega) = \sum_{n=-\infty}^{+\infty} g(n)\exp(\mathrm{i}n\omega),$$

for $\omega \in (-\pi, \pi)$.

Let $\mathcal{H} = L^2(-\pi, \pi)$, $C_1 = L^2(-\Omega, \Omega)$ and C_2 be the set of all members $G(\omega)$ of $\mathcal{H}$ whose Fourier coefficients satisfy $g(n) = f(n)$ for $n = M, M+1, \ldots, N$. The metric projection of a function $G(\omega) \in \mathcal{H}$ onto C_1 is $\Omega G(\omega)$. The metric projection onto C_2 is implemented by passing from $G(\omega)$ to the sequence of its Fourier coefficients $\mathcal{F}^{-1}G = g$, then replacing those coefficients for $n = M, M+1, \ldots, N$ with $f(n)$ and calculating the resulting Fourier series; that is, the metric projection of G onto C_2 is $\mathcal{F}(Df + (I - D)\mathcal{F}^{-1}G)$.

We begin the GP iteration with the function $F^0(\omega) = 0$ for all $\omega \in (-\pi, \pi)$. For $k = 0, 1, \ldots$, having calculated F^k with f^k its sequence of Fourier coefficients, we define F^{k+1} by

$$F^{k+1} = \Omega\mathcal{F}(Df + (I - D)\mathcal{F}^{-1}F^k).$$

It would appear that, in order to implement this algorithm, we must calculate the entries of the sequence $\{(I - D)\mathcal{F}^{-1}F^k\}$ for all integers n not in the set $\{M, M+1, \ldots, N\}$; this is not the case, fortunately. Note that

$$F^{k+1} - F^k = \Omega\mathcal{F}D(f - f^k) = \Omega\mathcal{F}a^k,$$

where the entries of the sequence $D(f - f^k) = a^k$ are zero, except for $n = M, \ldots, N$. Since $F^0 = 0$ it follows that each F^k has the form $F^k = \Omega\mathcal{F}b^k$, for some sequence b^k with $b^k(n) = 0$ for n not in the set $\{M, M+1, \ldots, N\}$. From this we conclude that the limit F^∞ has the form

$$F^\infty(\omega) = \Omega \sum_{n=M}^{N} c_n \exp(\mathrm{i}n\omega)$$

for appropriate c_n. The coefficients c_n can then be determined by equating the Fourier coefficients of both sides of this equation. To do this we must solve the finite system of linear equations

$$f(m) = \sum_{n=M}^{N} c_n \frac{\sin\Omega(m-n)}{\pi(m-n)}, \tag{7.4}$$

where $m = M, \ldots, N$. This, of course, can also be done iteratively, if we desire.

A different approach is frequently used, resulting in a slightly different extrapolation. This second approach formulates the problem entirely in terms of finite vectors and interprets the Fourier transform as a linear transformation between finite vectors, as is done with the fast Fourier transform (FFT) algorithm.

From the discussion above we see that for an arbitrary data vector d and an arbitrary choice of the band $[-\Omega, \Omega]$ in $[-\pi, \pi]$ there is a function $F_\Omega(\omega)$ supported on $[-\Omega, \Omega]$ that is consistent with the data in the vector d. The function F_Ω has the form

$$F_\Omega(\omega) = \Omega \sum_{n=M}^{N} c_n \exp(\mathrm{i}n\omega). \tag{7.5}$$

The coefficients c_n solve the equation (7.4). To perform data extrapolation one now evaluates the Fourier transform of F_Ω at the desired points. Note that this method applies equally to uniformly and nonuniformly spaced data and is easily extended to higher dimensions. This noniterative implementation of the GP extrapolation is not new; it was presented in [23], and has been rediscovered several times since then (see [57, p 209]).

