Bounds on the Largest Singular Value of a Matrix and the Convergence of Simultaneous and Block-Iterative Algorithms for Sparse Linear Systems

Charles Byrne (Charles_Byrne@uml.edu)
Department of Mathematical Sciences
University of Massachusetts Lowell
Lowell, MA 01854, USA
http://faculty.uml.edu/cbyrne/cbyrne.html

January 21, 2009

Abstract

We obtain the following upper bounds for the eigenvalues of the matrix $A^\dagger A$. For any a in the interval $[0,2]$ let

$$c_{aj} = \sum_{i=1}^{I} |A_{ij}|^a,$$

$$r_{ai} = \sum_{j=1}^{J} |A_{ij}|^{2-a},$$

and c_a and r_a the maxima of the c_{aj} and r_{ai}, respectively. Then no eigenvalue of the matrix $A^\dagger A$ exceeds the maximum of

$$\sum_{j=1}^{J} c_{aj}|A_{ij}|^{2-a},$$

over all i, nor the maximum of

$$\sum_{i=1}^{I} r_{ai}|A_{ij}|^a,$$

over all j. Therefore, no eigenvalue of $A^\dagger A$ exceeds $c_a r_a$.

Using these bounds, it follows that, for the matrix G with entries

$$G_{ij} = A_{ij}\sqrt{\alpha_i}\sqrt{\beta_j},$$

no eigenvalue of $G^\dagger G$ exceeds one, provided that, for some a in the interval $[0, 2]$, we have

$$\alpha_i \leq r_{ai}^{-1},$$

and

$$\beta_j \leq c_{aj}^{-1}.$$

Using this result, we obtain convergence theorems for several iterative algorithms for solving the problem $Ax = b$, including the CAV, BICAV, CARP1, SART, SIRT, and the block-iterative DROP and SART methods.

This article will appear in *International Transactions in Operations Research*, in 2009.

1 Introduction and Notation

We are concerned here with iterative methods for solving, at least approximately, the system of I linear equations in J unknowns symbolized by $Ax = b$. In the applications of interest to us, such as medical imaging, both I and J are quite large, making the use of iterative methods the only feasible approach. It is also typical of such applications that the matrix A is sparse, that is, has relatively few non-zero entries. Therefore, iterative methods that exploit this sparseness to accelerate convergence are of special interest to us.

The *algebraic reconstruction technique* (ART) of Gordon, et al. [12] is a *sequential* method; at each step only one equation is used. The current vector x^{k-1} is projected orthogonally onto the hyperplane corresponding to that single equation, to obtain the next iterate x^k. The iterative step of the ART is

$$x_j^k = x_j^{k-1} + \overline{A_{ij}}\Big(\frac{b_i - (Ax^{k-1})_i}{\sum_{t=1}^J |A_{it}|^2}\Big), \tag{1.1}$$

where $i = k(\mathrm{mod}\, I)$. The sequence $\{x^k\}$ converges to the solution closest to x^0 in the consistent case, but only converges subsequentially to a limit cycle in the inconsistent case.

Cimmino's method [10] is a *simultaneous* method, in which all the equations are used at each step. The current vector x^{k-1} is projected orthogonally onto each of the hyperplanes and these projections are averaged to obtain the next iterate x^k. The iterative step of Cimmino's method is

$$x_j^k = \frac{1}{I}\sum_{i=1}^I \Big(x_j^{k-1} + \overline{A_{ij}}\Big(\frac{b_i - (Ax^{k-1})_i}{\sum_{t=1}^J |A_{it}|^2}\Big)\Big),$$

which can also be written as

$$x_j^k = x_j^{k-1} + \sum_{i=1}^I \overline{A_{ij}}\Big(\frac{b_i - (Ax^{k-1})_i}{I\sum_{t=1}^J |A_{it}|^2}\Big). \tag{1.2}$$

Landweber's iterative scheme [16](see also [3, 4, 5] with

$$x^k = x^{k-1} + B^\dagger(d - Bx^{k-1}), \tag{1.3}$$

converges to the least-squares solution of $Bx = d$ closest to x^0, provided that the largest singular value of B does not exceed one. If we let B be the matrix with entries

$$B_{ij} = A_{ij}/\sqrt{I\sum_{t=1}^{J}|A_{it}|^2},$$

and define

$$d_i = b_i/\sqrt{I\sum_{t=1}^{J}|A_{it}|^2},$$

then, since the trace of the matrix $BB^\dagger$ is one, convergence of Cimmino's method follows. However, using the trace in this way to estimate the largest singular value of a matrix usually results in an estimate that is far too large, particularly when A is large and sparse, and therefore in an iterative algorithm with unnecessarily small step sizes.

