Notes on iterative fixed point algorithms

Charles Byrne (Charles_Byrne@uml.edu)
Department of Mathematical Sciences
University of Massachusetts at Lowell
Lowell, MA 01854, USA

September 19, 2002

Abstract

An operator G on Hilbert space $\mathcal{H}$ is said to be ν-inverse strongly monotone (ν-ism) if, for some $\nu > 0$ and all x and y in $\mathcal{H}$, we have

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

For any closed convex set C in $\mathcal{H}$ let P_C denote the metric projection operator onto C. If G is ν-ism, $T = P_C(I - \gamma G)$ has fixed points and $\gamma \in (0, 2\nu)$, then the orbit sequence $\{T^k x\}$ converges weakly to a fixed point of T, for any x. Such fixed points solve the variational inequality problem: find $\hat{c} \in C$ such that $\langle G\hat{c}, c - \hat{c}\rangle \geq 0$, for all $c \in C$.

A special case is the CQ algorithm for the split feasibility problem (SFP): 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. The CQ algorithm has the iterative step

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

with $\gamma \in (0, 2/\rho(A^T A))$, where the spectral radius $\rho(A^T A)$ is the largest eigenvalue of the matrix $A^T A$. The CQ algorithm converges to a minimizer of the function $f(x) = \frac{1}{2}||P_Q Ax - Ax||^2$ over the set C, provided such constrained minimizers exist.

If N is a nonexpansive map on $\mathcal{H}$ and $B = (1-\alpha)I + \alpha N$, for some $\alpha \in (0, 1)$, then the orbit sequence $\{B^k x\}$ converges weakly to a fixed point of N, if such points exist, for any x; this is the Krasnoselski/Mann (KM) approach to finding fixed points of N. An operator T has the form of B above if and only if the map $I - T$ is $\frac{1}{2\alpha}$-ism. This connects the KM theory to that of monotone maps. Similar convergence results hold for more general paracontractive maps, but only for finite dimensional spaces.

1 Introduction and Overview

The *split feasibility problem* (SFP) [7] 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 [6] the CQ algorithm for solving the SFP was presented. The CQ algorithm has the iterative step

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

where P_C and P_Q denote the metric projections (nearest point maps) onto C and Q, respectively, and $\gamma \in (0, 2/\rho(A^TA))$, for $\rho(A^TA)$ the spectral radius of the matrix A^TA, 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 $f(x) = \frac{1}{2}||P_QAx - Ax||^2$ over the set C, provided such constrained minimizers exist. Therefore the CQ algorithm is an iterative constrained optimization method. This fact, as well as the form of the iterative step of the CQ algorithm, suggest that the CQ algorithm is related to iterative solutions of the variational inequality problem for monotone operators.

Let $\mathcal{H}$ be a real Hilbert space. An operator $G : \mathcal{H} \to \mathcal{H}$ is *monotone* [17],[19] if, for all x and y,

$$\langle Gx - Gy, x - y\rangle \geq 0. \tag{1.2}$$

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

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

and

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

Adding, we obtain

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

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

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

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

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

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

$$x^{k+1} = P_C(I - \gamma G)x^k \tag{1.3}$$

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 characterized as the unique element of C for which

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

for all $c \in C$. Therefore $\hat{c} = P_C(I - \gamma G)\hat{c}$ if and only if $\langle c - \hat{c}, \hat{c} - (\hat{c} - \gamma G\hat{c}) \rangle = \gamma \langle c - \hat{c}, G\hat{c} \rangle \geq 0$, 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 (1.3).

Say that G is λ-*Lipschitz continuous* (λ-Lip) if, for all x and y in $\mathcal{H}$,

$$||Gx - Gy|| \leq \lambda ||x - y||$$

and β-*strongly monotone* (β-sm) if

$$\langle Gx - Gy, x - y \rangle \geq \beta ||x - y||^2.$$

Then the operator $I - \gamma G$ is a strict contraction for $\gamma \in (0, \frac{2\beta}{\lambda^2})$. The operator $T = P_C(I - \gamma G)$ is then also a strict contraction and the Banach-Picard fixed point theorem guarantees that the sequence $\{T^k x\}$ converges to the unique fixed point of T for any x.