The form of the estimator in equation (7.5) suggests an extension, called the PDFT estimator, involving the use of a prior estimate, $P(\omega) \geqslant 0$, of the magnitude function $|F(\omega)|$. Suppose now that the data that comprise the vector d are the values $f(t_n)$, $n = 1, \ldots, N$, for some possibly nonequispaced points t_n. The PDFT estimate, $F_{\mathrm{PDFT}}(\omega)$, has the form

$$F_{\mathrm{PDFT}}(\omega) = P(\omega) \sum_{n=1}^{N} a_n \exp(\mathrm{i}t_n\omega), \tag{7.6}$$

where the coefficients a_n solve the system of equations $Pa = d$, with P the matrix whose entries are $p(t_m - t_n)$, the Fourier transform of $P(\omega)$ evaluated at the points $t_m - t_n$, $m, n = 1, \ldots, N$. This estimate, which can also be viewed as a data extrapolation method, was first discussed in [24]. The PDFT was applied to image processing in [25] and to phase retrieval in [27]. Variants of the PDFT that are nonlinear in the data and related to maximum entropy and maximum likelihood estimation were discussed in [26].

The PDFT estimate of F is the unique function G consistent with the data that minimizes the weighted energy

$$\int |G(\omega)|^2 P(\omega)^{-1}\, \mathrm{d}\omega.$$

If the data are equispaced and $\Omega = \pi$ then the PDFT becomes the well known discrete-time Fourier transform (DFT). The PDFT can be implemented iteratively by discretizing the function to be estimated and representing the Fourier transform by means of a matrix. The prior profile $P(\omega)$ then becomes a finite vector of weights. The ART method can then be used to calculate a minimum weighted norm solution. This approach is particularly useful for large data sets encountered in image reconstruction.

We turn now to a special case of the CFP, called the split feasibility problem.

8. The split feasibility problem

The *split feasibility problem* (SFP) [28] is to find $c \in C$ with $Ac \in Q$, if such points exist, where A is a real M by N matrix and C and Q are nonempty, closed convex sets in R^N and R^M, respectively. In [21] the CQ algorithm for solving the SFP was presented. The CQ algorithm has the iterative step

$$x^{k+1} = P_C(x^k - \gamma A^{\mathrm{T}}(I - P_Q)Ax^k), \tag{8.1}$$

where $\gamma \in (0, 2/\rho(A^{\mathrm{T}}A))$, for $\rho(A^{\mathrm{T}}A)$ the spectral radius of the matrix $A^{\mathrm{T}}A$, which is also its largest eigenvalue.

The CQ algorithm converges to a solution of the SFP, for any starting vector x^0, whenever the SFP has solutions. When the SFP has no solutions, the CQ algorithm converges to a minimizer of the function

$$f(x) = \tfrac{1}{2}\|P_Q Ax - Ax\|^2$$

over the set C, provided such constrained minimizers exist. Therefore the CQ algorithm is an iterative constrained optimization method. In fact, convergence of the CQ algorithm is a consequence of theorem 2.1.

The function $f(x)$ is convex and differentiable on R^N and its derivative is the operator

$$\nabla f(x) = A^{\mathrm{T}}(I - P_Q)Ax;$$

see [2].

Lemma 8.1. *The derivative operator ∇f is λ-Lipschitz continuous for $\lambda = \rho(A^{\mathrm{T}}A)$, therefore it is ν-ism for $\nu = \frac{1}{\lambda}$.*

Proof. We have

$$\begin{aligned}\|\nabla f(x) - \nabla f(y)\|^2 &= \|A^{\mathrm{T}}(I - P_Q)Ax - A^{\mathrm{T}}(I - P_Q)Ay\|^2 \\ &\leqslant \lambda\|(I - P_Q)Ax - (I - P_Q)Ay\|^2.\end{aligned}$$

Also

$$\begin{aligned}\|(I - P_Q)Ax - (I - P_Q)Ay\|^2 &= \|Ax - Ay\|^2 + \|P_Q Ax - P_Q Ay\|^2 \\ &\quad - 2\langle P_Q Ax - P_Q Ay, Ax - Ay\rangle\end{aligned}$$

and, since P_Q is fne,

$$\langle P_Q Ax - P_Q Ay, Ax - Ay\rangle \geqslant \|P_Q Ax - P_Q Ay\|^2.$$

Therefore,

$$\begin{aligned}\|\nabla f(x) - \nabla f(y)\|^2 &\leqslant \lambda(\|Ax - Ay\|^2 - \|P_Q Ax - P_Q Ay\|^2) \\ &\leqslant \lambda\|Ax - Ay\|^2 \leqslant \lambda^2\|x - y\|^2.\end{aligned}$$

