
ELSEVIER Linear Algebra and its Applications 349 (2002) 125–154

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

A probabilistic analysis of asynchronous iteration

John C. Strikwerda

Computer Sciences Department, University of Wisconsin at Madison, Madison, WI 53706, USA

Received 7 March 2001; accepted 29 November 2001

Submitted by R.A. Brualdi

Abstract

We present and analyze a probabilistic model for asynchronous iteration of linear systems. The model is similar in spirit to the chaotic model proposed by Chazan and Miranker in 1969, but with the choice of components and delays being based on probability distributions. We give sufficient conditions and necessary conditions for the expected value of the error to converge to zero. In addition we give sufficient conditions for the variance of the error to converge to zero. These conditions are all weaker than the strong condition of Chazan and Miranker. We also give numerical results of simulations illustrating the theoretical results. © 2002 Published by Elsevier Science Inc.

AMS classification: 65F10

Keywords: Asynchronous iteration; Parallel computing

1. Introduction

In their seminal paper [3] Chazan and Miranker studied chaotic relaxation, now usually called asynchronous iteration, for the solution of linear systems. In asynchronous iteration the order in which components of the solution are updated is not fixed and the selection of previous values of the components used in the updates is also not fixed. This is intended to be a model for parallel computation in which different processors work independently and have access to data values in either shared or distributed memory.

E-mail address: strikwerda@cs.wisc.edu (J.C. Strikwerda).

0024-3795/02/$ - see front matter
PII: S0024-3795(02)00258-6

Frommer and Szyld [4] have an excellent survey of asynchronous iterations and the mathematical theory of the models developed to date. They also discuss the applicability of results about linear processes to nonlinear processes. We refer readers to [4] for references to recent applications of asynchronous methods and related literature. Other references of interest are [1,2].

The model of Chazan and Miranker is interesting because of the great generality allowed and because there is a simple necessary and sufficient condition for the system to converge for chaotic updates. Their analysis was based on determining conditions under which every one of the possible sequences of iterations is convergent.

In this paper we consider a probabilistic model for asynchronous iteration. The advantage of this model, we believe, is that it gives more reasonable conditions for the convergence of the solution in practical situations.

By considering the probability of different events, extreme cases that are non-convergent, but occur with negligible probability, are not given significant influence on the convergence criteria. Thus if the circumstances that cause non-convergence occur with negligible probability, then the expected result could be a convergent process. As we show, there are matrices which have non-convergent events according to the Chazan and Miranker model, but which are convergent according to the probabilistic model.

The structure of this paper is as follows. We first present the model for asynchronous iteration for a shared memory model. This model is the easiest to analyze of the methods considered in this paper. We consider both the expectation of the error and the variance of the error. Both necessary and sufficient conditions for the convergence of these quantities are given. We next consider a model for a distributed memory model, then we look at under or over-relaxation. Finally, we present some numerical results that illustrate the results of this paper and present some conclusions.

2. A probabilistic model for asynchronous iteration

We consider the solution of linear systems of the form

$$(I - B)\vec{x} = \vec{d}, \tag{2.1}$$

where I is the N by N identity matrix, B is an N by N matrix with real entries, $\vec{d}$ is the data vector, and $\vec{x}$ is the vector of unknowns. Both $\vec{x}$ and $\vec{d}$ have N components. We assume that the matrix $I - B$ is non-singular. (See [8] for information on the case when $I - B$ is singular.) The iteration begins with an initial vector $\vec{x}^0$ and proceeds by computing successive iterates $\vec{x}^\nu$. Successive iterates are obtained by updating different components at each step.

In asynchronous iteration, the choice of which component of the iterate $\vec{x}^\nu$ is updated next is not fixed. In our model this choice is governed by a probabilistic

process. Moreover, the iterates that are used to update this component need not be the values computed most recently. Due to latency in computer networks, the past iterates may be out of date by some amount. We model the delay in the data usage by another probabilistic process.

Our model is intended to simulate distributed computation such as arises on networks of computers, including connections using the Internet. Because of network delays, variations in the speed of the computers, computer availability, and similar issues, the tasks assigned to the different computers take different amounts of time to complete. We consider the situation where there is no attempt to synchronize the computation.

We do not address the question of whether asynchronous computation is a reasonable strategy for computation. That question would require an analysis that uses the results that are obtained here. One part of the computation that we do not need to model here is the length of time between the successive updates. This would have to be done to compare computational methods since computational speed refers to the time of the computation, not the number of iterations, however they are counted.

We now describe our model in some detail. This model is for a shared or common memory model as illustrated in Fig. 1. The several processors obtain data from the common memory and return values to the memory. There is no attempt to synchronize the computation. In an actual implementation of such an iterative process there would be one processor which checks on the convergence behavior of the method.

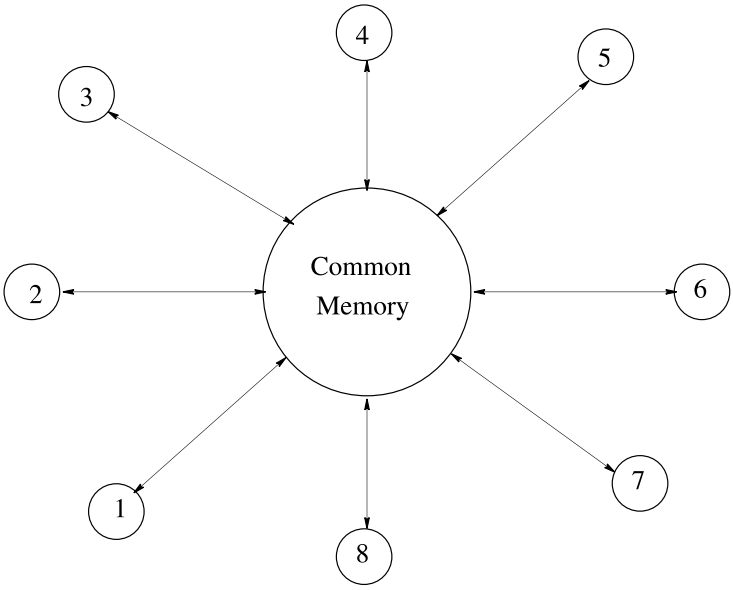

Fig. 1. Illustration of shared memory model.

The iteration counter ν is increased by 1 each time the data are read from the common memory. We assume that at each step the data used for the update come from the solution vector stored in common memory at some previous step. Also, only one component is updated per step.

The probability that the ℓth component is updated on any given step is p_ℓ. We also consider the choice of delay to be probabilistic, that is, the probability that the delay is q steps in the past is s_q. We assume that both of these probability distributions are independent of the step ν and that the delay probabilities are independent of the component ℓ. Then we may express the model with these formulas:

$$x_\ell^{\nu+1} = \sum_{m=1}^{N} b_{\ell,m} x_m^{\nu-q} + d_\ell \quad \text{with probability } p_\ell s_q,$$
$$x_\ell^{\nu+1} = x_\ell^{\nu} \quad \text{with probability } 1 - p_\ell.$$

Notice that the probability that component ℓ is updated with delay q is $p_\ell s_q$ and the updated value of the component uses the solution at step $\nu - q$.

Also note that the components x_ℓ^ν need not be single components, but could be collections of components. Much of our analysis applies equally to the case of single components or collections of components. For simplicity, we consider in detail only the case of single components.

The following relations obviously must be satisfied for the model to be well-defined:

$$\sum_{\ell=1}^{N} p_\ell = 1, \quad \sum_{q=0}^{\infty} s_q = 1. \tag{2.2}$$

In addition, we require each of the probabilities p_ℓ to be non-zero.

In comparison with the Chazan and Miranker model, we note that they strictly enforced that each component is updated infinitely often. Here we rely on the results of probability theory that with probability 1 each component will be updated infinitely often. Also, in the Chazan and Miranker model it is required that no delay may exceed a certain limit, here we allow arbitrary delays but the greater delays occur with less probability.

Having defined the model, we now begin the analysis of the model. Without loss of generality we take $\vec{d} = \vec{0}$ so that $\vec{x}$ is the error in the iteration. We also assume that $\vec{x}^\nu = \vec{x}^0$ for all negative values of ν.

The analysis depends on the recurrence relation for the probability density functions for the error at each step. We work with the probability density functions $\beta^\nu(\vec{x})$ for $\vec{x} \in R^N$. That is, the probability that the error $\vec{x}^\nu$ is in a measurable set G at step ν is given by the integral

$$\text{Prob}(\vec{x}^\nu \in G) = \int_G \beta^\nu(\vec{y})\, \mathrm{d}y.$$

The probability distribution at step $\nu+1$ is related to the probability densities at previous steps and depends on the update probabilities of the components and the delay probabilities. In order to compute the probability distribution at the new time level, we need the following integration result.