With $G = A^T(I - P_Q)A$ the iterative step in equation (1.1) has the form of equation (1.3). The function $f(x) = \frac{1}{2}||P_Q Ax - Ax||^2$ is convex and differentiable and its derivative is $\nabla f(x) = A^T(I - P_Q)Ax$. Consequently G is monotone. In a number of particular cases of interest, the CQ algorithm converges, even when the operator G is not strongly monotone and fixed points of $P_C(I - \gamma G)$ are not unique. This experience with the CQ algorithm suggests there may be weaker conditions on a general monotone operator G on $\mathcal{H}$ under which the sequence $\{T^k x\}$ will converge to a fixed point of T, provided such points exist.

In [6] it was shown that the operator $G = A^T(I - P_Q)A$ is λ-Lip for $\lambda = \rho(A^T A)$, the spectral radius of the matrix $A^T A$, which is also the largest eigenvalue of $A^T A$. In addition, we have

$$\langle Gx - Gy, x - y \rangle \geq \nu ||Gx - Gy||^2 \tag{1.4}$$

for $\nu = 1/\rho(A^TA)$ and for all x and y. Generally, say that monotone operator G on $\mathcal{H}$ is *ν-inverse strongly monotone* (ν-ism) [12] if , for some $\nu > 0$, G satisfies the inequality (1.4) for all x and y. The condition given in inequality (1.4) is the desired weaker condition. In fact, we have the following result.

Theorem 1.1 *Let $G : \mathcal{H} \to \mathcal{H}$ be a ν-inverse strongly monotone operator and $T = P_C(I - \gamma G)$. Then the sequence $\{T^k x\}$ converges weakly to a fixed point of T, if such points exist, for all x and any $\gamma \in (0, 2\nu)$.*

The convergence of the CQ algorithm is then a consequence of Theorem 1.1.

Standard references on the VIP do not give this theorem, but it is not new. Golshtein and Tretyakov [12] include this result, for the finite dimensional case, and credit Dolidze [9]. The surprising thing is that Theorem 1.1 is also a consequence of a much better known result in the theory of fixed points for nonexpansive operators.

An operator $N : \mathcal{H} \to \mathcal{H}$ is *nonexpansive* (ne) if, for all x and y, we have

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

Note that if N is ne then $I - N$ is monotone. The literatures on monotone operators and nonexpansive maps seem to have relatively small intersection, with the result that discoveries in one area are not quickly made part of the other. In the study of monotone operators and the VIP, the focus is on the properties of the monotone operator G, not those of $I - \gamma G$, while, for those investigating nonexpansive maps, N is the focus, not its complement $I - N$.

The fixed point set Fix(N) of the nonexpansive map $N = I$, the identity map, is all of $\mathcal{H}$ and $\{N^k x\}$ converges in a single step to a fixed point of N, for every x. In contrast, the ne map $N = -I$ has Fix(N)= $\{0\}$ and $\{N^k x\}$ fails to converge for any nonzero x. Therefore, the existence of fixed points for a ne map N is not sufficient for (weak) convergence of the orbit sequence $\{N^k x\}$. A simple modification of the iterative step remedies the situation, however.

The Krasnoselski/Mann (KM) approach is to include relaxation in the iterative step. Define

$$Bx = (1 - \alpha)x + \alpha Nx, \tag{1.5}$$

for some $\alpha \in (0, 1)$. Then we have the following result:

Theorem 1.2 *The orbit sequence $\{B^k x\}$ converges weakly to a fixed point of N, if such points exist, for any x.*

In these notes we shall look more closely at operators B as defined by equation (1.5), the goal being to demonstrate that Theorem 1.1 is a special case of Theorem 1.2.