This completes the proof. □

If $\gamma \in (0, 2/\lambda)$ then $B = P_C(I - \gamma A^{\mathrm{T}}(I - P_Q)A)$ is av and, by Dolidze's theorem 2.3, the orbit sequence $\{B^k x\}$ converges weakly to a fixed point of B, whenever such points exist. If z is a fixed point of B, then $z = P_C(z - \gamma A^{\mathrm{T}}(I - P_Q)Az)$. Therefore, for any c in C we have

$$\langle c - z, z - (z - \gamma A^{\mathrm{T}}(I - P_Q)Az)\rangle \geqslant 0.$$

This tells us that

$$\langle c - z, A^{\mathrm{T}}(I - P_Q)Az\rangle \geqslant 0,$$

which means, according to the characterization in (2.3), that z minimizes $f(x)$ relative to the set C.

The CQ algorithm employs the relaxation parameter γ in the interval $(0, 2/L)$, where L is the largest eigenvalue of the matrix $A^{\mathrm{T}}A$. Choosing the best relaxation parameter in any algorithm is a nontrivial procedure. Generally speaking, we want to select γ near to $1/L$. In practice, it would be helpful to have a quick method for estimating L. In [21] we presented such a method, particularly useful for sparse matrices.

In image reconstruction from projections the matrix A is quite large and ϵ-sparse; that is, most of its elements do not exceed ϵ in absolute value, where ϵ denotes a small positive quantity. In [21] it was shown that if A is normalized so that each row has length one, then the spectral radius of $A^{\mathrm{T}}A$ does not exceed the maximum number of nonzero elements in any column of A. A similar upper bound on $\rho(A^{\mathrm{T}}A)$ can be obtained for nonnormalized, ϵ-sparse A.

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

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

Lent has shown that the eigenvalues of the matrix $G^{\mathrm{T}}G$ do not exceed one [50]. This result suggested the following proposition, whose proof was given in [21].

Proposition 8.1. *Let A be an M by N matrix. For each $m = 1, \ldots, M$ let $v_m = \sum_{n=1}^{N} A_{mn}^2 > 0$. For each $n = 1, \ldots, N$ let $\sigma_n = \sum_{m=1}^{M} e_{mn} v_m$, 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 $A^{\mathrm{T}}A$ do not exceed σ. If A is normalized so that the Euclidean length of each of its rows is one, then the eigenvalues of $A^{\mathrm{T}}A$ do not exceed s, the maximum number of nonzero elements in any column of A.*

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

In image reconstruction from projection data that includes scattering we often encounter matrices A most of whose entries are small, if not exactly zero. A slight modification of the proof above provides us with a useful upper bound for L, the largest eigenvalue of $A^{\mathrm{T}}A$, in such cases. We assume that the rows of A have been normalized to have length one. For $\epsilon > 0$ let s be the largest number, in any column of A, of entries whose magnitudes exceed ϵ. Then we have

$$L \leqslant s + MN\epsilon^2 + 2\epsilon(MNs)^{1/2}.$$

The proof of this result is similar to that for the proposition above.

9. The Landweber algorithms

It is easy to find important examples of the SFP: if $C \subseteq R^N$ and $Q = \{b\}$ then solving the SFP amounts to solving the linear system of equations $Ax = b$; if C is a proper subset of R^N, such as the nonnegative cone, then we seek solutions of $Ax = b$ that lie within C, if there are any. The SFP is currently of some interest in dynamic PET medical image reconstruction, for reasons discussed in detail in [21]. Generally, we cannot solve the SFP in closed form and iterative methods are needed.

A number of well known iterative algorithms, such as the Landweber [46] and projected Landweber methods (see [7]), are particular cases of the CQ algorithm.