Lemma 2.1. *Integration on a hyperplane. Consider the integral of a function $f(\cdot)$ on the hyperplane given by $\vec{b}\cdot\vec{y}=z$ with the natural (inherited) measure on this plane μ. In addition, consider the function $g(z)$ integrated over all values of z. Then the following formula holds*:

$$\int g(z)\int_{\vec{b}\cdot\vec{y}=z} f(\vec{y})\,\mathrm{d}\mu\,\mathrm{d}z = \|\vec{b}\|\int g(\vec{b}\cdot\vec{x})f(\vec{x})\,\mathrm{d}x.$$

Proof. First, make an orthogonal change of coordinates in $\vec{y}$ so that the unit vector in the direction of $\vec{b}$ is a coordinate direction, say the first coordinate. Call this new coordinate vector $\vec{w}$ and let Q be the orthogonal matrix for this change of coordinates, such that $\vec{y}=Q\vec{w}$. Then $\vec{b}\cdot\vec{y}=z$ becomes $\|\vec{b}\|w_1=z$. After rearranging, change back to the original coordinates.

$$\begin{aligned}
&\int g(z)\int_{\vec{b}\cdot\vec{y}=z} f(\vec{y})\,\mathrm{d}\mu\,\mathrm{d}z\\
&\quad=\int g(z)\int_{\|\vec{b}\|w_1=z} f(Q\vec{w})\,\mathrm{d}w_2\cdots\mathrm{d}w_N\,\mathrm{d}z\\
&\quad=\|\vec{b}\|\int_{w_1} g(\|\vec{b}\|w_1)\int_{w_2}\int_{w_3}\cdots\int_{w_N} f(Q\vec{w})\,\mathrm{d}w_2\cdots\mathrm{d}w_N\,\mathrm{d}w_1\\
&\quad=\|\vec{b}\|\int g(\vec{b}\cdot\vec{y})f(\vec{y})\,\mathrm{d}y. \qquad \square
\end{aligned}$$

From this result, using $g(\cdot)\equiv 1$, we have that the probability distribution on the plane $\vec{b}\cdot\vec{y}=z$ that is inherited from a distribution $f(\cdot)$ is

$$f_{\vec{b}}(z)=\frac{1}{\|\vec{b}\|}\int_{\vec{b}\cdot\vec{y}=z} f(\vec{y})\,\mathrm{d}y.$$

We now proceed with deriving the recursion relation for asynchronous iteration. Given that the component being updated is ℓ and the delay is q, then the probability distribution is equal to the product of the distribution for the new component and the distribution for the other components. The distribution for x_ℓ is

$$\frac{1}{\|\vec{b}_\ell\|}\int_{(B\vec{y})_\ell=x_\ell}\beta^{\nu-q}(\vec{y})\,\mathrm{d}y,$$

where $(B\vec{y})_\ell=\vec{b}_\ell\cdot\vec{y}$ and $\vec{b}_\ell$ is the vector that forms the ℓth row of the matrix B. The distribution for the other components is

$$\int_{-\infty}^{\infty} \beta^{\nu}(\vec{x})\ \mathrm{d}x_{\ell},$$

that is, only the ℓth component is integrated.

Putting all this together, we obtain the relation

$$\beta^{\nu+1}(\vec{x}) = \sum_{\ell,q} \frac{p_{\ell} s_q}{\|\vec{b}_{\ell}\|} \int_{(B\vec{y})_{\ell}=x_{\ell}} \beta^{\nu-q}(\vec{y})\ \mathrm{d}y \int_{-\infty}^{\infty} \beta^{\nu}(\vec{x})\ \mathrm{d}x_{\ell}. \tag{2.3}$$

Relation (2.3) is the basic relation from which we derive all information about the asynchronous iterative method.

3. Convergence of the solution in the mean

We begin by studying the convergence of the error. The error is a vector quantity, so the expected value of the error is also a vector. We first consider the expected value of the single coordinate x_1

$$E^{\nu+1}(x_1) = \int x_1 \beta^{\nu+1}(\vec{x})\ \mathrm{d}x.$$

Using the recursion (2.3) and the integration result of Lemma 2.1, we have

$$\begin{aligned}
E^{\nu+1}(x_1) &= \sum_{\ell,q} \frac{p_{\ell} s_q}{\|\vec{b}_{\ell}\|} \int x_1 \int_{(B\vec{y})_{\ell}=x_{\ell}} \beta^{\nu-q}(\vec{y})\ \mathrm{d}y \int_{-\infty}^{\infty} \beta^{\nu}(\vec{x})\ \mathrm{d}x_{\ell}\ \mathrm{d}x \\
&= \sum_{q} \frac{p_1 s_q}{\|\vec{b}_1\|} \int x_1 \int_{(B\vec{y})_1=x_1} \beta^{\nu-q}(\vec{y})\ \mathrm{d}y \int_{-\infty}^{\infty} \beta^{\nu}(\vec{x})\ \mathrm{d}x_1\ \mathrm{d}x \\
&\quad + \sum_{\ell\neq 1,q} \frac{p_{\ell} s_q}{\|\vec{b}_{\ell}\|} \int x_1 \int_{(B\vec{y})_{\ell}=x_{\ell}} \beta^{\nu-q}(\vec{y})\ \mathrm{d}y \int_{-\infty}^{\infty} \beta^{\nu}(\vec{x})\ \mathrm{d}x_{\ell}\ \mathrm{d}x \\
&= \sum_{q} p_1 s_q \int (B\vec{y})_1 \beta^{\nu-q}(\vec{y})\ \mathrm{d}y \int \beta^{\nu}(\vec{x})\ \mathrm{d}x \\
&\quad + \sum_{\ell\neq 1,q} p_{\ell} s_q \int \beta^{\nu-q}(\vec{y})\ \mathrm{d}y \int x_1 \beta^{\nu}(\vec{x})\ \mathrm{d}x \\
&= \sum_{q} p_1 s_q \int (B\vec{x})_1 \beta^{\nu-q}(\vec{x})\ \mathrm{d}x + (1-p_1) E^{\nu}(x_1) \\
&= \sum_{q} p_1 s_q (B E^{\nu-q}(\vec{x}))_1 + (1-p_1) E^{\nu}(x_1).
\end{aligned} \tag{3.1}$$

Combining this with similar results for $\ell = 2, \ldots, N$ we obtain

$$E^{\nu+1}(\vec{x}) = PB \sum_{q} s_q E^{\nu-q}(\vec{x}) + (I-P) E^{\nu}(\vec{x}), \tag{3.2}$$

where P is the diagonal matrix with entries p_ℓ. This is the recursion relation for the expectation of the error. Notice that all essential information about the process is in this formula, the matrix B, the component probabilities P, and the delay probabilities s_q.

We wish to find conditions under which the expectations converge as ν increases. To do this, we transform Eq. (3.2) using the Laplace transform. We form the sum

$$\begin{aligned}
&\sum_{\nu=0}^{\infty} z^{\nu+1} E^{\nu+1}(\vec{x}) \\
&\quad = z \sum_{\nu=0}^{\infty} z^{\nu} \sum_{q=0}^{\infty} s_q P B E^{\nu-q}(\vec{x}) + z(I-P) \sum_{\nu=0}^{\infty} z^{\nu} E^{\nu}(\vec{x}) \\
&\quad = z \sum_{q=0}^{\infty} s_q z^q \sum_{\nu=0}^{\infty} z^{\nu-q} P B E^{\nu-q}(\vec{x}) + z(I-P) \sum_{\nu=0}^{\infty} z^{\nu} E^{\nu}(\vec{x}) \\
&\quad = z \sum_{q=0}^{\infty} s_q z^q \sum_{\nu=0}^{\infty} z^{\nu-q} P B E^{\nu-q}(\vec{x}) + z(I-P) \sum_{\nu=0}^{\infty} z^{\nu} E^{\nu}(\vec{x}). \qquad (3.3)
\end{aligned}$$

Recall that we assume that $\vec{x}^{\nu} = \vec{x}^0$ for all negative values of ν.

In the subsequent analysis we need the generating function $s(z)$ of the delay probabilities, defined by

$$s(z) = \sum_{q=0}^{\infty} s_q z^q.$$

Note that, since the coefficients s_q are non-negative and $s(1) = 1$ by (2.2), $s(z)$ is an analytic function in the unit disk and is continuous on the closed unit disk.

If we define

$$\vec{E}(z) = \sum_{\nu=0}^{\infty} z^{\nu} E^{\nu}(\vec{x}),$$

then relation (3.3) becomes

$$\begin{aligned}
&\vec{E}(z) - \vec{E}(0) \\
&\quad = z \sum_{q} s_q z^q \sum_{\nu=0}^{\infty} z^{\nu-q} P B \vec{E}^{\nu-q}(z) + z(I-P)\vec{E}(z) \\
&\quad = z \sum_{q} s_q \sum_{\nu=0}^{q-1} z^{\nu} P B E^0(\vec{x}) + z \sum_{q} s_q z^q \sum_{\mu=0}^{\infty} z^{\mu} P B \vec{E}^{\mu}(z) + z(I-P)\vec{E}(z)
\end{aligned}$$

$$= z \sum_q s_q \frac{1-z^q}{1-z} PB\vec{E}(0) + zs(z)PB\vec{E}(z) + z(I-P)\vec{E}(z)$$
$$= z\frac{1-s(z)}{1-z} PB\vec{E}(0) + zs(z)PB\vec{E}(z) + z(I-P)\vec{E}(z). \tag{3.4}$$

Solving for $\vec{E}(z)$ we obtain

$$(I - z[I - P + s(z)PB])\vec{E}(z) = z\frac{1-s(z)}{1-z} PB\vec{E}(0) + \vec{E}(0) \tag{3.5}$$

or

$$\vec{E}(z) = (I - z[I - P + s(z)PB])^{-1} \left(z\frac{1-s(z)}{1-z} PB + I \right) \vec{E}(0). \tag{3.6}$$

Define the matrices

$$M(z) = I - z[I - P + s(z)PB]$$

and

$$N(z) = z\frac{1-s(z)}{1-z} PB + I.$$

Because $s(z)$ is an analytic function for z with $|z| < 1$, $M(z)$ is also an analytic function for z in the unit circle.

From (3.6), we have that the values of the expectation are given by

$$E^\nu(\vec{x}) = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \frac{1}{z^{\nu+1}} M(z)^{-1} N(z)\, \mathrm{d}z\ \vec{E}^0(\vec{x}), \tag{3.7}$$

where Γ is any simple curve that circles the origin in the positive direction. The convergence rate of $E^\nu(\vec{x})$ depends on the points where $M(z)$ is singular.