In the next section we examine the CQ algorithm in more detail. Then we pass on to nonexpansive maps and the several important subclasses of ne maps. We examine the proof of convergence associated with the KM approach, to see how the properties of the involved operators play a role. In the final sections we look at similar fixed point methods that apply to the more general class of paracontractive maps, using Korpolevich's solution of VIP(G, C) to provide an important example. Wherever possible, we shall present results for infinite dimensional $\mathcal{H}$; where not yet possible we shall consider briefly the issues involved in extending the results to such spaces. In the KM approach the relaxation parameters can be permitted to change with each iteration; in the Appendix 1 we include proof of convergence for this more general scheme.

2 The SFP and the CQ algorithm

Let $C_1, ..., C_M$ be closed nonempty convex subsets of Hilbert space $\mathcal{H}$. The *convex feasibility problem* (CFP) is to find a member of their intersection, if such elements exist [3]. One approach is to cycle through these sets repeatedly, applying the metric projection onto the current set at each step. The SFP is a special case of the CFP, with $C_1 = C$ and $C_2 = A^{-1}(Q)$. In applications the metric projections P_C and P_Q are usually easily calculated, while the metric projection onto $A^{-1}(Q)$ is not. The CQ algorithm was originally obtained using one oblique and one metric projection, along with a generalized form of relaxation, according to the theorem given in [5].

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 [6] it was shown that if A is normalized so that each row has length one, then the spectral radius of A^TA does not exceed the maximum number of nonzero elements in any column of A. A similar upper bound on $\rho(A^TA)$ can be obtained for non-normalized, ϵ-sparse A (see Appendix 2). The proof of convergence of the CQ algorithm reveals that the larger the γ the faster the convergence, generally speaking, so reasonably good upper bounds on the spectral radius should helpful in accelerating convergence for large, ϵ-sparse problems.

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 [6]. 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 and projected Landweber methods (see [4]) and the Gerchberg-Papoulis bandlimited extrapolation (or super-resolution) technique, are particular cases of the CQ algorithm.

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

$$x^{k+1} = x^k + \gamma A^T(b - Ax^k). \tag{2.1}$$

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 $k = 0, 1, ...$ let

$$x^{k+1} = P_C(x^k + \gamma A^T(b - Ax^k)). \tag{2.2}$$

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 aproximate solution of $Ax = b$, while the projected Landweber method will converge to a minimizer, over the set C, of the function $||b - Ax||$, whenever such a minimizer exists. Examples of the Landweber method include the Gerchberg-Papoulis iterative procedure for band-limited extrapolation and super-resolution [11], [16] and 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

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

This is a particular case of the Landweber algorithm, with $\gamma = 1$. The convergence of SART will follow from Theorem 1.1, 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 $u = (u_1, ..., u_N)^T$ with $u_j = (A_{+j})^{1/2}$ and $v = (v_1, ..., v_M)^T$, 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.

It turns out that the CQ algorithm is itself a particular case of a more general iterative method for finding fixed points of nonexpansive operators on Hilbert space. Relating these algorithms to each other is the main topic of these notes.

3 Varieties of nonexpansive operators

A nonexpansive operator N on $\mathcal{H}$ need not have any fixed points, as the example $Nx = x + x_0$, for $x_0 \neq 0$ shows. An operator S on $\mathcal{H}$ is a *σ-strict contraction* (σ-sc) if there is $\sigma \in (0, 1)$ such that, for all x and y, $||Sx - Sy|| \leq \sigma ||x - y||$. If S is σ-sc then, according to the Banach-Picard theorem, the orbit sequence $\{S^k x\}$ converges to the unique fixed point of S, for any x. Another result along these lines is Schauder's fixed point theorem [10]: if T is continuous on $\mathcal{H}$ and its range is contained within a compact set, then T has a fixed point.