The Landweber algorithm. *With x^0 arbitrary and $k = 0, 1, \ldots$ let*

$$x^{k+1} = x^k + \gamma A^{\mathrm{T}}(b - Ax^k). \tag{9.1}$$

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*

$$x^{k+1} = P_C(x^k + \gamma A^{\mathrm{T}}(b - Ax^k)). \tag{9.2}$$

From the convergence theorem for the CQ algorithm it follows that the Landweber algorithm converges to a solution of $Ax = b$ and the projected Landweber algorithm converges to a solution of $Ax = b$ in C, whenever such solutions exist. When there are no solutions of the desired type, the Landweber algorithm converges to a least squares approximate solution of $Ax = b$, while, by corollary 4.1, the projected Landweber method will converge to a minimizer, over the set C, of the function $\|b - Ax\|$, whenever such a minimizer exists. The GP iterative procedure for bandlimited extrapolation and super-resolution is an example of the Landweber algorithm.

Another example of the Landweber method is the *simultaneous algebraic reconstruction technique* (SART) [1] for solving $Ax = b$, for nonnegative matrix A. Let A be an M by N matrix with nonnegative 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:

$$B_{ij} = A_{ij}/(A_{i+})^{1/2}(A_{+j})^{1/2},$$
$$z_j = x_j(A_{+j})^{1/2},$$

and

$$c_i = b_i/(A_{i+})^{1/2}.$$

Then the SART iterative step can be written as

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

This is a particular case of the Landweber algorithm, with $\gamma = 1$. The convergence of SART follows from theorem 2.1, once we know that the largest eigenvalue of $B^{\mathrm{T}}B$ is less than two; in fact, we showed it is one [21].

10. Generalized projections onto convex sets

There is a large amount of literature dealing with iterative algorithms based on other measures of distance between vectors; these so-called *generalized* or *Bregman* distances take the form

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

where f is a convex differentiable function endowed with other properties that permit the development of a satisfactory theory [10, 5, 29]. Given a closed nonempty convex set C contained within the domain of f and a vector x in the domain of ∇f, the projection of x onto C, relative to D_f, denoted $P_C^f x$, minimizes the function $D_f(c, x)$ over all $c \in C$. Assuming these projections can be defined, the operators P_C^f can then be used in much the same way as the metric projections to create iterative algorithms [29].

An important example of a Bregman distance is the Kullback–Leibler distance [45] for which $f(x)$ is the Shannon entropy function, defined for nonnegative vectors x by

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

The associated Bregman distance is

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

where

$$KL(x_j, z_j) = x_j \log \frac{x_j}{z_j} + z_j - x_j.$$

Projections onto convex sets C using this distance are called *entropic* projections [6, 32, 36].

To illustrate the use of entropic projections we consider the problem of finding a nonnegative solution to the system $y = Px$, where y is a vector with positive entries and P is an I by J matrix with positive entries. For each i let C_i be the set of all nonnegative vectors w with $(Pw)_i = y_i$. We cannot calculate the entropic projection onto C_i in closed form, but if we use instead the distance defined by

$$D_i(x, z) = \sum_{j=1}^{J} P_{ij} KL(x_j, z_j)$$

we find that the projection of x onto C_i relative to D_i is the vector whose entries are $x_j y_i/(Px)_i$ [18, 20, 22]. A simultaneous algorithm can then be formulated by taking at each step the weighted arithmetic mean of these projections: with $s_j = \sum_{i=1}^{I} P_{ij}$ let

$$x_j^{k+1} = x_j^k s_j^{-1} \sum_{i=1}^{I} P_{ij} y_i/(Px^k)_i.$$

This algorithm is the *expectation maximization maximum likelihood* (EMML) method widely studied in medical imaging [33, 47, 59, 48, 12–14, 52].

We can derive a second simultaneous algorithm by taking a weighted geometric mean:

$$\log x_j^{k+1} = \log x_j^k + s_j^{-1} \sum_{i=1}^{I} P_{ij} \log(y_i/(Px^k)_i).$$

This algorithm is the *simultaneous multiplicative* ART method (SMART) [42, 12].