The appearance of the term

$$I\sum_{t=1}^{J}|A_{it}|^2$$

in the denominator of Equation (1.2) suggested to Censor et al. [8] that, when A is sparse, this denominator might be replaced with

$$\sum_{t=1}^{J}s_t|A_{it}|^2,$$

where s_t denotes the number of non-zero entries in the tth column of A. The resulting iterative method is the *component-averaging* (CAV) iteration. Convergence of the CAV method was established by showing that no singular value of the matrix B exceeds one, where B has the entries

$$B_{ij} = A_{ij}/\sqrt{\sum_{t=1}^{J}s_t|A_{it}|^2}.$$

In this paper we extend a result of van der Sluis and van der Vorst [18] to obtain upper bounds on the eigenvalues of the matrix $A^\dagger A$; as a corollary, we have that no eigenvalue of $A^\dagger A$ exceeds the maximum of the numbers

$$p_i = \sum_{t=1}^{J}s_t|A_{it}|^2.$$

Convergence of CAV then follows, as does convergence of several other methods, including the ART, Landweber's method, the SART [1], the block-iterative CAV (BICAV) [9], the CARP1 method of Gordon and Gordon [13], and a block-iterative variant of CARP1 obtained from the DROP method of Censor et al. [7]. Convergence of most of these methods was also established in [15], using a unifying framework of a block-iterative Landweber algorithm, but without deriving upper bounds for the largest eigenvalue of a general $A^\dagger A$.

For a positive integer N with $1 \leq N \leq I$, we let $B_1, ..., B_N$ be not necessarily disjoint subsets of the set $\{i = 1, ..., I\}$; the subsets B_n are called *blocks*. We then let A_n be the matrix and b^n the vector obtained from A and b, respectively, by removing all the rows except for those whose index i is in the set B_n. For each n, we let s_{nt} be the number of non-zero entries in the tth column of the matrix A_n, s_n the maximum of the s_{nt}, s the maximum of the s_t, and $L_n = \rho(A_n^\dagger A_n)$ be the spectral radius, or largest eigenvalue, of the matrix $A_n^\dagger A_n$, with $L = \rho(A^\dagger A)$. We denote by A_i the ith row of the matrix A, and by ν_i the length of A_i, so that

$$\nu_i^2 = \sum_{j=1}^{J} |A_{ij}|^2.$$

2 Some Upper Bounds for L

For the iterative algorithms we shall consider here, having a good upper bound for the largest eigenvalue of the matrix $A^\dagger A$ is important. In the applications of interest, principally medical image processing, the matrix A is large; even calculating $A^\dagger A$, not to mention computing eigenvalues, is prohibitively expensive. In addition, the matrix A is typically sparse, but $A^\dagger A$ will not be, in general. In this section we present upper bounds for L that are particularly useful when A is sparse and do not require the calculation of $A^\dagger A$.

In [18] van der Sluis and van der Vorst show that certain rescaling of the matrix A results in none of the eigenvalues of $A^\dagger A$ exceeding one. A modification of their proof leads to upper bounds on the eigenvalues of the original $A^\dagger A$. For any a in the interval $[0, 2]$ let

$$c_{aj} = c_{aj}(A) = \sum_{i=1}^{I} |A_{ij}|^a,$$

$$r_{ai} = r_{ai}(A) = \sum_{j=1}^{J} |A_{ij}|^{2-a},$$

and c_a and r_a the maxima of the c_{aj} and r_{ai}, respectively. We prove the following theorem.

Theorem 2.1 *For any a in the interval $[0,2]$, no eigenvalue of the matrix $A^\dagger A$ exceeds the maximum of*

$$\sum_{j=1}^{J} c_{aj}|A_{ij}|^{2-a},$$

over all i, nor the maximum of

$$\sum_{i=1}^{I} r_{ai}|A_{ij}|^{a},$$

over all j. Therefore, no eigenvalue of $A^\dagger A$ exceeds $c_a r_a$.