In the Krasnoselski/Mann approach to finding fixed points of a ne operator N we can allow the relaxation parameters to vary with each step:

Theorem 3.1 *(Combettes [8], Proposition 9) Let N be a nonexpansive operator on $\mathcal{H}$. For $k = 0, 1, 2, \ldots$, let $\alpha_k \in (0, 1)$ and let the sequence $\{x^k\}$ be defined by*

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

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

The proof of this theorem is given in the Appendix.

For any operator T on $\mathcal{H}$ we have the identity

$$||x - y||^2 - ||Tx - Ty||^2 = 2\langle (I - T)x - (I - T)y, x - y\rangle \, - ||(I - T)x - (I - T)y||^2$$

from which we see immediately that $T = N$ is ne if and only if $G = I - N$ is $\frac{1}{2}$-ism. We say that an ne operator N has *associated parameter value* $\alpha \in (0, 1]$ if $I - N$ is $\frac{1}{2\alpha}$-ism. This is equivalent to having $N = (1 - \alpha)I + \alpha M$ for some ne operator M. So every ne operator has $\alpha = 1$ as an associated parameter value; it may have a lower associated value, however.

An operator $F : \mathcal{H} \to \mathcal{H}$ is called *firmly nonexpansive* (fne) if, for every x and y in $\mathcal{H}$, we have

$$\langle Fx - Fy, x - y\rangle \, \geq ||Fx - Fy||^2.$$

An operator $R : \mathcal{H} \to \mathcal{H}$ is called *relaxed firmly nonexpansive* (rfne) if there is a fne operator F and some $\alpha \in (0, 1)$ such that $R = (1 - \alpha)I + \alpha F$.

Theorem 3.1 involves relaxed nonexpansive operators of the form $B = (1 - \alpha)I + \alpha N$ for some ne operator N and $\alpha \in (0, 1)$; such operators are called *averaged* (av) [3]. The class of averaged operators is a larger class than that of firmly nonexpansive operators. It is not difficult to show that the product of finitely many averaged operators is again averaged.

The following characterizations are easily demonstrated and are helpful in this regard. The operator N is nonexpansive (respectively, averaged, firmly nonexpansive, relaxed firmly nonexpansive) if and only if N has an associated parameter value of $\alpha = 1$ (respectively, < 1, $\leq \frac{1}{2}$, $< \frac{1}{2}$).

Let C be a closed, convex nonempty set in $\mathcal{H}$ and P_C the associated metric projection operator. As noted earlier, for each $x \in \mathcal{H}$ the projection $P_C x$ is characterized as the unique element of C for which

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

for all $c \in C$. It follows that, for any x and y in $\mathcal{H}$,

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

that is, $P_C(\cdot)$ is a fne operator. However, over-relaxed metric projections $A_C = (1-\beta)I + \beta P_C$ are averaged but not generally firmly nonexpansive when $\beta \in (1, 2)$. For example, let $\beta = 1.9$ and $C = \{0\}$. Since P_C is fne, it is averaged. Therefore, if G is ν-ism and $\gamma \in (0, 2\nu)$, the product $P_C(I - \gamma G)$ is averaged.

4 A convergence proof for KM iteration

In this section we examine the proof of the convergence theorem for the KM approach to finding fixed points of ne maps, to appreciate the role played by the properties of B and $G = I - B$.

Theorem 4.1 *Let N be a nonexpansive operator on Hilbert space $\mathcal{H}$ with Fix(N) nonempty. Let $\alpha \in (0, 1)$ and $B = (1-\alpha)I + \alpha N$. Let x^0 be arbitrary and $x^{k+1} = Bx^k$. Then the sequence $\{x^k\}$ converges weakly to a fixed point of N.*

Proof: Let $Nz = z$. With $G = I - B$ we have

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

Since G is $\frac{1}{2\alpha}$-ism it follows that

$$||z - x^k||^2 - ||z - x^{k+1}||^2 \geq (\frac{1}{\alpha} - 1)||Gz - Gx^k||^2.$$