The SMART algorithm minimizes the function $KL(Px, y)$ over all nonnegative vectors x, while the EMML algorithm minimizes the function $KL(y, Px)$ over those same vectors.

Whenever $y = Px$ has nonnegative solutions the SMART algorithm converges to that solution minimizing the function $KL(x, x^0)$; the EMML algorithm also converges to a solution, but no characterization of the limit is known.

Sequential algorithms based on these projections have also been developed. The *multiplicative* ART (MART) [39] has the following iterative step. For $k = 0, 1, \ldots$ and $i = i(k) = k(\text{mod } I) + 1$ let

$$x_j^{k+1} = x_j^k (y_i/(Px^k)_i)^{s_j^{-1} P_{ij}}.$$

Whenever nonnegative solutions of $y = Px$ exist the MART algorithm converges to the same solution as the SMART, for the same x^0. We can accelerate the convergence of MART in this case by using the *rescaled* MART (RMART) iteration [16, 17]:

$$x_j^{k+1} = x_j^k (y_i/(Px^k)_i)^{m_i^{-1} s_j^{-1} P_{ij}},$$

with m_i the maximum, over j, of the values $s_j^{-1} P_{ij}$. The SMART is clearly the simultaneous analogue of the MART, hence the name.

The accelerated sequential analogue of the EMML algorithm is the REMART method [16, 17]:

$$x_j^{k+1} = (1 - m_i^{-1} s_j^{-1} P_{ij}) x_j^k + m_i^{-1} s_j^{-1} x_j^k P_{ij} y_i/(Px^k)_i.$$

When there are nonnegative solutions of $y = Px$ the REMART converges to a solution, not necessarily the same one obtained by the EMML, even when the starting vectors are the same.

Between sequential and simultaneous methods are the so-called *block-iterative* algorithms, in which only some of the equations (or convex sets) are employed at each step of the iteration. A block-iterative version of the EMML, called the *rescaled block-iterative* EMML (RBI-EMML) [15] has been applied recently to hyperspectral imaging [53]. Related procedures are the RAMLA method of Browne and De Pierro [11] and the *ordered subset method* in [43].

11. Interior point optimization algorithms

The entropy-based methods discussed in the previous section are *interior point* methods in that the vectors that occur in the calculations always lie within the positive cone of R^J. A more general method of this sort is the *interior point algorithm* (IPA) [18–20]. The IPA is designed to minimize a convex differentiable function f over the domain of a second convex differentiable function h. The iterative step of the IPA is to solve

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

for x^{k+1}, where $\gamma > 0$ is chosen so that the function $h - \gamma f$ is convex. Note the similarities between this iterative step and the iterative step

$$x^{k+1} = P_C(x^k - \gamma \nabla f(x^k))$$

of the algorithm we considered earlier. Other conditions are required for convergence; see [20] for details. Applications of the IPA to medical imaging were discussed in [54].

12. Summary

A number of iterative algorithms used in signal processing and image reconstruction are particular cases of the KM approach to finding fixed points of ne operators on Hilbert space. These include GP bandlimited extrapolation, the Landweber methods for finding constrained solutions of linear systems of equations, simultaneous and sequential methods for solving the convex feasibility problem, the CQ algorithm for the split feasibility problem, Cimmino's method, the ART and the SART algorithm of Anderson and Kak. Similar algorithms can be developed by employing generalized POCS. Algorithms based on entropic projections, such as the EMML and MART, are obtained in this manner.

Acknowledgments

I wish to thank Heinz Bauschke, Patrick Combettes, Boris Polyak and Isao Yamada for helpful correspondences and an anonymous referee for many helpful suggestions.