Define the matrices

$$T_\nu = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \frac{1}{z^{\nu+1}} M(z)^{-1} N(z)\, \mathrm{d}z \tag{3.8}$$

and so, from (3.7) we have

$$E^\nu(\vec{x}) = T_\nu\ \vec{E}^0(\vec{x}).$$

Thus the convergence of the expectation is governed by the sequence of matrices T_ν.

In particular, we have the following results.

Theorem 3.1. *A necessary condition that the expected value of the solution converge for all initial values $\vec{x}_0$ is that $M(z)$ be non-singular in $|z| < 1$.*

Proof. By standard results from the calculus of residues, if $M(z)$ is singular for a single value ξ with $|\xi| < 1$, then the matrices T_ν satisfy

$$T_\nu = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \frac{1}{z^{\nu+1}} M(z)^{-1} N(z)\, \mathrm{d}z = \mathrm{O}(|\xi|^{-\nu}).$$

If there are other points at which $M(z)$ is singular, then they add similar terms to the sequence of the T_ν. Thus if $M(z)$ is singular inside the unit circle, then the sequence of matrices T_ν is unbounded. □

Theorem 3.2. *If the matrix $M(z)$ is analytic and non-singular in $|z| \leqslant R$ with $1 < R$, then the expected value of the solution converges and*

$$\|E^\nu(\vec{x})\| = \mathrm{O}(R^{-\nu}).$$

Proof. From (3.8), we have

$$\begin{aligned}\|T_\nu\| &\leqslant \frac{1}{2\pi}\int_{|z|=R} \frac{1}{R^{\nu+1}}\|M(z)^{-1}N(z)\|\ R\ \mathrm{d}\theta \\ &= R^{-\nu}\frac{1}{2\pi}\int_{|z|=R}\|M(z)^{-1}N(z)\|\ \mathrm{d}\theta\end{aligned}$$

and the proof is immediate. □

We do not state other theorems for the convergence of the expectation of the error, rather we state without proof a few sufficient conditions for convergence.

Assume that $M(z)$ is non-singular inside the unit circle. Then for $z = \mathrm{e}^{\mathrm{i}\theta}$ we have

$$T_\nu = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{-\nu\mathrm{i}\theta}M(\mathrm{e}^{\mathrm{i}\theta})^{-1}N(\mathrm{e}^{\mathrm{i}\theta})\ \mathrm{d}\theta. \tag{3.9}$$

If the integrand is in L^1 in an appropriate matrix norm then by the Riemann–Lebesgue theorem the values of T_ν converge to 0. This implies that if $M(z)$ has isolated singularities on the unit circle and there are constants α and C with $0 \leqslant \alpha < 1$ such that

$$(1-|z|)^\alpha\|M(z)^{-1}N(z)\| < C \tag{3.10}$$

for all z inside and on the unit circle, then the expectation of the solution will converge. We also see that if $M(z)$ has a pole on the unit circle, then the expectation of the solution will not converge.

We also have, from results in the L^2 theory of Fourier series, Parseval's relation for the series of T_ν

$$\begin{aligned}\sum_{\nu=0}^{\infty}(T_\nu)^*T_\nu &= \frac{1}{2\pi}\int_0^{2\pi}\left[M(\mathrm{e}^{\mathrm{i}\theta})^{-1}N(\mathrm{e}^{\mathrm{i}\theta})\right]^*M(\mathrm{e}^{\mathrm{i}\theta})^{-1}N(\mathrm{e}^{\mathrm{i}\theta})\ \mathrm{d}\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi}N(\mathrm{e}^{\mathrm{i}\theta})^*M(\mathrm{e}^{\mathrm{i}\theta})^{-*}M(\mathrm{e}^{\mathrm{i}\theta})^{-1}N(\mathrm{e}^{\mathrm{i}\theta})\ \mathrm{d}\theta.\end{aligned}$$

The integral will be finite if relation (3.10) holds for α with $0 \leqslant \alpha < 1/2$.

In the analysis of the variance, the following condition will be seen to be important.

Definition 3.1. The matrix B is said to satisfy the weighting assumption if there exists a diagonal matrix D with positive diagonal entries such that

$$B^{\mathrm{T}} D B < D. \tag{3.11}$$

One application that will be useful is the following.

Theorem 3.3. *If the matrix B satisfies the weighting assumption, then the matrix $M(z)$ is non-singular for $|z| \leqslant 1$ and any diagonal probability matrix P.*

Proof. Let α be an eigenvalue of the matrix $I - P + \zeta PB$, where ζ is a complex number satisfying $|\zeta| \leqslant 1$ and let v be the associated eigenvector. Let D be a diagonal matrix such that (3.11) is satisfied and define $W = P^{-1}D$. Then we have

$$\zeta PBv = \alpha v - (I - P)v$$

and by using the weighting assumption we obtain

$$\begin{aligned} |\zeta|^2(\vec{v}, WP\vec{v}) > |\zeta|^2(B\vec{v}, WPB\vec{v}) &= |\zeta|^2(PB\vec{v}, P^{-1}WPB\vec{v}) \\ &= (\alpha\vec{v} - (I - P)\vec{v}, P^{-1}W(\alpha\vec{v} - (I - P)\vec{v})) \\ &= \sum_k |\alpha - (1 - p_k)|^2 p_k^{-1} w_k |v_k|^2 \\ &\geqslant \min_k \left| \frac{\alpha - (1 - p_k)}{p_k} \right|^2 (\vec{v}, WP\vec{v}). \end{aligned}$$

Thus,

$$1 \geqslant |\zeta| > \min_k \left| \frac{\alpha - (1 - p_k)}{p_k} \right|$$

and in particular, for some index k we have

$$p_k > |\alpha - (1 - p_k)|$$

but this means that α is inside a circle of radius p_k centered at $1 - p_k$. Thus α has magnitude less than 1.

The matrix $M(z)$ is

$$M(z) = I - z[I - P + s(z)PB]$$

and so if μ is an eigenvalue of $M(z)$, then

$$\mu = 1 - z\alpha,$$

where α is an eigenvalue as above with $\zeta = s(z)$ (recall $|s(z)| \leqslant 1$). But since α has magnitude less than 1, μ cannot be zero for $|z| \leqslant 1$. Thus $M(z)$ is non-singular for all z inside and on the unit circle. □

For later results we need the following observation.

Theorem 3.4. *If the expectation of the error converges, then the quantity*

$$S^{\nu}(\vec{x}) = \sum_{q=0}^{\infty} s_q E^{\nu-q}(\vec{x})$$

also converges. This sum represents the average expected value to be used in updating the solution.

Proof. The proof follows from the observation that

$$\begin{aligned} S^{\nu}(\vec{x}) &= \frac{1}{2\pi \mathrm{i}} \int_{\Gamma} \frac{s(z)}{z^{\nu+1}} M(z)^{-1} N(z)\, \mathrm{d}z\ \vec{E}^{0}(\vec{x}) \\ &= \sum_{q=0}^{\nu} s_{\nu-q} T_q \vec{E}^{0}(\vec{x}) + \sum_{q=\nu+1}^{\infty} s_q \vec{E}^{0}(\vec{x}). \end{aligned} \tag{3.12}$$

Since the sequence T_ν is convergent to 0, for any ε there is an N such that for $q \geqslant N$, $\|T_q\| \leqslant \varepsilon$. We may also assume that N is sufficiently large that

$$\sum_{r=N}^{\infty} s_r \leqslant \varepsilon.$$

Hence, for $\nu \geqslant N$ the last summation in (3.12) is bounded by ε times the magnitude of the initial expectation and

$$\begin{aligned} \left\| \sum_{q=0}^{\nu} s_{\nu-q} T_q \right\| &\leqslant \sum_{q=0}^{\nu} s_{\nu-q} \|T_q\| = \sum_{q=0}^{\nu-N} s_{\nu-q} \|T_q\| + \sum_{q=\nu-N+1}^{\nu} s_{\nu-q} \|T_q\| \\ &\leqslant \sum_{r=N}^{\infty} s_r \max_q(\|T_q\|) + \sum_{r=0}^{\infty} s_r\ \varepsilon \leqslant \varepsilon \left(\max_q(\|T_q\|) + 1 \right). \end{aligned}$$

Thus, $S^{\nu}(\vec{x})$ is convergent whenever $E^{\nu}(\vec{x})$ is convergent. □

For the analysis of the distributed memory model we will need the following result.

Theorem 3.5.

$$\sum_{\nu=0}^{\infty} \|S^{\nu}(\vec{x})\|^2 \tag{3.13}$$

converges if and only if

$$\sum_{\nu=0}^{\infty} \|E^{\nu}(\vec{x})\|^2 \tag{3.14}$$

converges.