Proof: Let $A^\dagger A v = \lambda v$, and let $w = Av$. Then we have

$$\|A^\dagger w\|^2 = \lambda \|w\|^2.$$

Applying Cauchy's Inequality, we obtain

$$\Big|\sum_{i=1}^{I} \overline{A_{ij}} w_i\Big|^2 \leq \Big(\sum_{i=1}^{I} |A_{ij}|^{a/2}|A_{ij}|^{1-a/2}|w_i|\Big)^2$$

$$\leq \Big(\sum_{i=1}^{I} |A_{ij}|^{a}\Big)\Big(\sum_{i=1}^{I} |A_{ij}|^{2-a}|w_i|^2\Big).$$

Therefore,

$$\|A^\dagger w\|^2 \leq \sum_{j=1}^{J} \Big(c_{aj}(\sum_{i=1}^{I} |A_{ij}|^{2-a}|w_i|^2)\Big) = \sum_{i=1}^{I} \Big(\sum_{j=1}^{J} c_{aj}|A_{ij}|^{2-a}\Big)|w_i|^2$$

$$\leq \max_i \Big(\sum_{j=1}^{J} c_{aj}|A_{ij}|^{2-a}\Big)\|w\|^2.$$

The remaining two assertions follow in similar fashion. ∎

The following corollary is central to our discussion.

Corollary 2.1 *For each $i = 1, 2, ..., I$, let*

$$p_i = \sum_{j=1}^{J} s_j |A_{ij}|^2,$$

and let p be the maximum of the p_i. Then $L \leq p$.

Proof: Take $a = 0$. Then, using the convention that $0^0 = 0$, we have $c_{0j} = s_j$. ∎

Corollary 2.2 *Selecting $a = 1$, we have*

$$L = \|A\|_2^2 \leq \|A\|_1 \|A\|_\infty = c_1 r_1.$$

Corollary 2.3 *Selecting $a = 2$, we have*

$$L = \|A\|_2^2 \leq \|A\|_F^2,$$

where $\|A\|_F$ denotes the Frobenius norm of A, which is the Euclidean norm of the vectorized A.

Corollary 2.4 *Let G be the matrix with entries*

$$G_{ij} = A_{ij}\sqrt{\alpha_i}\sqrt{\beta_j},$$

where

$$\alpha_i \leq \Big(\sum_{j=1}^{J} s_j \beta_j |A_{ij}|^2\Big)^{-1},$$

for all i. Then $\rho(G^\dagger G) \leq 1$.

Proof: We have

$$\sum_{j=1}^{J} s_j |G_{ij}|^2 = \alpha_i \sum_{j=1}^{J} s_j \beta_j |A_{ij}|^2 \leq 1,$$

for all i. The result follows from Corollary 2.1. ∎

Corollary 2.5 *If $\sum_{j=1}^{J} s_j |A_{ij}|^2 \leq 1$ for all i, then $L \leq 1$.*

Corollary 2.6 *If $0 < \gamma_i \leq p_i^{-1}$ for all i, then the matrix B with entries $B_{ij} = \sqrt{\gamma_i} A_{ij}$ has $\rho(B^\dagger B) \leq 1$.*

Proof: We have

$$\sum_{j=1}^{J} s_j |B_{ij}|^2 = \gamma_i \sum_{j=1}^{J} s_j |A_{ij}|^2 = \gamma_i p_i \leq 1.$$

Therefore, $\rho(B^\dagger B) \leq 1$, according to the theorem. ∎

Corollary 2.7 *([2]; [17], Th. 4.2) If $\sum_{j=1}^{J} |A_{ij}|^2 = 1$ for each i, then $L \leq s$.*

Proof: For all i we have

$$p_i = \sum_{j=1}^{J} s_j |A_{ij}|^2 \leq s \sum_{j=1}^{J} |A_{ij}|^2 = s.$$

Therefore,

$$L \leq p \leq s.$$

∎

Corollary 2.8 *If, for some a in the interval $[0,2]$, we have*

$$\alpha_i \le r_{ai}^{-1}, \tag{2.1}$$

for each i, and

$$\beta_j \le c_{aj}^{-1}, \tag{2.2}$$

for each j, then, for the matrix G with entries

$$G_{ij} = A_{ij}\sqrt{\alpha_i}\sqrt{\beta_j},$$

no eigenvalue of $G^\dagger G$ exceeds one.