Since $||Gz-Gx^k|| = ||x^k-x^{k+1}||$, we have that $\{||z-x^k||\}$ is a decreasing sequence and the sequence $\{||x^k-x^{k+1}||\}$ converges to zero. The sequence $\{x^k\}$ is then bounded, so we can extract a weak cluster point, x^*. If $\mathcal{H}$ is finite dimensional, then we conclude that $Bx^* = x^*$, so, by using x^* in place of z, we have that $\{x^k\}$ converges to x^*. If $\mathcal{H}$ is infinite dimensional, however, we need to work a bit harder. Since B is ne, G is *demi-closed*, which means here that $Gx^* = 0$; see the Appendix for details. Since $\{||z - x^k||\}$ is decreasing, we say that $\{x^k\}$ is Fejér monotone with respect to the set Fix(N). For such sequences, if every weak cluster point is in Fix(N), then the sequence converges weakly. This completes the proof. ∎

Corollary 4.1 *Let G be ν-ism and $\gamma \in (0, 2\nu)$. Let $T = P_C(I - \gamma G)$. Then the orbit sequence $\{T^k x\}$ converges for any x to a solution of VIP(G, C), whenever such solutions exist.*

Corollary 4.2 *The sequence produced by the CQ algorithm converges to a minimizer of* $||P_Q Ax - Ax||$*, relative to* $x \in C$*, whenever such relative minimizers exist, for* $\gamma \in (0, 2/\rho(A^T A))$*.*

A close look at the proof of the theorem above shows that we have not used the full power of the definition of G being ν-ism; we only considered the case in which one of the two vectors is a fixed point. In the next section we consider going beyond nonexpansive maps in finding fixed points.

5 Beyond nonexpansive maps

Say that an operator $P : \mathcal{H} \to \mathcal{H}$ is *paracontractive* (pc) if, for every z such that $Pz = z$, we have $Px = x$ or

$$||z - Px|| < ||z - x||.$$

Consider the orbit sequence $\{x^k\}$ defined by $x^{k+1} = Px^k$. Mimicking the proof in the previous section, and assuming that $x^{k+1} \neq x^k$, we have

$$||z - x^k|| - ||z - x^{k+1}|| > 0,$$

so that $\{||z - x^k||\}$ is decreasing and has limit $d(z) \geq 0$. The sequence $\{x^k\}$ is bounded, so let x^* be a weak cluster point. If $\mathcal{H}$ is finite dimensional, then $||z - x^*|| = d(z) = ||z - Px^*||$, so that $Px^* = x^*$. Using x^* in place of z, we conclude that $\{x^k\}$ converges to x^*. If $\mathcal{H}$ is infinite dimensional, we cannot argue in this manner, since P need not be weakly continuous. Since P need not be nonexpansive, $I - P$ need not be demi-closed, so we cannot complete the proof as we did for the KM approach.

An interesting example of a pc operator occurs in Korpolevich's method [13] for solving the VIP(G, C) in the finite dimensional case. If G is monotone and λ-Lip, but not sm or ism, Korpolevich uses the two-step iteration defined as follows: with $\gamma = 1/\lambda$ let

$$y^k = P_C(x^k - \gamma G x^k)$$

and

$$x^{k+1} = P_C(x^k - \gamma G y^k).$$

We consider here only the case in which $C = R^N$. Then the operator $S = \gamma G$ is a strict contraction, with $||Sx - Sy|| \leq \sigma ||x - y||$ for $\sigma = \gamma\lambda$. The operator $T = I - S(I - S)$ is then paracontractive, as we now show.