Proof. The proof follows from the observation that

$$\sum_{\nu=0}^{\infty} \|S^{\nu}(\vec{x})\|^2 = \big(E^0(\vec{x}), \mathscr{S}E^0(\vec{x})\big),$$

where

$$\mathscr{S} = \frac{1}{2\pi}\int_0^{2\pi} |s(\mathrm{e}^{\mathrm{i}\theta})|^2 \big[M(\mathrm{e}^{\mathrm{i}\theta})^{-1}N(\mathrm{e}^{\mathrm{i}\theta})\big]^* M(\mathrm{e}^{\mathrm{i}\theta})^{-1}N(\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta. \tag{3.15}$$

Similarly,

$$\sum_{\nu=0}^{\infty} \|E^{\nu}(\vec{x})\|^2 = \left(E^0(\vec{x}), \sum_{\nu=0}^{\infty} T_{\nu}^* T_{\nu} E^0(\vec{x})\right)$$

and

$$\sum_{\nu=0}^{\infty} T_{\nu}^* T_{\nu} = \frac{1}{2\pi}\int_0^{2\pi} \big[M(\mathrm{e}^{\mathrm{i}\theta})^{-1}N(\mathrm{e}^{\mathrm{i}\theta})\big]^* M(\mathrm{e}^{\mathrm{i}\theta})^{-1}N(\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta. \tag{3.16}$$

The difference between the integrals in (3.15) and (3.16) is only the factor of $|s(\mathrm{e}^{\mathrm{i}\theta})|^2$ which is bounded by 1 according to (2.2). This factor might be zero for some value of θ, however, in that case

$$M(\mathrm{e}^{\mathrm{i}\theta}) = I - \mathrm{e}^{\mathrm{i}\theta}[I - P]$$

is non-singular. Thus, if $|s(\mathrm{e}^{\mathrm{i}\theta})|^2$ is zero for some value of θ, then the integrand of (3.16) is not unbounded there. Thus the integral in (3.16) is finite if and only if the integral in (3.15) is finite. □

4. Convergence of the variance

In addition to having the expectation converge, it is important to have some estimate of the variance of the error V^{ν} at each step. The variance is given by

$$V^{\nu} = E^{\nu}(|\vec{x} - E^{\nu}(\vec{x})|^2) = E^{\nu}(|\vec{x}|^2) - |E^{\nu}(\vec{x})|^2. \tag{4.1}$$

We obtain estimates on $E^{\nu}(|\vec{x}|^2)$ and, as we show, the natural estimates are on the boundedness of the sum

$$\sum_{\nu=0}^{\infty} E^{\nu}(|\vec{x}|^2). \tag{4.2}$$

Also, from (4.1) we have that

$$|E^\nu(\vec{x})|^2 \leqslant E^\nu(|\vec{x}|^2).$$

So if the expectation of the sum of squares of the error converges to zero, so does the expectation of the error.

The convergence of the sum (4.2) implies that $\vec{x}$ converges to zero almost surely.

Theorem 4.1. *If the sum* (4.2) *converges, then* $\vec{x}^\nu$ *converges to zero almost surely.*

Proof. For any positive value of ε we have

$$\text{Prob}(|\vec{x}^\nu| > \varepsilon) = \int_{|\vec{x}|>\varepsilon} \beta^\nu(\vec{x})\, \mathrm{d}x < \int_{|\vec{x}|>\varepsilon} \left|\frac{\vec{x}}{\varepsilon}\right|^2 \beta^\nu(\vec{x})\, \mathrm{d}x \leqslant \frac{1}{\varepsilon^2} E^\nu(|\vec{x}|^2).$$

Thus, if ν is chosen large enough that $E^\nu(|\vec{x}|^2) < \varepsilon^3$, then

$$\text{Prob}(|\vec{x}^\nu| > \varepsilon) < \varepsilon,$$

and so,

$$\text{Prob}(|\vec{x}^\nu| \leqslant \varepsilon) \geqslant 1 - \varepsilon$$

for ν sufficiently large. □

We now state our first theorem on the convergence of the variance.

Theorem 4.2. *Sufficient conditions for the convergence of the variance for the shared memory model are*:

(1) *the matrix B satisfies the weighting assumption, Definition* 3.1;
(2) *the expected value of the delay,* $s'(1)$, *is finite*;
(3) *the initial variance is finite.*

Under these conditions the sum (4.2) *is finite.*

Proof. We begin by considering the expected value of $|\vec{x}|^2$, which is

$$E^{\nu+1}(|\vec{x}|^2) = \int |\vec{x}|^2 \beta^{\nu+1}(\vec{x})\, \mathrm{d}x = \sum_{j=1}^{N} \int x_j^2 \beta^{\nu+1}(\vec{x})\, \mathrm{d}x.$$

We first consider the expected value of the one component x_1^2, i.e.,

$$E^{\nu+1}(x_1^2) = \int x_1^2 \beta^{\nu+1}(\vec{x})\, \mathrm{d}x.$$

Similar to the derivation (3.1), we obtain

$$\begin{aligned} E^{\nu+1}(x_1^2) &= \sum_{\ell,q} \frac{p_\ell s_q}{\|b_\ell\|} \int x_1^2 \int_{(B\vec{y})_\ell = x_\ell} \beta^{\nu-q}(\vec{y})\, \mathrm{d}y \int_{-\infty}^{\infty} \beta^\nu(\vec{x})\, \mathrm{d}x_\ell\, \mathrm{d}x \\ &= \sum_{q} \frac{p_1 s_q}{\|b_1\|} \int x_1^2 \int_{(B\vec{y})_1 = x_1} \beta^{\nu-q}(\vec{y})\, \mathrm{d}y \int_{-\infty}^{\infty} \beta^\nu(\vec{x})\, \mathrm{d}x_1\, \mathrm{d}x \end{aligned}$$

$$+\sum_{\ell\neq 1,q}\frac{p_\ell s_q}{\|b_\ell\|}\int\int_{(B\vec{y})_\ell=x_\ell}\beta^{\nu-q}(\vec{y})\,\mathrm{d}y\int_{-\infty}^{\infty}x_1^2\beta^\nu(\vec{x})\,\mathrm{d}x_\ell\,\mathrm{d}x$$
$$=\sum_q p_1 s_q\int(B\vec{x})_1^2\beta^{\nu-q}(\vec{x})\,\mathrm{d}x+(1-p_1)E^\nu(x_1^2).$$

We now combine this with similar results for $\ell=2,\ldots,N$, but first note that we can multiply these relations by positive weights w_ℓ obtaining

$$E^{\nu+1}(\vec{x},W\vec{x})=\sum_q s_q E^{\nu-q}(B\vec{x},WPB\vec{x})+E^\nu(\vec{x},W\vec{x})-E^\nu(\vec{x},WP\vec{x}),\tag{4.3}$$

where W is the diagonal matrix whose diagonal entries are the w_ℓ.

As with the expectation of the error, we use the Laplace transform. However, whereas linearity can be used to give formulas for the expectation of the error, for the expectation of the square of the error other techniques are used. Thus the Laplace variable t is restricted to being a positive variable.

From (4.3) we obtain

$$\begin{aligned}\sum_{\nu=0}^{\infty}t^{\nu+1}E^{\nu+1}(\vec{x},W\vec{x})&=t\sum_{\nu=0}^{\infty}t^\nu\sum_q s_q E^{\nu-q}(B\vec{x},WPB\vec{x})\\&\quad+t\sum_{\nu=0}^{\infty}t^\nu E^\nu(\vec{x},W\vec{x})-t\sum_{\nu=0}^{\infty}t^\nu E^\nu((\vec{x},WP\vec{x}))\\&=t\sum_q s_q t^q\sum_{\nu=0}^{\infty}t^{\nu-q}E^{\nu-q}(B\vec{x},WPB\vec{x})\\&\quad+t\sum_{\nu=0}^{\infty}t^\nu E^\nu(\vec{x},W\vec{x})-t\sum_{\nu=0}^{\infty}t^\nu E^\nu((\vec{x},WP\vec{x}))\\&=t\sum_q s_q t^q\sum_{\nu=0}^{\infty}t^{\nu-q}E^{\nu-q}(B\vec{x},WPB\vec{x})\\&\quad+t\sum_{\nu=0}^{\infty}t^\nu E^\nu(\vec{x},W\vec{x})-t\sum_{\nu=0}^{\infty}t^\nu E^\nu((\vec{x},WP\vec{x})).\end{aligned}\tag{4.4}$$

We define

$$W(t)=\sum_{\nu=0}^{\infty}t^\nu E^\nu(\vec{x},W\vec{x}),$$

$$BWPB(t) = \sum_{\nu=0}^{\infty} t^{\nu} E^{\nu}(B\vec{x}, WPB\vec{x}),$$

and similarly for other summations in (4.4). We obtain from (4.4)

$$\begin{aligned}
&W(t) - W(0) \\
&\quad = t \sum_{q} s_q t^q \sum_{\nu=0}^{\infty} t^{\nu-q} E^{\nu-q}(B\vec{x}, WPB\vec{x}) + tW(t) - tWP(t) \\
&\quad = t \sum_{q} s_q \sum_{\nu=0}^{q-1} t^{\nu} E^{0}(B\vec{x}, WPB\vec{x}) \\
&\qquad + t \sum_{q} s_q t^q \sum_{\mu=0}^{\infty} t^{\mu} E^{\mu}(B\vec{x}, WPB\vec{x}) + tW(t) - tWP(t) \\
&\quad = t \sum_{q} s_q \frac{1-t^q}{1-t} BWPB(0) + ts(t)BWPB(t) + tW(t) - tWP(t) \\
&\quad = t \frac{1-s(t)}{1-t} BWPB(0) + ts(t)BWPB(t) + tW(t) - tWP(t). \qquad (4.5)
\end{aligned}$$

Recall that $\vec{x}^{\nu} = \vec{x}^{0}$ for all negative values of ν.

After rearranging, we obtain the important relation

$$\begin{aligned}
tWP(t) + (1-t)W(t) = t \frac{1-s(t)}{1-t} BWPB(0) \\
+ W(0) + ts(t)BWPB(t). \qquad (4.6)
\end{aligned}$$

Note that the limit of $(1-s(t))/(1-t)$ as t approaches 1 is $s'(1)$. This is the expected delay as is seen by

$$E(q) = \sum_{q=0}^{\infty} q s_q = s'(1).$$

If the initial variance is finite and if the expected delay is finite, then the first two terms on the right-hand side are finite for $t = 1$. We now assume that these two assumptions hold.