Proof: We calculate $c_{aj}(G)$ and $r_{ai}(G)$ and find that

$$c_{aj}(G) \le \Big(\max_i \alpha_i^{a/2}\Big)\beta_j^{a/2}\sum_{i=1}^I |A_{ij}|^a = \Big(\max_i \alpha_i^{a/2}\Big)\beta_j^{a/2}c_{aj}(A),$$

and

$$r_{ai}(G) \le \Big(\max_j \beta_j^{1-a/2}\Big)\alpha_i^{1-a/2}r_{ai}(A).$$

Therefore, applying the inequalities (2.1) and (2.2), we have

$$c_{aj}(G)r_{ai}(G) \le 1,$$

for all i and j. Consequently, $\rho(G^\dagger G) \le 1$. ∎

The next theorem ([2]) provides another upper bound for L that is useful when A is sparse. For each i and j, we let $e_{ij} = 1$, if A_{ij} is not zero, and $e_{ij} = 0$, if $A_{ij} = 0$. Let $0 < \nu_i = \sqrt{\sum_{j=1}^J |A_{ij}|^2}$, $\sigma_j = \sum_{i=1}^I e_{ij}\nu_i^2$, and σ be the maximum of the σ_j.

Theorem 2.2 *([2]) No eigenvalue of $A^\dagger A$ exceeds σ.*

Proof: Let $A^\dagger Av = cv$, for some non-zero vector v and scalar c. With $w = Av$, we have

$$w^\dagger AA^\dagger w = cw^\dagger w.$$

Then

$$\Big|\sum_{i=1}^I \overline{A_{ij}}w_i\Big|^2 = \Big|\sum_{i=1}^I \overline{A_{ij}}e_{ij}\nu_i\frac{w_i}{\nu_i}\Big|^2 \le \Big(\sum_{i=1}^I |A_{ij}|^2\frac{|w_i|^2}{\nu_i^2}\Big)\Big(\sum_{i=1}^I \nu_i^2 e_{ij}\Big)$$
$$= \Big(\sum_{i=1}^I |A_{ij}|^2\frac{|w_i|^2}{\nu_i^2}\Big)\sigma_j \le \sigma\Big(\sum_{i=1}^I |A_{ij}|^2\frac{|w_i|^2}{\nu_i^2}\Big).$$

Therefore, we have

$$cw^\dagger w = w^\dagger AA^\dagger w = \sum_{j=1}^J\Big|\sum_{i=1}^I \overline{A_{ij}}w_i\Big|^2 \le \sigma\sum_{j=1}^J\Big(\sum_{i=1}^I |A_{ij}|^2\frac{|w_i|^2}{\nu_i^2}\Big) = \sigma\sum_{i=1}^I |w_i|^2 = \sigma w^\dagger w.$$

We conclude that $c \le \sigma$. ∎

Corollary 2.9 *Let the rows of A have Euclidean length one. Then no eigenvalue of $A^\dagger A$ exceeds the maximum number of non-zero entries in any column of A.*

Proof: We have $\nu_i^2 = \sum_{j=1}^J |A_{ij}|^2 = 1$, for each i, so that $\sigma_j = s_j$ is the number of non-zero entries in the jth column of A, and $\sigma = s$ is the maximum of the σ_j. ▮

When the rows of A have length one, it is easy to see that $L \leq I$, so the choice of $\gamma = \frac{1}{I}$ in the Landweber algorithm, which gives Cimmino's algorithm [10], is acceptable, although perhaps much too small.

The proof of Theorem 2.2 is based on results presented by Arnold Lent in informal discussions with Gabor Herman, Yair Censor, Rob Lewitt and me at MIPG in Philadelphia in the late 1990's.

3 The Basic Convergence Theorem

The following theorem is a basic convergence result concerning block-iterative algorithms.

Theorem 3.1 *Let $L_n \leq 1$, for $n = 1, 2, ..., N$. If the system $Ax = b$ is consistent, then, for any starting vector x^0, and with $n = n(k) = k(\mathrm{mod}\, N)$ and $\lambda_k \in [\epsilon, 2-\epsilon]$ for all k, the sequence $\{x^k\}$ with iterative step*

$$x^k = x^{k-1} + \lambda_k A_n^\dagger (b^n - A_n x^{k-1}) \tag{3.1}$$

converges to the solution of $Ax = b$ for which $\|x - x^0\|$ is minimized.

We begin with the following lemma.

Lemma 3.1 *Let T be any (not necessarily linear) operator on R^J, and $S = I - T$, where I denotes the identity operator. Then, for any x and y, we have*

$$\|x - y\|^2 - \|Tx - Ty\|^2 = 2\langle Sx - Sy, x - y\rangle - \|Sx - Sy\|^2. \tag{3.2}$$

The proof is a simple calculation and we omit it here.