Lemma 5.1 *Let $Tz = z$. Then $Sz = 0$ and*

$$2\langle S(I-S)x, x-z\rangle \geq ||S(I-S)x||^2 + (1-\sigma^2)||Sx||^2.$$

Proof: Since $Tz = z$, it follows immediately that $S(I-S)z = 0$. Then

$$||Sz|| = ||S(I-S)z - Sz|| \leq \sigma||(I-S)z - z|| = \sigma||Sz||,$$

from which we conclude that $Sz = 0$. Now we have

$$\begin{aligned} 2\langle S(I-S)x, x-z\rangle &= 2\langle S(I-S)x - S(I-S)z, x - Sx - z + Sz\rangle + 2\langle S(I-S)x, Sx\rangle \\ &\geq 2\langle S(I-S)x, Sx\rangle, \end{aligned}$$

because S is monotone. Also

$$\sigma^2||Sx||^2 \geq ||S(I-S)x - Sx||^2 = ||S(I-S)x||^2 - 2\langle S(I-S)x, Sx\rangle + ||Sx||^2.$$

The result follows by combining these two inequalities. ∎

Lemma 5.2 *For all x we have $\sigma^2||Sx||^2 \geq (||Sx|| - ||S(I-S)x||)^2$.*

Proof: By Cauchy's inequality we have

$$2||Sx||||S(I-S)x|| \geq 2\langle S(I-S)x, Sx\rangle \geq (1-\sigma^2)||Sx||^2 + ||S(I-S)x||^2.$$

The result follows. ∎

Lemma 5.3 *For all x we have $||S(I-S)x|| \geq (1-\sigma)||Sx||$.*

Proof: If $||Sx|| \geq ||S(I-S)x||$ then, by Lemma 5.2 we have $\sigma||Sx|| \geq ||Sx|| - ||S(I-S)x||$. If $||Sx|| < ||S(I-S)x||$ then $||S(I-S)x|| \geq ||Sx|| \geq (1-\sigma)||Sx||$. ∎

Lemma 5.4 *For all x we have $(1+\sigma)||Sx|| \geq ||S(I-S)x|| \geq (1-\sigma)||Sx||$.*

Proof: The second inequality was established in Lemma 5.3, so we need only prove the first one. If $||Sx|| - ||S(I-S)x|| < 0$ then $\sigma||Sx|| \geq ||S(I-S)x|| - ||Sx||$. If $||Sx|| \geq ||S(I-S)x||$ then $(1+\sigma)||Sx|| \geq ||Sx|| \geq ||S(I-S)x||$. ∎

Combining these inequalities, we find that the operator $S(I-S)$ satisfies a weak form of inverse strong monotonicity. Proceeding then as in the proof of Theorem 1.2 we obtain the following result.

Proposition 5.1 *Let S be monotone and σ-sc. Let z be a fixed point of the operator $T = I - S(I - S)$. Then*

$$2\langle S(I-S)x, x-z\rangle \geq \frac{2}{1+\sigma}||S(I-S)x||^2.$$

Therefore the operator T is paracontractive.

If G is the gradient operator for a convex differentiable function and G is λ-Lip, then, as Baillon and Haddad [2] have shown, G is $\frac{1}{\lambda}$-ism, so we do not need Korpolevich's approach in this case.

6 Summary

The CQ algorithm was designed originally to solve the split feasibility problem. At the suggestion of a referee, I was led to consider its behavior in the case in which the SFP had no solution and discovered that the CQ algorithm minimized $||P_Q Ax - Ax||$, over x in C, whenever such minimizers exist. So the CQ algorithm is an iterative constrained optimization method. This fact, along with the form of the CQ algorithm, led me to consider possible connections with the variational inequality problem and, eventually, to a strengthening of an iterative VIP method. This stronger VIP result, involving inverse strongly monotonicity, turned out to be known, but not well. Surprisingly, this result for the VIP can be reformulated as a special case of a much better known convergence theorem for nonexpansive maps, the Krasnoselski/Mann theorem.

7 Appendix 1: Proof of Theorem 3.1

To prove the theorem we need a couple of preliminary results. Although the operator $I - N$ is strongly continuous, it may not be weakly continuous. It does, however, have a useful, but somewhat weaker, property called *demi-closed*[15]. Specifically, we have

Lemma 7.1 *If the sequence $\{z^k\}$ converges weakly to z and $\{(I-N)z^k\}$ converges strongly to zero, then $(I-N)z = 0$ or $Nz = z$.*

For a proof see [15] or [18].