If B satisfies the weighting assumption for some diagonal matrix D, then we define the (diagonal) matrix W by $D = WP$. By the weighting assumption, there is a positive value η such that

$$B^{\mathrm{T}} WPB \leqslant (1-\eta)WP$$

and this implies that

$$BWPB(t) \leqslant (1-\eta)WP(t).$$

Using this in relation (4.6), we obtain

$$\begin{aligned} tWP(t) + (1-t)W(t) &= t\frac{1-s(t)}{1-t}BWPB(0) + W(0) + ts(t)BWPB(t) \\ &\leqslant t\frac{1-s(t)}{1-t}BWPB(0) + W(0) + (1-\eta)ts(t)WP(t). \end{aligned}$$

Rearranging this gives

$$\begin{aligned} &\eta tWP(t) + (1-\eta)t(1-s(t))WP(t) + (1-t)W(t) \\ &\quad\leqslant t\frac{1-s(t)}{1-t}BWPB(0) + W(0), \end{aligned} \tag{4.7}$$

where each of the terms on both sides is a non-negative quantity.

Taking the limit as t approaches 1 gives the result

$$\eta WP(1) + \lim_{t\to 1^-}(1-t)W(t) \leqslant s'(1)BWPB(0) + W(0), \tag{4.8}$$

which implies that $WP(1)$ is bounded. Since $WP(1)$ is a series of non-negative terms, we have that the terms $E^\nu(\vec{x}, WP\vec{x})$ tend to zero as ν increases. We also have that $W(1)$ is bounded. Moreover,

$$E^\nu(|\vec{x}|^2) \leqslant (wp)^{-1}_{\min}E^\nu(\vec{x}, WP\vec{x}) \to 0 \tag{4.9}$$

as ν increases. Similarly, we conclude from $\sum E^\nu(\vec{x}, WP\vec{x})$ being finite that

$$\sum_{\nu=0}^{\infty} E^\nu(|\vec{x}|^2)$$

is finite. This concludes the proof of the theorem. □

It is interesting that the probabilities p_ℓ do not play an important role in the convergence of the variance.

Theorem 4.3. *A necessary condition for the sum*

$$W(1) = \sum_{\nu=0}^{\infty} E^\nu(\vec{x}, W\vec{x})$$

and the sum of variances

$$\sum_{\nu=0}^{\infty} V^\nu(\vec{x})$$

to be finite for all initial data $\vec{x}^0$ is that $s'(1)$ be finite.

Proof. An examination of relation (4.7) shows that if $W(1)$ is to be finite, then it is necessary that $s'(1)$ be finite. □

Also, Theorem 3.3 shows that if the conditions of Theorem 4.2 are satisfied then the matrix $M(z)^{-1}N(z)$ is bounded for all z inside and on the unit circle.

Notice that it is possible for the expectation of the error to converge, but that the sum of variances not converge, in the sense of Theorem 4.3. An example is given by the delay distributions generated with

$$s(z) = 1 - \sqrt{1 - z}$$

with B suitably chosen, e.g., $\|B\|$ being sufficiently small. Then by (3.10) the expected value of the error may converge, but by Theorem 4.3 the sum of the variances is not finite.

5. Asynchronous iteration with distributed memory

In this section we look at a model for distributed memory for asynchronous iteration. As illustrated in Fig. 2 the different processors communicate with each other for the data and there is no common source for the data. Each processor stores the component or set of components which it updates. In this model the update of a component uses components that may have different delays. The iteration counter is increased each time some component is updated. The update of a component depends on components $x_k^{q_k}$ where the delays are chosen independently. If the delays are the N-tuple $Q = (q_k)$ then the probability of the delay is

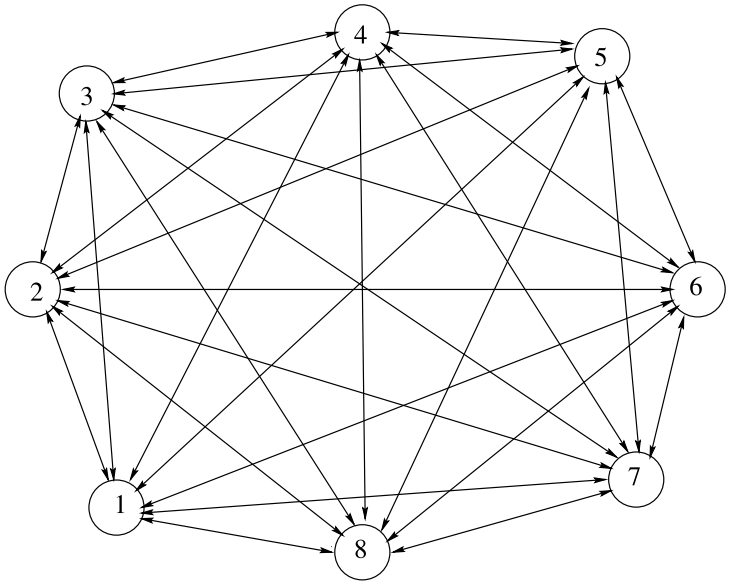

Fig. 2. Illustration of distributed memory model.

$$s_Q = \prod_{k=1}^{N} s_{q_k}.$$

We define the expectation $E^{\nu-Q}(\vec{x})$ by

$$E^{\nu-Q}(\vec{x}) = \begin{pmatrix} E^{\nu-q_1}(\vec{x}) \\ E^{\nu-q_2}(\vec{x}) \\ \vdots \\ E^{\nu-q_N}(\vec{x}) \end{pmatrix}.$$

For the distributed memory model, the iterative process can be written as

$$x_\ell^{\nu+1} = \sum_{m=1}^{N} b_{\ell,m} x_m^{\nu-q_m} + d_\ell \quad \text{with probability } p_\ell s_Q,$$
$$x_\ell^{\nu+1} = x_\ell^{\nu} \quad \text{with probability } 1 - p_\ell.$$

Using ideas similar to those used in the derivation in Section 2, the probability distribution for this case is seen to satisfy the recursion

$$\beta^{\nu+1}(\vec{x}) = \sum_{\ell,Q} \frac{p_\ell s_Q}{\|b_\ell\|} \int_{(B\vec{y})_\ell = x_\ell} \beta^{\nu-Q}(\vec{y})\, \mathrm{d}y \int_{-\infty}^{\infty} \beta^{\nu}(\vec{x})\, \mathrm{d}x_\ell,$$

where

$$\beta^{\nu-Q}(\vec{y}) = \prod_{i=1}^{N} \int_{z_i = y_i} \beta^{\nu-q_i}(\vec{z})\, \mathrm{d}\hat{z}_i \quad \text{for } Q = (q_1, q_2, \ldots, q_N).$$

The integration with respect to $\mathrm{d}\hat{z}_i$ means that all variables other than z_i are integrated over $\mathbb{R}$.

We begin with determining the convergence of the expectation of the error. The recursion formula is

$$\begin{aligned} E^{\nu+1}(x_1) &= p_1 \sum_{Q} \frac{s_Q}{\|b_1\|} \int x_1 \int_{(B\vec{y})_1 = x_1} \beta^{\nu-Q}(\vec{y})\, \mathrm{d}y \int_{-\infty}^{\infty} \beta^{\nu}(\vec{x})\, \mathrm{d}\hat{x}_1\, \mathrm{d}x \\ &\quad + \sum_{Q,\ell\neq 1} \frac{p_\ell s_Q}{\|b_\ell\|} \int x_1 \int_{(B\vec{y})_\ell = x_\ell} \beta^{\nu-Q}(\vec{y})\, \mathrm{d}y \int_{-\infty}^{\infty} \beta^{\nu}(\vec{x})\, \mathrm{d}\hat{x}_\ell\, \mathrm{d}x \\ &= p_1 \sum_{Q} s_Q \int (B\vec{x})_1 \beta^{\nu-Q}(\vec{x})\, \mathrm{d}x + (1-p_1) E^{\nu}(x_1). \end{aligned}$$

We examine in more detail the integral in the last summation

$$\int (B\vec{x})_1 \beta^{\nu-Q}(\vec{x})\, \mathrm{d}x = \int \sum_{k=1}^{N} B_{1,k} \int x_k \prod_{i=1}^{N} \int_{x_i = x_i} \beta^{\nu-q_i}(\vec{z})\, \mathrm{d}\hat{z}_i\, \mathrm{d}x$$

$$= \sum_{k=1}^{N} B_{1,k} E^{\nu - q_k} x_k = (B E^{\nu - Q}(\vec{x}))_1.$$

Hence, we have for the first component of the error

$$E^{\nu+1}(x_1) = p_1 \sum_{Q} s_Q (B E^{\nu - Q}(\vec{x}))_1 + (1 - p_1) E^{\nu}(x_1)$$

and combining this with similar relations for the other components, we have

$$E^{\nu+1}(\vec{x}) = PB \sum_{Q} s_Q E^{\nu - Q}(\vec{x}) + (1 - P) E^{\nu}(\vec{x}).$$

Finally, we note that

$$\sum_{Q} s_Q E^{\nu - Q}(\vec{x}) = \begin{pmatrix} \sum_Q s_Q E^{\nu - q_1}(x_1 \\ \sum_Q s_Q E^{\nu - q_2}(x_2) \\ \vdots \\ \sum_Q s_Q E^{\nu - q_N}(x_N) \end{pmatrix} = \begin{pmatrix} \sum_Q \prod s_{q_j} E^{\nu - q_1}(x_1) \\ \sum_Q \prod s_{q_j} E^{\nu - q_2}(x_2) \\ \vdots \\ \sum_Q \prod s_{q_j} E^{\nu - q_N}(x_N) \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{q_1} s_{q_1} E^{\nu - q_1}(x_1) \\ \sum_{q_2} s_{q_2} E^{\nu - q_2}(x_2) \\ \vdots \\ \sum_{q_N} s_{q_N} E^{\nu - q_N}(x_N) \end{pmatrix} = \sum_{q=0}^{\infty} s_q E^{\nu - q}(\vec{x}).$$

Thus, we have the same recurrence relation for this case as the case of shared memory

$$E^{\nu+1}(\vec{x}) = PB \sum_{q} s_q E^{\nu - q}(\vec{x}) + (1 - P) E^{\nu}(\vec{x})$$

and so all the results for that case concerning the convergence of the expectation hold for this case as well.