Proof of Theorem 3.1: Let $Az = b$. Applying Equation (3.2) to the operator

$$Tx = x + \lambda_k A_n^\dagger (b^n - A_n x),$$

we obtain

$$\|z - x^{k-1}\|^2 - \|z - x^k\|^2 = 2\lambda_k \|b^n - A_n x^{k-1}\|^2 - \lambda_k^2 \|A_n^\dagger b^n - A_n^\dagger A_n x^{k-1}\|^2. \tag{3.3}$$

Since $L_n \leq 1$, it follows that

$$\|A_n^\dagger b^n - A_n^\dagger A_n x^{k-1}\|^2 \leq \|b^n - A_n x^{k-1}\|^2.$$

Therefore,

$$\|z - x^{k-1}\|^2 - \|z - x^k\|^2 \geq (2\lambda_k - \lambda_k^2)\|b^n - A_n x^{k-1}\|^2,$$

from which we draw several conclusions:

- the sequence $\{\|z - x^k\|\}$ is decreasing;
- the sequence $\{\|b^n - A_n x^{k-1}\|\}$ converges to zero.

In addition, for fixed $n = 1, ..., N$ and $m \to \infty$,

- the sequence $\{\|b^n - A_n x^{mN+n-1}\|\}$ converges to zero;
- the sequence $\{x^{mN+n}\}$ is bounded.

Let $x^{*,1}$ be a cluster point of the sequence $\{x^{mN+1}\}$; then there is subsequence $\{x^{m_r N+1}\}$ converging to $x^{*,1}$. The sequence $\{x^{m_r N+2}\}$ is also bounded, and we select a cluster point $x^{*,2}$. Continuing in this fashion, we obtain cluster points $x^{*,n}$, for $n = 1, ..., N$. From the conclusions reached previously, we can show that $x^{*,n} = x^{*,n+1} = x^*$, for $n = 1, 2, ..., N-1$, and $Ax^* = b$. Replacing the generic solution $\hat{x}$ with the solution x^*, we see that the sequence $\{\|x^* - x^k\|\}$ is decreasing. But, subsequences of this sequence converge to zero, so the entire sequence converges to zero, and so $x^k \to x^*$.

Now we show that x^* is the solution of $Ax = b$ that minimizes $\|x - x^0\|$. Since $x^k - x^{k-1}$ is in the range of $A^\dagger$ for all k, so is $x^* - x^0$, from which it follows that x^* is the solution minimizing $\|x - x^0\|$. Another way to get this result is to use Equation (3.3). Since the right side of Equation (3.3) is independent of the choice of solution, so is the left side. Summing both sides over the index k reveals that the difference

$$\|x - x^0\|^2 - \|x - x^*\|^2$$

is independent of the choice of solution. Consequently, minimizing $\|x - x^0\|$ over all solutions x is equivalent to minimizing $\|x - x^*\|$ over all solutions x; the solution to the latter problem is clearly $x = x^*$. ∎

4 Simultaneous Iterative Algorithms

In this section we apply the previous theorems to obtain convergence of several simultaneous iterative algorithms for linear systems.

4.1 The General Simultaneous Iterative Scheme

In this section we are concerned with simultaneous iterative algorithms having the following iterative step:

$$x_j^k = x_j^{k-1} + \lambda_k \sum_{i=1}^{I} \gamma_{ij} \overline{A_{ij}} (b_i - (Ax^{k-1})_i), \tag{4.1}$$

with $\lambda_k \in [\epsilon, 1]$ and the choices of the parameters γ_{ij} that guarantee convergence. Although we cannot prove convergence for this most general iterative scheme, we are able to prove the following theorems for the separable case of $\gamma_{ij} = \alpha_i \beta_j$.

Theorem 4.1 *If, for some a in the interval $[0, 2]$, we have*

$$\alpha_i \le r_{ai}^{-1}, \tag{4.2}$$

for each i, and

$$\beta_j \le c_{aj}^{-1}, \tag{4.3}$$

for each j, then the sequence $\{x^k\}$ given by Equation (4.1) converges to the minimizer of the proximity function

$$\sum_{i=1}^{I} \alpha_i |b_i - (Ax)_i|^2$$

for which

$$\sum_{j=1}^{J} \beta_j^{-1} |x_j - x_j^0|^2$$

is minimized.