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] Aubin J-P 1993 *Optima and Equilibria: an Introduction to Nonlinear Analysis* (Berlin: Springer)
[3] Baillon J and Haddad G 1977 Quelques proprietes des operateurs angle-bornes et n-cycliquement monotones *Isr. J. Math.* **26** 137–50
[4] Bauschke H and Borwein J 1996 On projection algorithms for solving convex feasibility problems *SIAM Rev.* **38** 367–426
[5] Bauschke H H and Borwein J M 1997 Legendre functions and the method of random Bregman projections *J. Convex Anal.* **4** 27–67
[6] Ben-Tal A, Charnes A and Teboulle M 1989 Entropic means *J. Math. Anal. Appl.* **139** 537–51
[7] Bertero M and Boccacci P 1998 *Introduction to Inverse Problems in Imaging* (Bristol: Institute of Physics Publishing)
[8] Bertsekas D P 1997 A new class of incremental gradient methods for least squares problems *SIAM J. Optim.* **7** 913–26
[9] Borwein J and Lewis A 2000 *Convex Analysis and Nonlinear Optimization* (*Canadian Mathematical Society Books in Mathematics*) (New York: Springer)
[10] 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
[11] Browne J and De Pierro A 1996 A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography *IEEE Trans. Med. Imaging* **15** 687–99
[12] Byrne C L 1993 Iterative image reconstruction algorithms based on cross-entropy minimization *IEEE Trans. Image Process.* **2** 96–103
[13] Byrne C L 1995 Erratum and addendum to 'Iterative image reconstruction algorithms based on cross-entropy minimization' *IEEE Trans. Image Process.* **4** 225–6
[14] Byrne C L 1996 Iterative reconstruction algorithms based on cross-entropy minimization *Image Models (and their Speech Model Cousins)* (*The IMA Volumes in Mathematics and its Applications* vol 80) ed S E Levinson and L Shepp (New York: Springer) pp 1–11
[15] Byrne C L 1996 Block-iterative methods for image reconstruction from projections *IEEE Trans. Image Process.* **5** 792–4
[16] Byrne C L 1997 Convergent block-iterative algorithms for image reconstruction from inconsistent data *IEEE Trans. Image Process.* **6** 1296–304
[17] Byrne C L 1998 Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative (RBI) methods *IEEE Trans. Image Process.* **7** 100–9
[18] Byrne C L 1999 Iterative projection onto convex sets using multiple Bregman distances *Inverse Problems* **15** 1295–313
[19] Byrne C L 2000 Block-iterative interior point optimization methods for image reconstruction from limited data *Inverse Problems* **16** 1405–19
[20] 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
[21] Byrne C L 2002 Iterative oblique projection onto convex sets and the split feasibility problem *Inverse Problems* **18** 441–53
[22] Byrne C and Censor Y 2001 Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback–Leibler distance minimization *Ann. Oper. Res.* **105** 77–98
[23] Byrne C and Fitzgerald R 1979 A unifying model for spectrum estimation *Proc. RADC Workshop on Spectrum Estimation (Griffiss AFB, Rome, NY, Oct. 1979)*
[24] Byrne C and Fitzgerald R 1982 Reconstruction from partial information, with applications to tomography *SIAM J. Appl. Math.* **42** 933–40
[25] Byrne C, Fitzgerald R, Fiddy M, Hall T and Darling A 1983 Image restoration and resolution enhancement *J. Opt. Soc. Am.* **73** 1481–7