Let $S \subseteq \mathcal{H}$. A sequence $\{x^k\}$ is Fejér monotone with respect to the set S if the sequence $\{||s - x^k||\}$ is non-increasing for every $s \in S$. The following result is Proposition 2 of [8].

Lemma 7.2 *Let the sequence $\{x^k\}$ be Fejér monotone with respect to the set S. If $\{x^k\}$ has weak cluster points and all such points belong to S then $\{x^k\}$ converges weakly to a member of S.*

For every operator T we have the identity

$$||x-y||^2 - ||Tx-Ty||^2 = 2\langle (I-T)x-(I-T)y, x-y\rangle - ||(I-T)x-(I-T)y||^2.$$

From this we have

Lemma 7.3 *The operator N is nonexpansive if and only if the operator $I-N$ has the property*

$$\langle (I-N)z-(I-N)x, z-x\rangle \geq \frac{1}{2}||(I-N)z-(I-N)x||^2.$$

Finally, we have

Lemma 7.4 *Let the sequence $\{x^k\}$ be defined by*

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

for $k=0,1,...$ and $\alpha_k \in (0,1)$. Then the sequence $\{||(I-N)x^k||\}$ is nonincreasing.

Proof: We have

$$||(I-N)x^k||^2 = ||(I-N)x^k - (I-N)x^{k+1} + (I-N)x^{k+1}||^2$$

$$= ||(I-N)x^{k+1}||^2 + 2\langle (I-N)x^k-(I-N)x^{k+1}, (I-N)x^{k+1}\rangle + ||(I-N)x^k-(I-N)x^{k+1}||^2.$$

We show that

$$2\langle (I-N)x^k-(I-N)x^{k+1}, (I-N)x^{k+1}\rangle + ||(I-N)x^k-(I-N)x^{k+1}||^2 \geq 0,$$

or, equivalently,

$$2\langle (I-N)x^k-(I-N)x^{k+1}, (I-N)x^{k}\rangle - ||(I-N)x^k-(I-N)x^{k+1}||^2 \geq 0.$$

Equivalently, using $(I-N)x^k = 1/\alpha_k(x^k - x^{k+1})$ on the left side of the inequality above, we must show that

$$\frac{2}{\alpha_k}\langle (I-N)x^k-(I-N)x^{k+1}, x^k - x^{k+1}\rangle \geq ||(I-N)x^k-(I-N)x^{k+1}||^2.$$

But this follows from Lemma 7.3, since $\alpha_k \in (0,1)$. ∎

We turn now to the proof of Theorem 3.1.

Proof: Let z be a fixed point of N. For each k let $H_k = \alpha_k(I - N)x^k$. Then a simple calculation shows that

$$||z - x^k||^2 - ||z - x^{k+1}||^2 \geq 2\langle(H_k z - H_k x^k, z - x^k\rangle - ||H_k z - H_k x^k||^2 \geq 0.$$

That the right side above is nonnegative follows from Lemma 7.3 and the definition of H_k. The sequence $\{x^k\}$ is then Fejér monotone with respect to the set Fix(N). Using Lemma 7.3 again we have

$$||z - x^k||^2 - ||z - x^{k+1}||^2 \geq \alpha_k(1 - \alpha_k)||x^k - Nx^k||^2. \tag{7.1}$$

Therefore the sequence $\{x^k\}$ is bounded, so has weak cluster points. Summing both sides of the inequality (7.1) with respect to k, we see that

$$\sum_{k=1}^{\infty} \alpha_k(1 - \alpha_k)||x^k - Nx^k||^2 < +\infty.$$

Since $\{||x^k - Nx^k||\}$ is a nonincreasing sequence we conclude that $||x^k - Nx^k|| \to 0$.