Proof: For each i and j, let

$$G_{ij} = \sqrt{\alpha_i}\sqrt{\beta_j} A_{ij},$$

$$z_j = x_j / \sqrt{\beta_j},$$

and

$$d_i = \sqrt{\alpha_i} b_i.$$

Then $Ax = b$ if and only if $Gz = d$. From Corollary 2.8 we have that $\rho(G^\dagger G) \le 1$. Convergence then follows from Theorem 3.1. ∎

Corollary 4.1 *Let $\gamma_{ij} = \alpha_i \beta_j$, for positive α_i and β_j. If*

$$\alpha_i \leq \Big(\sum_{j=1}^{J} s_j \beta_j |A_{ij}|^2 \Big)^{-1}, \tag{4.4}$$

for each i, then the sequence $\{x^k\}$ in (4.1) converges to the minimizer of the proximity function

$$\sum_{i=1}^{I} \alpha_i |b_i - (Ax)_i|^2$$

for which

$$\sum_{j=1}^{J} \beta_j^{-1} |x_j - x_j^0|^2$$

is minimized.

Proof: We know from Corollary 2.4 that $\rho(G^\dagger G) \leq 1$. ∎

4.2 Some Convergence Results

We obtain convergence for several known algorithms as corollaries to the previous theorems.

The SIRT Algorithm:

Corollary 4.2 *([18]) For some a in the interval $[0, 2]$ let $\alpha_i = r_{ai}^{-1}$ and $\beta_j = c_{aj}^{-1}$. Then the sequence $\{x^k\}$ in (4.1) converges to the minimizer of the proximity function*

$$\sum_{i=1}^{I} \alpha_i |b_i - (Ax)_i|^2$$

for which

$$\sum_{j=1}^{J} \beta_j^{-1} |x_j - x_j^0|^2$$

is minimized.

For the case of $a = 1$, the iterative step becomes

$$x_j^k = x_j^{k-1} + \sum_{i=1}^{I} \Bigg(\frac{\overline{A_{ij}}(b_i - (Ax^{k-1})_i)}{(\sum_{t=1}^{J} |A_{it}|)(\sum_{m=1}^{I} |A_{mj}|)} \Bigg),$$

which was considered in [14]. The SART algorithm [1] is a special case, in which it is assumed that $A_{ij} \geq 0$, for all i and j.

The CAV Algorithm:

Corollary 4.3 *If $\beta_j = 1$ and α_i satisfies*

$$0 < \alpha_i \leq \Big(\sum_{j=1}^{J} s_j |A_{ij}|^2 \Big)^{-1},$$

for each i, then the algorithm with the iterative step

$$x^k = x^{k-1} + \lambda_k \sum_{i=1}^{I} \alpha_i (b_i - (Ax^{k-1})_i) A_i^\dagger \tag{4.5}$$

converges to the minimizer of

$$\sum_{i=1}^{I} \alpha_i |b_i - (Ax^{k-1})_i|^2$$

for which $\|x - x^0\|$ is minimized.

When

$$\alpha_i = \Big(\sum_{j=1}^{J} s_j |A_{ij}|^2 \Big)^{-1},$$

for each i, this is the relaxed *component-averaging* (CAV) method of Censor et al. [8].

The Landweber Algorithm: When $\beta_j = 1$ and $\alpha_i = \alpha$ for all i and j, we have the relaxed Landweber algorithm. The convergence condition in Equation (2.1) becomes

$$\alpha \leq \Big(\sum_{j=1}^{J} s_j |A_{ij}|^2 \Big)^{-1} = p_i^{-1}$$

for all i, so $\alpha \leq p^{-1}$ suffices for convergence. Actually, the sequence $\{x^k\}$ converges to the minimizer of $\|Ax - b\|$ for which the distance $\|x - x^0\|$ is minimized, for any starting vector x^0, when $0 < \alpha < 1/L$. Easily obtained estimates of L are usually over-estimates, resulting in overly conservative choices of α. For example, if A is first normalized so that $\sum_{j=1}^{J} |A_{ij}|^2 = 1$ for each i, then the trace of $A^\dagger A$ equals I, which tells us that $L \leq I$. But this estimate, which is the one used in Cimmino's method [10], is far too large when A is sparse.