5.1. Convergence of the variance

As in the shared memory model we consider the expected value of $|\vec{x}|^2$. We begin with the expected value of x_1^2

$$E^{\nu+1}(x_1^2) = \int x_1^2 \beta^{\nu+1}(\vec{x})\, \mathrm{d}x$$

$$= \sum_{\ell, Q} \frac{p_\ell s_Q}{\|b_\ell\|} \int x_1^2 \int_{(B\vec{y})_\ell = x_\ell} \beta^{\nu - Q}(\vec{y})\, \mathrm{d}y \int_{-\infty}^{\infty} \beta^{\nu}(\vec{x})\, \mathrm{d}x_\ell\, \mathrm{d}x$$

$$= \sum_{Q} \frac{p_1 s_Q}{\|b_1\|} \int x_1^2 \int_{(B\vec{y})_1 = x_1} \beta^{\nu - Q}(\vec{y})\, \mathrm{d}y \int_{-\infty}^{\infty} \beta^{\nu}(\vec{x})\, \mathrm{d}x_1\, \mathrm{d}x$$

$$
\begin{aligned}
&+ \sum_{\ell \neq 1,Q} \frac{p_\ell S_Q}{\|b_\ell\|} \int x_1^2 \int_{(B\vec{y})_\ell = x_\ell} \beta^{\nu-Q}(\vec{y})\, \mathrm{d}y \int_{-\infty}^{\infty} \beta^{\nu}(\vec{x})\, \mathrm{d}x_\ell\, \mathrm{d}x \\
&= \sum_Q p_1 s_Q \int (B\vec{x})_1^2 \beta^{\nu-Q}(\vec{x})\, \mathrm{d}x + (1-p_1) E^{\nu}(x_1^2) \\
&= \sum_Q p_1 s_Q E^{\nu-Q}((B\vec{x})_1^2) + (1-p_1) E^{\nu}(x_1^2). \qquad (5.1)
\end{aligned}
$$

We look in more detail at the summation in this last expression.

$$
\begin{aligned}
&\sum_Q s_Q E^{\nu-Q}((B\vec{x})_1^2) \\
&\quad = \sum_Q \sum_{k=1}^{N} s_Q \int (B_{1,k}x_k)^2 \beta^{\nu-Q}(\vec{x})\, \mathrm{d}x \\
&\qquad + \sum_Q \sum_{k\neq j}^{N} s_Q \int (B_{1,k}x_k)(B_{1,j}x_j) \beta^{\nu-Q}(\vec{x})\, \mathrm{d}x \\
&\quad = \sum_{k=1}^{N} \sum_q s_q (B_{1,k})^2 E^{\nu-q}(x_k^2) \\
&\qquad + \sum_{q_k, q_j} \sum_{k\neq j}^{N} s_{q_k} s_{q_j} B_{1,k} E^{\nu-q_k}(x_k) B_{1,j} E^{\nu-q_j}(x_j) \\
&\quad = \sum_{k=1}^{N} \sum_q s_q (B_{1,k})^2 E^{\nu-q}(x_k^2) + \sum_{k\neq j}^{N} S^{\nu}(x_k) B_{1,k} B_{1,j} S^{\nu}(x_j),
\end{aligned}
$$

where $S^{\nu}(\vec{x})$ is defined as in Theorem 3.4.

Let $\mathscr{D}$ be the matrix $\mathrm{diag}(B^{\mathrm{T}}WPB)$ and let $\mathscr{C} = B^{\mathrm{T}}WPB - \mathscr{D}$. Then from (5.1) and using weighting as done previously, we have

$$
\begin{aligned}
E^{\nu+1}(\vec{x}, W\vec{x}) = &\sum_{q=0}^{\infty} s_q E^{\nu-q}(\vec{x}, \mathscr{D}\vec{x}) + E^{\nu}(\vec{x}, W\vec{x}) \\
&- E^{\nu}(\vec{x}, WP\vec{x}) + (S^{\nu}(\vec{x}), \mathscr{C}S^{\nu}(\vec{x})).
\end{aligned}
$$

As shown in Theorems 3.4 and 3.5, the convergence of $S^{\nu}(\vec{x})$ follows from that of $E^{\nu}(\vec{x})$.

This leads to the relation

$$
tWP(t) + (1-t)W(t) = t\frac{1-s(t)}{1-t}\mathscr{D}(0) + ts(t)\mathscr{D}(t) + W(0) + \mathscr{C}(t).
$$

For now assume that the value of $\mathscr{C}(1)$ is finite. A sufficient condition for convergence of the variance is that $\mathscr{D} < WP$. Since these matrices are diagonal, this means that

$$\sum_{j}^{N} w_j p_j |B_{j,k}|^2 < w_k p_k \tag{5.2}$$

holds for each index k. But since the matrix

$$B^{(2)} = \left(|B_{j,k}|^2\right)$$

is a matrix of non-negative elements, there are weights w_k such that condition (5.2) is satisfied precisely when

$$\rho(B^{(2)}) < 1, \tag{5.3}$$

where $\rho(\cdot)$ is the spectral radius function for matrices.

We now examine the condition that $\mathscr{C}(1)$ be finite. This term is

$$\mathscr{C}(1) = \sum_{\nu=0}^{\infty} (S^\nu(\vec{x}), \mathscr{C}S^\nu(\vec{x}))$$

and this can be bounded by a constant multiple of

$$\sum_{\nu=0}^{\infty} \|S^\nu(\vec{x})\|^2.$$

By Theorem 3.5 this sum is finite if and only if the similar sum for $E^\nu(\vec{x})$ (i.e., (3.14)) is finite. Thus if the sum (3.14) is finite, then $\mathscr{C}(1)$ will be finite. The sum (3.14) is finite if the matrix $M(z)$ satisfies (3.10) for $\alpha < 1/2$.

We state this result as a theorem.

Theorem 5.1. *Sufficient conditions for the convergence of the variance for the distributed memory model are*:
(1) *the matrix B satisfies* $\rho(B^{(2)}) < 1$;
(2) *the expected value of the delay*, $s'(1)$, *is finite*;
(3) *the matrix* $M(z)$ *satisfies* (3.10) *for* $\alpha < \frac{1}{2}$;
(4) *the initial variance is finite*.
Under these conditions the quantity (4.2) *is finite*.

6. Over- and under-relaxation

In the theory of general linear methods (see e.g. [5]) over-relaxing or under-relaxing an iterative method may improve the rate of convergence. We now consider briefly the effect of relaxing the iterative methods presented here.

We consider the modifications resulting from relaxing the iteration with the formula

$$x_\ell^{\nu+1} = (1-\omega)x_\ell^\nu + \omega\left(\sum_{m=1}^{N} b_{\ell,m}x_m^{\nu-q} + d_\ell\right) \quad \text{with probability } p_\ell s_q,$$
$$x_\ell^{\nu+1} = x_\ell^\nu \quad \text{with probability } 1-p_\ell.$$

Note that this modification does not increase the storage or communication costs of the method. We consider only the case of shared memory for simplicity.

As was done previously, the recurrence relation for the expectation of the error can be shown to satisfy

$$\begin{aligned} E^{\nu+1}(\vec{x}) &= (I-\omega)PE^\nu(\vec{x}) + \omega PB\sum_q s_q E^{\nu-q}(\vec{x}) + (I-P)E^\nu(\vec{x}) \\ &= \omega PB\sum_q s_q E^{\nu-q}(\vec{x}) + (I-\omega P)E^\nu(\vec{x}). \end{aligned} \tag{6.1}$$

Thus, the earlier results carry over with the matrix ωP replacing the matrix P.

Theorem 6.1. *A necessary condition that the expected value of the solution converge is that the matrix*

$$M_\omega(z) = I - z[I - \omega P + s(z)\omega PB]$$

be non-singular in $|z| < 1$.

Theorem 6.2. *If the matrix B satisfies the weighting assumption, then the matrix* $M_\omega(z)$ *is non-singular for* $|z| \leqslant 1$ *for* ω *satisfying*

$$\omega \max_k p_k < 1.$$

The proof proceeds as with Theorem 6.2 with ωp_k replacing p_k. The conclusion is obtained from the requirement that the estimate

$$\omega p_k > |\alpha - (1-\omega p_k)|$$

give the conclusion that α be inside the unit circle for every value of k.