[26] Byrne C and Fitzgerald R 1984 Spectral estimators that extend the maximum entropy and maximum likelihood methods *SIAM J. Appl. Math.* **44** 425–42
[27] Byrne C and Fiddy M 1987 Estimation of continuous object distributions from Fourier magnitude measurements *J. Opt. Soc. Am.* A **4** 412–7
[28] Censor Y and Elfving T 1994 A multiprojection algorithm using Bregman projections in a product space *Numer. Algorithms* **8** 221–39
[29] Censor Y and Zenios S A 1997 *Parallel Optimization: Theory, Algorithms and Applications* (New York: Oxford University Press)
[30] Censor Y, Iusem A N and Zenios S A 1998 An interior point method with Bregman functions for the variational inequality problem with paramonotone operators *Math. Program.* **81** 373–400
[31] Combettes P 2000 Fejér monotonicity in convex optimization *Encyclopedia of Optimization* ed C A Floudas and P M Pardalos (Boston, MA: Kluwer)
[32] De Pierro A 1991 Multiplicative iterative methods in computed tomography *Mathematical Methods in Tomography* (*Lecture Notes in Mathematics* vol 1497) ed G T Herman, A K Louis and F Natterer (New York: Springer)
[33] Dempster A P, Laird N M and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Stat. Soc.* B **37** 1–38
[34] Dolidze Z O 1982 Solution of variational inequalities associated with a class of monotone maps *Ekonomika i Matem. Metody* **18** 925–7 (in Russian)
[35] Dugundji J 1970 *Topology* (Boston: Allyn and Bacon)
[36] Eggermont P 1990 Multiplicative iterative algorithms for convex programming *Linear Algebr. Appl.* **130** 25–42
[37] Gerchberg R W 1974 Super-restoration through error energy reduction *Opt. Acta* **21** 709–20
[38] Golshtein E and Tretyakov N 1996 *Modified Lagrangians and Monotone Maps in Optimization* (New York: Wiley)
[39] 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
[40] 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
[41] Herman G T and Meyer L 1993 Algebraic reconstruction techniques can be made computationally efficient *IEEE Trans. Med. Imaging* **12** 600–9
[42] Holte S, Schmidlin P, Linden A, Rosenqvist G and Eriksson L 1990 Iterative image reconstruction for positron emission tomography: a study of convergence and quantitation problems *IEEE Trans. Nucl. Sci.* **37** 629–35
[43] Hudson H M and Larkin R S 1994 Accelerated image reconstruction using ordered subsets of projection data *IEEE Trans. Med. Imaging* **13** 601–9
[44] Kaczmarz S 1937 Angenäherte Auflösung von Systemen linearer Gleichungen *Bull. Acad. Polonaise Sci. Lett.* A **35** 355–7
[45] Kullback S and Leibler R 1951 On information and sufficiency *Ann. Math. Stat.* **22** 79–86
[46] Landweber L 1951 An iterative formula for Fredholm integral equations of the first kind *Am. J. Math.* **73** 615–24
[47] Lange K and Carson R 1984 EM reconstruction algorithms for emission and transmission tomography *J. Comput. Assist. Tomogr.* **8** 306–16
[48] 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
[49] Leahy R and Byrne C 2000 Guest editorial: recent development in iterative image reconstruction for PET and SPECT *IEEE Trans. Med. Imaging* **19** 257–60
[50] Lent A 1998 private communication
[51] Mann W 1953 Mean value methods in iteration *Proc. Am. Math. Soc.* **4** 506–10
[52] McLachlan G J and Krishnan T 1997 *The EM Algorithm and Extensions* (New York: Wiley)
[53] Meidunas E 2001 Re-scaled block iterative expectation maximization maximum likelihood (RBI-EMML) abundance estimation and sub-pixel material identification in hyperspectral imagery *MS Thesis* Department of Electrical Engineering, University of Massachusetts at Lowell, Lowell, MA
[54] Narayanan M, Byrne C and King M 2001 An interior point iterative maximum-likelihood reconstruction algorithm incorporating upper and lower bounds with application to SPECT transmission imaging *IEEE Trans. Med. Imaging* **20** 342–53
[55] Papoulis A 1975 A new algorithm in spectral analysis and band-limited extrapolation *IEEE Trans. Circuits Syst.* **22** 735–42
[56] Rockafellar R 1970 *Convex Analysis* (Princeton, NJ: Princeton University Press)
[57] Stark H and Yang Y 1998 *Vector Space Projections: a Numerical Approach to Signal and Image Processing, Neural Nets and Optics* (New York: Wiley)

[58] Tanabe K 1971 Projection method for solving a singular system of linear equations and its applications *Numer. Math.* **17** 203–14

[59] Vardi Y, Shepp L A and Kaufman L 1985 A statistical model for positron emission tomography *J. Am. Stat. Assoc.* **80** 8–20

[60] Yamada I 2001 The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications* ed D Butnariu, Y Censor and S Reich (Amsterdam: Elsevier) pp 473–504

[61] 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

[62] Zeidler E 1990 *Nonlinear Functional Analysis and its Applications: II/B-Nonlinear Monotone Operators* (Berlin: Springer)