The Simultaneous DROP Algorithm:

Corollary 4.4 *Let* $0 < w_i \leq 1$,

$$\alpha_i = w_i \nu_i^{-2} = w_i \Big(\sum_{j=1}^{J} |A_{ij}|^2 \Big)^{-1}$$

and $\beta_j = s_j^{-1}$, for each i and j. Then the simultaneous algorithm with the iterative step

$$x_j^k = x_j^{k-1} + \lambda_k \sum_{i=1}^{I} \left(\frac{w_i \overline{A_{ij}}(b_i - (Ax^{k-1})_i)}{s_j \nu_i^2} \right), \tag{4.6}$$

converges to the minimizer of the function

$$\sum_{i=1}^{I} \left| \frac{w_i(b_i - (Ax)_i)}{\nu_i} \right|^2$$

for which the function

$$\sum_{j=1}^{J} s_j |x_j - x_j^0|^2$$

is minimized.

For $w_i = 1$, this is the CARP1 algorithm of [13] (see also [11, 8, 9]). The simultaneous DROP algorithm of [7] requires only that the weights w_i be positive, but dividing each w_i by their maximum, $\max_i\{w_i\}$, while multiplying each λ_k by the same maximum, gives weights in the interval $(0, 1]$. For convergence of their algorithm, we need to replace the condition $\lambda_k \leq 2 - \epsilon$ with $\lambda_k \leq \frac{2-\epsilon}{\max_i\{w_i\}}$.

The denominator in CAV is

$$\sum_{t=1}^{J} s_t |A_{it}|^2,$$

while that in CARP1 is

$$s_j \sum_{t=1}^{J} |A_{it}|^2.$$

It was reported in [13] that the two methods differed only slightly in the simulated cases studied.

5 Block-iterative Algorithms

The methods discussed in the previous section are *simultaneous*, that is, all the equations are employed at each step of the iteration. We turn now to *block-iterative methods*, which employ only some of the equations at each step. When the parameters are appropriately chosen, block-iterative methods can be significantly faster than simultaneous ones.

5.1 The Block-Iterative Landweber Algorithm

For a given set of blocks, the block-iterative Landweber algorithm has the following iterative step: with $n = k(\mathrm{mod}\, N)$,

$$x^k = x^{k-1} + \gamma_n A_n^\dagger (b^n - A_n x^{k-1}). \tag{5.1}$$

The sequence $\{x^k\}$ converges to the solution of $Ax = b$ that minimizes $\|x - x^0\|$, whenever the system $Ax = b$ has solutions, provided that the parameters γ_n satisfy the inequalities $0 < \gamma_n < 1/L_n$. This follows from Theorem 3.1 by replacing the matrices A_n with $\sqrt{\gamma_n} A_n$ and the vectors b^n with $\sqrt{\gamma_n} b^n$.

If the rows of the matrices A_n are normalized to have length one, then we know that $L_n \le s_n$. Therefore, we can use parameters γ_n that satisfy

$$0 < \gamma_n \le \Big(s_n \sum_{j=1}^J |A_{ij}|^2 \Big)^{-1}, \tag{5.2}$$

for each $i \in B_n$.

5.2 The BICAV Algorithm

We can extend the block-iterative Landweber algorithm as follows: let $n = k(\mathrm{mod}\, N)$ and

$$x^k = x^{k-1} + \lambda_k \sum_{i \in B_n} \gamma_i (b_i - (Ax^{k-1})_i) A_i^\dagger. \tag{5.3}$$

It follows from Theorem 2.1 that, in the consistent case, the sequence $\{x^k\}$ converges to the solution of $Ax = b$ that minimizes $\|x - x^0\|$, provided that, for each n and each $i \in B_n$, we have

$$\gamma_i \le \Big(\sum_{j=1}^J s_{nj} |A_{ij}|^2 \Big)^{-1}.$$

The BICAV algorithm [9] uses

$$\gamma_i = \Big(\sum_{j=1}^J s_{nj} |A_{ij}|^2 \Big)^{-1}.$$

The iterative step of BICAV is

$$x^k = x^{k-1} + \lambda_k \sum_{i \in B_n} \left(\frac{b_i - (Ax^{k-1})_i}{\sum_{t=1}^J s_{nt} |A_{it}|^2} \right) A_i^\dagger. \tag{5.4}$$

5.3 A Block-Iterative CARP1

The obvious way to obtain a block-iterative version of CARP1 would be to replace the denominator term

$$s_j \sum_{t=1}^{J} |A_{it}|^2$$

with

$$s_{nj} \sum_{t=1}^{J} |A_{it}|^2.$$

However, this is problematic, since we cannot redefine the vector of unknowns using $z_j = x_j \sqrt{s_{nj}}$, since this varies with n. In [7], this issue is resolved by taking τ_j to be not less than the maximum of the s_{nj}, and using the denominator

$$\tau_j \sum_{t=1}^{J} |A_{it}|^2 = \tau_j \nu_i^2.$$

A similar device is used in [15] to obtain a convergent block-iterative version of SART. The iterative step of DROP is

$$x_j^k = x_j^{k-1} + \lambda_k \sum_{i \in B_n} \left(\overline{A_{ij}} \frac{(b_i - (Ax^{k-1})_i)}{\tau_j \nu_i^2} \right). \tag{5.5}$$

Convergence of the DROP (*diagonally-relaxed orthogonal projection*) iteration follows from their Theorem 11. We obtain convergence as a corollary of our previous results.