We now consider the convergence of the variance. The derivation of the recursion relation is similar to that done before, but involves several new terms. In the following derivation, let $\vec{t}_\ell$ stand for $(x_1, \ldots, t_\ell, \ldots, x_N)$.

$$\begin{aligned} E^{\nu+1}(x_1^2) = &\sum_{\ell,q} \frac{p_\ell s_q}{\|(1-\omega)\vec{e}_\ell + \omega\vec{b}_\ell\|} \\ &\times \int x_1^2 \int_{((1-\omega)\vec{t}+\omega B\vec{y})_\ell = x_\ell} \beta^{\nu-q}(\vec{y})\beta^\nu(\vec{t}_\ell)\,\mathrm{d}y\,\mathrm{d}t\,\mathrm{d}x \end{aligned}$$

$$
\begin{aligned}
&= \sum_q p_1 s_q \int\int ((1-\omega)\vec{u} + \omega B\vec{x})_1^2 \beta^{\nu-q}(\vec{x})\beta^{\nu}(\vec{u})\,\mathrm{d}x\,\mathrm{d}u \\
&\quad + (1-p_1)E^{\nu}(x_1^2) \\
&= p_1(1-\omega)^2 \int u_1^2 \beta^{\nu}(\vec{u})\,\mathrm{d}u \\
&\quad + 2p_1(1-\omega)\omega \sum_q s_q \int\int u_1 (B\vec{x})_1 \beta^{\nu-q}(\vec{x})\beta^{\nu}(\vec{u})\,\mathrm{d}x\,\mathrm{d}u \\
&\quad + p_1\omega^2 \int (B\vec{x})_1^2 \beta^{\nu-q}(\vec{x})\,\mathrm{d}x + (1-p_1)E^{\nu}(x_1^2) \\
&= p_1(1-\omega)^2 E^{\nu}(x_1^2) + 2(1-\omega)\omega E^{\nu}(x_1) p_1 \sum_q s_q E^{\nu-q}((B\vec{x})_1) \\
&\quad + \omega^2 p_1 \sum_q s_q E^{\nu-q}((B\vec{x})_1^2) + (1-p_1)E^{\nu}(x_1^2).
\end{aligned}
$$

Using the weighting as before, we have

$$
\begin{aligned}
E^{\nu+1}(\vec{x}, W\vec{x}) &= (1-\omega)^2 E^{\nu}(\vec{x}, WP\vec{x}) + 2\omega(1-\omega)E^{\nu}(\vec{x})WPBS^{\nu}(\vec{x}) \\
&\quad + \omega^2 \sum_q s_q E^{\nu-q}(B\vec{x}, WPB\vec{x}) + E^{\nu}(\vec{x}, W\vec{x}) - E^{\nu}(\vec{x}, WP\vec{x}) \\
&= -\omega(2-\omega)E^{\nu}(\vec{x}, WP\vec{x}) + \omega^2 \sum_q s_q E^{\nu-q}(B\vec{x}, WPB\vec{x}) \\
&\quad + E^{\nu}(\vec{x}, W\vec{x}) + 2\omega(1-\omega)E^{\nu}(\vec{x})WPBS^{\nu}(\vec{x}).
\end{aligned}
$$

This last term involves the product of $E^{\nu}(\vec{x})$ and $S^{\nu}(\vec{x})$, and it is finite under various assumptions on $M_{\omega}(z)$.

Then, using the same notational conventions as before, we have

$$
\begin{aligned}
&\omega(2-\omega)tWP(t) + (1-t)W(t) \\
&\quad = t\frac{1-s(t)}{1-t}BWPB(0) + \omega^2 ts(t)BWPB(t) + W(0) + \mathscr{S}(t). \qquad (6.2)
\end{aligned}
$$

As in the previous analyses, the convergence follows if the term involving $BWPB(t)$ on the right-hand side is bounded above by the term $WP(t)$ on the left-hand side. If

$$B^{\mathrm{T}}WPB \leqslant \rho^2 WP,$$

then we have the variance converges if $\omega(2-\omega) > \omega^2\rho^2$. This reduces to the relation

$$\omega < \frac{2}{1+\rho^2}.$$

This result shows that if $B^{\mathrm{T}}DB < \rho^2 D$ for some weighting matrix and for some ρ larger than 1, then under-relaxation can be used to get a convergent method. Also, if

ρ is less than 1, then over-relaxation is possible and may give faster rates of convergence.

It is also required that the term $\mathscr{S}(1)$ in (6.2) be finite and this will be so if the matrix $M_\omega(z)$ is non-singular in $|z| \leqslant 1$.

It is not clear how to estimate an optimal, or ever a better, value of ω. The convergence rate depends on the values where $M_\omega(z)$ is singular or its behavior for $|z| = 1$, however, this must be constrained by having the variance converge.

7. Numerical illustrations

In this section we illustrate the results of the previous theory for the shared memory model with computational results. All simulations were done using Matlab [6]. The matrix elements of the matrix B were chosen from uniform distributions in the range $[-\gamma, \gamma]$ for different values of γ. For a given randomly generated matrix, several cases were run, each for a specified number of iterations. We do not discuss practical stopping criteria in this paper. In each case shown here, the matrix was non-convergent in the sense of Chazan and Miranker, that is,

$$\rho(|B|) > 1,$$

so there do exist iteration sequences that are non-convergent, but the set of these sequences has probability zero. (The matrix $|B|$ is that formed by taking the absolute values of each of the elements of B.) Each figure displays the logarithm of the error $|x^\nu|$ as a function of ν.

The first case we present is a case for which the expectation of the error is convergent, but the variance of the error is non-convergent. The generating function for the delays, $s(z)$, is given by

$$s(z) = 1 - \sqrt{1 - z}.$$

Notice that $s'(1) = \infty$. The probabilities p_ℓ are uniform (i.e., $p_\ell = 1/N$). The matrix is a 5 by 5 matrix with spectral radius 0.49. Fig. 3 shows the result of a single run. What is displayed is the logarithm of $|x^\nu|$ as a function of ν for 20,000 steps. This figure shows that although the expectation of the error is converging to 0, the non-convergence of the variance makes it difficult to determine a stopping criteria for the iterations. It is not clear whether the error itself tends to zero in any sense, but this convergence is not useful in a practical computational sense.

Fig. 4 shows a typical calculation with a 5 by 5 matrix with spectral radius 0.90 and non-uniform probabilities p_ℓ. The generating function for the delays is

$$s(z) = z + \tfrac{2}{3}(1 - z)^{3/2}.$$

Seven cases are shown in Fig. 4 and for each case there is a definite convergence rate shown, with occasional intermittent jumps in the magnitude of the error. Each of the runs is obviously converging to the solution.

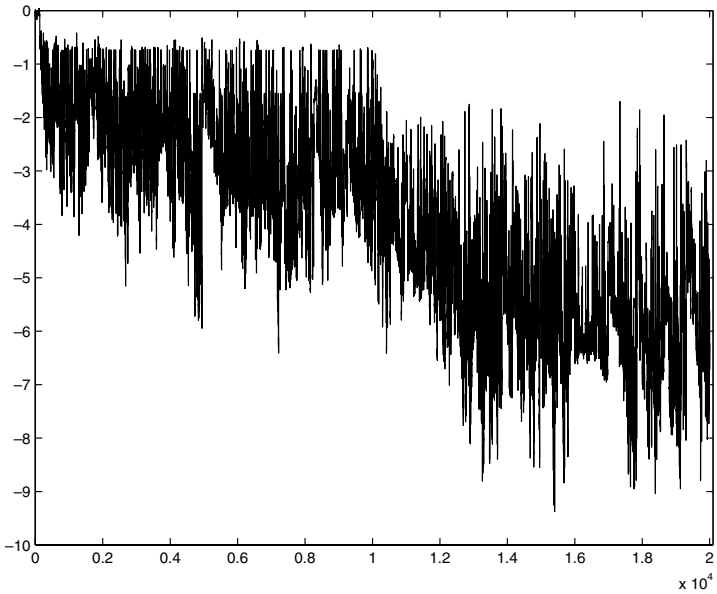

Fig. 3. A case with non-convergent variance.

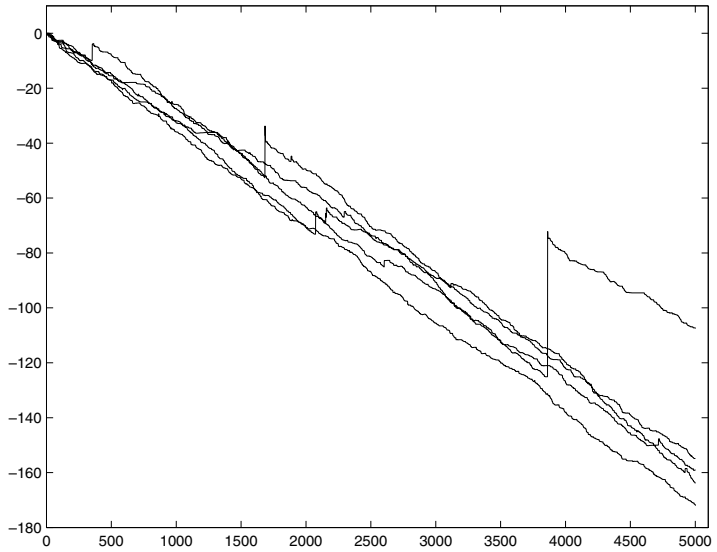

Fig. 4. A case with convergent variance.