Let x^* be an arbitrary weak cluster point of $\{x^k\}$ and let the subsequence $\{x^{k_j}\}$ converge weakly to x^*. Since $\{(I - N)x^{k_j}\} \to 0$, it follows from Lemma 7.1that $Nx^* = x^*$. Since x^* was arbitrarily chosen, every weak cluster point of the sequence lies in Fix(N). Since the sequence $\{x^k\}$ is Fejér monotone with respect to Fix(N), we conclude from Lemma 7.2 that $\{x^k\}$ converges weakly to a member of Fix(N). ▮

8 Appendix 2: A helpful eigenvalue bound

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. 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 appendix we present such a method, particularly useful for sparse matrices.

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

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

Lent has shown that the eigenvalues of the matrix $G^T G$ do not exceed one [14]. The following proposition and its proof are straightforward extensions of Lent's work.

Proposition 8.1 *Let A be an M by N matrix. For each $m = 1, ..., M$ let $\nu_m = \sum_{n=1}^N A_{mn}^2 > 0$. For each $n = 1, ..., N$ 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 $A^T Av = cv$ for some nonzero vector v. We show that $c \leq \sigma$. We have $AA^T Av = cAv$ 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

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

Therefore,

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

We also have

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

The result follows immediately. ∎

If we normalize A so that its rows have length one, then the trace of the matrix AA^T is $\text{tr}(AA^T) = M$, 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 $s << M$.

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 $A^T A$, in such cases. For $\epsilon > 0$ let σ be the largest number of entries in any column of A whose magnitudes exceed ϵ. Then we have

$$L \leq \sigma + MN\epsilon^2 + 2\epsilon(MN\sigma)^{1/2}.$$

Acknowledgments

I wish to thank Heinz Bauschke, Patrick Combettes, Boris Polyak and Isao Yamada for helpful correspondences.

References

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

[2] Baillon, J., and Haddad, G. (1977) Quelques proprietes des operateurs angle-bornes et n-cycliquement monotones, *Israel J. of Mathematics*, **26**, pp. 137-150.

[3] Bauschke, H., and Borwein, J. (1996) On projection algorithms for solving convex feasibility problems, *SIAM Review*, **38 (3)**, pp. 367–426.

[4] Bertero, M. and Boccacci, P. (1998) *Introduction to Inverse Problems in Imaging*, Institute of Physics Publishing, Bristol, UK.

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

[6] Byrne, C. (2002) Iterative oblique projection onto convex sets and the split feasibility problem, *Inverse Problems*, **18**, pp. 441-453.

[7] Censor, Y. and Elfving, T. (1994) A multiprojection algorithm using Bregman projections in a product space, *Numerical Algorithms*, **8**, pp. 221–239.

[8] Combettes, P. (2000) Fejér monotonicity in convex optimization, in *Encyclopedia of Optimization*, C.A. Floudas and P. M. Pardalos, Eds., Kluwer Publ., Boston, MA .

[9] Dolidze, Z.O. (1982) Solution of variational inequalities associated with a class of monotone maps, *Ekonomika i Matem. Metody*, **18 (5)**, pp. 925–927 (in Russian).

[10] Dugundji, J. (1970) *Topology*, Allyn and Bacon, Inc., Boston.

[11] Gerchberg, R. W. (1974) Super-restoration through error energy reduction, *Optica Acta*, **21**, pp. 709–720.

[12] Golshtein, E., and Tretyakov, N. (1996) *Modified Lagrangians and Monotone Maps in Optimization*, John Wiley, NY.

[13] Korpolevich, G. (1976) The extragradient method for finding saddle points and other problems, *Ekonomika i Matem. Metody*, **12**, pp. 747–756 (in Russian).

[14] Lent, A. (1998) private communication.

[15] Opial, Z. (1967) Weak convergence of the sequence of successive approximations for nonexpansive mappings, *Bull. Amer. Math. Soc.*, **73**, pp. 591–597.

[16] Papoulis, A. (1975) A new algorithm in spectral analysis and band-limited extrapolation, *IEEE Transactions on Circuits and Systems*, **22**, pp. 735–742.

[17] Rockafellar, R. (1970) *Convex Analysis*, Princeton University Press.

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

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