The change of variables is $z_j = x_j \sqrt{\tau_j}$, for each j. Using our eigenvalue bounds, it is easy to show that the matrices C_n with entries

$$(C_n)_{ij} = \left(\frac{A_{ij}}{\sqrt{\tau_j} \nu_i} \right),$$

for all $i \in B_n$ and all j, have $\rho(C_n^\dagger C_n) \leq 1$. The resulting iterative scheme, which is equivalent to Equation (5.5), then converges, whenever $Ax = b$ is consistent, to the solution minimizing the proximity function

$$\sum_{i=1}^{I} \left| \frac{b_i - (Ax)_i}{\nu_i} \right|^2$$

for which the function

$$\sum_{j=1}^{J} \tau_j |x_j - x_j^0|^2$$

is minimized.

Acknowledgments

I wish to thank Professor Dan Gordon, Department of Computer Science, University of Haifa, for numerous helpful conversations.

References

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

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

[3] Byrne, C. (2004) "A unified treatment of some iterative algorithms in signal processing and image reconstruction."*Inverse Problems* **20**, pp. 103–120.

[4] Byrne, C. (2005) *Signal Processing: A Mathematical Approach*, AK Peters, Publ., Wellesley, MA.

[5] Byrne, C. (2007) *Applied Iterative Methods*, AK Peters, Publ., Wellesley, MA.

[6] Byrne, C. (2009) "Block-iterative algorithms." *International Transactions in Operations Research*, to appear.

[7] Censor, Y., Elfving, T., Herman, G.T., and Nikazad, T. (2008) "On diagonally-relaxed orthogonal projection methods." *SIAM Journal on Scientific Computation*, **30(1)**, pp. 473–504.

[8] Censor, Y., Gordon, D., and Gordon, R. (2001) "Component averaging: an efficient iterative parallel algorithm for large and sparse unstructured problems." *Parallel Computing*, **27**, pp. 777–808.

[9] Censor, Y., Gordon, D., and Gordon, R. (2001) "BICAV: A block-iterative, parallel algorithm for sparse systems with pixel-related weighting." *IEEE Transactions on Medical Imaging*, **20**, pp. 1050–1060.

[10] Cimmino, G. (1938) "Calcolo approssimato per soluzioni dei sistemi di equazioni lineari."*La Ricerca Scientifica XVI, Series II, Anno IX* **1**, pp. 326–333.

[11] Dines, K., and Lyttle, R. (1979) "Computerized geophysical tomography." *Proc. IEEE*, **67**, pp. 1065–1073.

[12] Gordon, R., Bender, R., and Herman, G.T. (1970) "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography."*J. Theoret. Biol.* **29**, pp. 471–481.

[13] Gordon, D., and Gordon, R.(2005) "Component-averaged row projections: A robust block-parallel scheme for sparse linear systems." *SIAM Journal on Scientific Computing*, **27**, pp. 1092–1117.

[14] Hager, B., Clayton, R., Richards, M., Comer, R., and Dziewonsky, A. (1985) "Lower mantle heterogeneity, dynamic typography and the geoid." *Nature*, **313**, pp. 541–545.

[15] Jiang, M., and Wang, G. (2003) "Convergence studies on iterative algorithms for image reconstruction." *IEEE Transactions on Medical Imaging*, **22(5)**, pp. 569–579.

[16] Landweber, L. (1951) "An iterative formula for Fredholm integral equations of the first kind."*Amer. J. of Math.* **73**, pp. 615–624.

[17] van der Sluis, A. (1969) "Condition numbers and equilibration of matrices." *Numer. Math.*, **14**, pp. 14–23.

[18] van der Sluis, A., and van der Vorst, H.A. (1990) "SIRT- and CG-type methods for the iterative solution of sparse linear least-squares problems." *Linear Algebra and its Applications*, **130**, pp. 257–302.