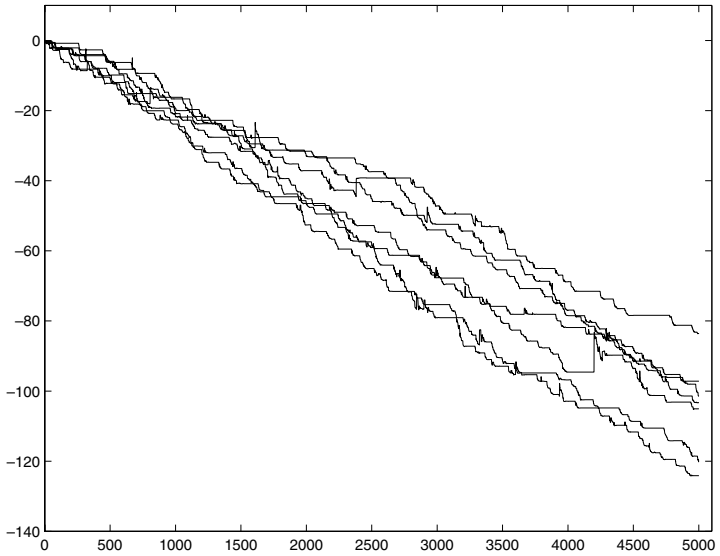

Fig. 5. A case with convergent variance.

Fig. 5 shows a typical calculation with a 10 by 10 matrix with spectral radius 0.61 and non-uniform probabilities p_ℓ. The generating function for the delays is

$$s(z) = \frac{1}{2 - z}.$$

Five cases are shown in Fig. 5 and each case exhibits a similar pattern of regions in which the norm of the error decreases very slowly or not at all followed by a rapid drop in the norm. As with the cases shown in Fig. 4, each case shows a convergent sequence. The difference in the behavior shown in Figs. 4 and 5 can be explained in part by the delay probabilities.

Numerous other tests were run and each of them exhibited behavior similar to that displayed in Figs. 4 and 5. Results for the distributed memory model were similar in appearance to those for the shared memory model.

8. Comparison of the convergence criteria

The conditions on the matrix B for the convergence of the variance given in Theorems 4.2 and 5.1 are both weaker than the condition of Chazan and Miranker, which requires that

$$\rho(|B|) < 1. \tag{8.1}$$

The matrix $|B|$ is that formed by taking the absolute values of each of the elements of B. Condition (8.1) is the necessary and sufficient condition that all of the iterative

sequences of Chazan and Miranker are convergent. This is true for both the shared memory and distributed memory models, see [7].

For the Jacobi algorithm, the standard synchronous method for solving the system (2.1), the convergence criterion is

$$\rho(B) < 1. \tag{8.2}$$

Here we discuss the relation of the weighting condition (3.11) and condition (5.3) to the two conditions (8.1) and (8.2).

A useful lemma in analyzing condition (8.1) is the following lemma of Baudet [1], see also [7].

Lemma 8.1. *The condition* $\rho(|B|) < 1$ *implies that there is a value* α *with* $0 < \alpha < 1$ *and a vector* $\vec{p}$ *with positive components such that* $|B|\vec{p} \leqslant \alpha\vec{p}$.

In fact, it is easily shown that the existence of α and $\vec{p}$ satisfying the conditions of Lemma 8.1 is equivalent to condition (8.1).

We now show that condition (8.1) implies the weighting condition and the condition on $B^{(2)}$ that are the conditions on the matrix B sufficient for the convergence of the variance.

Theorem 8.2. *Condition* (8.1) *implies the weighting condition*.

Proof. If (8.1) holds, then by Lemma 8.1 we have that there are positive vectors $\vec{w}$ and $\vec{v}$ such that

$$|B|\vec{v} < \vec{v} \quad \text{and} \quad |B|^{\mathrm{T}}\vec{w} < \vec{w}.$$

That is,

$$\sum_j |B_{i,j}|v_j < v_i \quad \text{and} \quad \sum_i w_i|B_{i,j}| < w_j.$$

We take the diagonal matrix D to have elements $d_i = w_i/v_i$. Let

$$Y_i = \sum_j B_{i,j}x_j.$$

Then

$$\begin{aligned}(\vec{x}, B^{\mathrm{T}}DB\vec{x}) = \sum_i d_i Y_i^2 = \sum_i d_i Y_i \sum_j B_{i,j}x_j = \sum_{i,j} \frac{w_i}{v_i} Y_i B_{i,j}x_j \\ \leqslant \left(\sum_{i,j} |B_{i,j}|\, x_j^2\, \frac{w_i}{v_j}\right)^{1/2} \left(\sum_{i,j} \frac{w_i}{v_i^2}\, Y_i^2\, |B_{i,j}|v_j\right)^{1/2}\end{aligned}$$

$$< \left(\sum_j x_j^2 \frac{w_j}{v_j} \right)^{1/2} \left(\sum_i Y_i^2 \frac{w_i}{v_i} \right)^{1/2}$$
$$= (\vec{x}, D\vec{x})^{1/2} (\vec{x}, B^{\mathrm{T}} D B \vec{x})^{1/2}.$$

This shows that

$$(\vec{x}, B^{\mathrm{T}} D B \vec{x}) < (\vec{x}, D\vec{x})$$

if (8.1) holds. □

If the weighting condition holds, then (5.3) on $B^{(2)}$ is also satisfied. This is most easily seen by taking x to be the vectors for which all components except one are 0.

Conditions (8.1) and (5.3) both use the spectral radius, and it is interesting to see how the weighting condition is related to conditions involving the spectral radius. With this in mind, it is useful to make the following definition.

Definition 8.1. For an N-dimensional matrix B, we define

$$\rho_{\pm}(B) = \max_{\vec{\theta}} \rho(S(\vec{\theta})B),$$

where $S(\vec{\theta}) = \mathrm{diag}(\mathrm{e}^{\mathrm{i}\theta_k})$ and the maximum is taken over all real values of $\vec{\theta}$ in $\mathbb{R}^N$.

Theorem 8.3. *The weighting condition implies that*

$$\rho_{\pm}(B) < 1. \tag{8.3}$$

Proof. The proof follows from the observations that the weighting assumption implies that $\rho(B) < 1$, and

$$B^{\mathrm{T}} D B = B^{\mathrm{T}} S(-\vec{\theta}) D S(\vec{\theta}) B = (S(\vec{\theta})B)^* D S(\vec{\theta}) B.$$

Thus, the weighting assumption implies that $\rho(S(\vec{\theta})B)$ is less than 1, for each $\vec{\theta}$. This proves the theorem. □

It is not clear if condition (8.3) implies the weighting condition. To compare condition (8.3) with (8.1), note that

$$\rho(|B|) = \max_{\sigma} \rho(B_\sigma),$$

where B_σ is any matrix obtained from B by altering some of the complex signs of the elements of B, i.e., multiplying elements by complex numbers of modulus 1. Thus, condition (8.1) requires that all possible complex sign changes of the matrix B have spectral radius less than 1, while (8.3) requires only the matrices resulting from complex sign changes of the rows (or columns) have spectral radius less than 1.

We show the relationship between the conditions in the following diagram, where the arrows show that the one condition implies the other.

$$\begin{array}{ccccc} \rho(|B|) < 1 & \Rightarrow & B^{\mathrm{T}}DB < D & \Rightarrow & \rho(B^{(2)}) < 1 \\ & & \Downarrow & & \\ & & \rho_{\pm}(B) < 1 & \Rightarrow & \rho(B) < 1. \end{array} \tag{8.4}$$

Note that the two conditions to the right are independent. There are matrices satisfying one and not the other. In particular, there are matrices B for which the synchronous method will not converge, but the distributed memory asynchronous method will converge.

9. Conclusions

The probabilistic analysis presented in this paper provides useful convergence criteria for asynchronous iteration that are weaker than the criteria of Chazan and Miranker. The analysis of Chazan and Miranker is concerned with the worst possible case whereas the analysis here considers the probability of the different outcomes, obtaining results on the expected outcomes.

We also note that whereas for synchronous iteration, the elapsed time for a computation is directly related to the iteration count, for the models used here, the elapsed time is not so simply determined. One way to model the elapsed time would be to let the time between iterations be a random variable with a distribution such as a Poisson distribution. With such a model it may be possible to analyze the overall efficiency of the different models.

The conditions obtained here are for the convergence of the expectation of the error and the convergence of the variance (or the expectation of the square of the magnitude of the error). The convergence of the variance is the more significant issue, since the expected value of the error converges if the variance converges.

The comparison of the shared memory model and the distributed memory models is interesting since they have the same conditions for the convergence of the expectation of the error, but the variance requires different conditions. That the distributed memory model can be be convergent, in the sense of this paper, for matrices B for which the synchronous is non-convergent, is rather interesting.

The convergence results given here are not completely satisfactory because they are not both necessary and sufficient, but further analysis may sharpen the results given here. The analysis is more complex than that for the model of Chazan and Miranker and for synchronous methods because of the combination of linear algebra and probability. However, the results are very satisfying because they do give new insights into asynchronous computation. In particular, they show that asynchronous computation does not require the strong condition of Chazan and Miranker in order to be successful for practical computation.

References

[1] G. Baudet, Asynchronous iterative methods for multiprocessors, J. ACM 25 (1978) 225–244.
[2] D.P. Bertsekas, J.N. Tsitsiklis, Parallel and Distributed Computing, Numerical Methods, Prentice-Hall, Englewood Cliffs, NJ, 1989.
[3] D. Chazan, W. Miranker, Chaotic relaxation, Linear Algebra Appl. 2 (1969) 199–222.
[4] A. Frommer, D.B. Szyld, On asynchronous iterations, J. Comput. Appl. Math. 123 (2000) 201–216.
[5] L.A. Hageman, D.M. Young, Applied Iterative Methods, Academic Press, New York, 1981.
[6] Matlab, The MathWorks, Natick, MA.
[7] J.C. Strikwerda, A convergence theorem for chaotic asynchronous relaxation, Linear Algebra Appl. 253 (1997) 15–24.
[8] D.B. Szyld, The mystery of asynchronous iterations convergence when the spectral radius is one, Research Report 98–102, Department of Mathematics, Temple University, 1998.