Models of Parallel Chaotic Iteration Methods

Rafael Bru*
Department of Mathematics
University of Connecticut
Storrs, Connecticut 06268

Ludwig Elsner
Fakultät für Mathematik
Universität Bielefeld
Postfach 8640
Bielefeld 1, Federal Republic of Germany

and

Michael Neumann†
Department of Mathematics
University of Connecticut
Storrs, Connecticut 06268

Submitted by Robert J. Plemmons

ABSTRACT

We consider two models of parallel multisplitting chaotic iterations for solving large nonsingular systems of equations $Ax = b$. In the first model each processor can carry out an arbitrary number of local iterations before the next global approximation to the solution is formed. In the second model any processor can update the global approximation which resides in the central processor at any time. This model is a generalization of a sequential iterative scheme due to Ostrowski called the free steering group Jacobi iterative scheme and a chaotic relaxation point iterative scheme due to Chazan and Miranker. We show that when A is a monotone matrix and all the splittings are weak regular, both models lead to convergent schemes.

*The work of this author was supported in part by a grant from Conselleria de Cultura, Educació i Ciència, de la Generalitat Valenciana, Spain, while on leave from Universidad Politécnica de Valencia.

†The work of this author was supported in part by NSF Grant No. DMS-8700604.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/88/$3.50

1. INTRODUCTION

The parallel multisplitting iteration algorithm for solving large nonsingular systems of equations

$$Ax = b \tag{1.1}$$

was recently suggested by O'Leary and White (1985), and it has been further investigated in White (1986a,b), and in Neumann and Plemmons [1987]. According to O'Leary and White, multiple splittings of the coefficient matrix A into

$$A = M_l - N_l, \ \det(M_l) \neq 0, \qquad l = 1, \ldots, k, \tag{1.2}$$

are formed, and the *parallel iterative multisplitting procedure*

$$y^{(i)} = \sum_{l=1}^{k} E_l M_l^{-1} N_l y^{(i-1)} + \sum_{l=1}^{k} E_l M_l^{-1} b \tag{1.3}$$

is executed. Here E_l, $1 \leqslant l \leqslant k$, are nonnegative diagonal matrices, which we may refer to as *masking* (or *weighting*) matrices, such that

$$\sum_{l=1}^{k} E_l = I. \tag{1.4}$$

A key idea here is this: There are k processors connected to a host processor. At each major stage of the iteration the lth processor computes only those entries of the vector

$$M_l^{-1} N_l y^{(i)} + M_l^{-1} b \tag{1.5}$$

which correspond to the nonzero diagonal entries of E_l. The processor then scales these entries so as to be able to deliver the vector

$$E_l M_l^{-1} N_l y^{(i)} + M_l^{-1} b \tag{1.6}$$

to a central or host processor.

O'Leary and White (1985) and White (1986a, b) further investigate classes of matrices A and types of splitting (1.2) which lead to a convergent

multisplitting procedure (1.3). They show, for example, that if all the splittings in (1.2) are weak *regular*, that is,

$$M_l^{-1} \geqslant 0 \quad \text{and} \quad M_l^{-1}N_l \geqslant 0, \qquad l = 1, \ldots, k, \tag{1.7}$$

and if A is *monotone*, that is, A is nonsingular and

$$A^{-1} \geqslant 0, \tag{1.8}$$

then the *spectral radius*

$$\rho\left(\sum_{l=1}^{k} E_l M_l^{-1} N_l\right) < 1.$$

This implies that $\lim_{i \to \infty} y^{(i)} = A^{-1}b$ from any starting vector $y^{(0)}$. O'Leary and White also investigate parallel multisplitting iterative technique relative to the case when A in (1.1) is hermitian positive definite. Systems (1.1) with such a coefficient matrix will not be considered in this paper.

One aspect of attaining maximum efficiency in the implimentation of (1.3) is to choose the splittings in (1.2) and the masking diagonal matrices E_l, $1 \leqslant l \leqslant k$, so that the workload carried by all processors is roughly equally distributed. When such a balance can be achieved, then the individual processors are ready to contribute towards their update of the global iterate $y^{(i)}$ at the same time, which, in turn, minimizes idle time. However, there are applications such as finite element methods [arising, for example, from problems in aeronautical engineering (Przemieniecki, 1963), storm surge forecasting (Thacker et al., 1980), and semiconductor analysis (Buturla et al., 1981); see also Axelsson and Barker (1984) and Bykat (1983) for further references], in which materials or shapes of elements lead to problems (1.1) which quite naturally divide into subproblems of unequal size. To avoid loss of time and efficiency in processor utilization, we shall consider here two models of the so-called *parallel chaotic* or *asynchroneous* iterations:

Model A: Each processor can carry out a varying number of *local iterations* until a mutual phase time is reached when all processors are ready to contribute towards the global iteration.

Model B: Each processor can update the global approximation, or retrieve any subset of the components of the global approximation residing in the host processor, at any time.

We shall formulate these models mathematically in Section 2, and there it will be shown that when the matrix A in (1.1) is monotone and all the

splittings in (1.2) are weak regular, then both models lead to convergent schemes. Roughly speaking, the only other provision which we shall require for their convergence is that each processor contributes at least one local iteration to the global approximation infinitely often.

To conclude this introductory section we mention that, interestingly, Model B can be viewed as a generalization of the "free steering group block Jacobi method" which was suggested, in the context of *sequential* iteration, by Ostrowski (1961). In Section 3 we shall present Ostrowski's *free steering method* and indicate how Model B extends it. In addition we shall show that a special case of Model B yields the *point chaotic relaxation iteration methods* which were suggested by Chazan and Miranker (1969). Finally, the merits of the standard parallel block Jacobi iteration method will also be discuss

2. TWO CHAOTIC MODELS

Let $A = M_l - N_l$, $1 \leqslant l \leqslant k$, be k splittings satisfying the conditions of (1.2), and for each l define the operator $F_l: R^n \to R^n$, where R^n is the n-dimensional real space, as follows:

$$F_l(x) = M_l^{-1}N_l^{-1}x + M_l^{-1}b. \tag{2.1}$$

This is called *one local iteration*. Observe that in general F_l is an affine operator. Furthermore, for a nonnegative integer μ let

$$F_l^\mu = \begin{cases} F_l \circ F_l \circ \cdots \circ F_l & \mu > 0, \\ I, & \mu = 0 \end{cases}.$$

Note that μ is the number of compositions of F_l with itself.

We are now ready to formulate the mathematical version of Model A in the previous section.

MODEL A. Let E_l, $1 \leqslant l \leqslant k$ be k nonnegative diagonal matrices satisfying (1.4), and let $A = M_l - N_l$ be k splittings satisfying (1.2). Starting with an arbitrary vector $z^{(0)}$, perform the iteration

$$z^{(i)} = \sum_{l=1}^{k} E_l F^{\mu_{l,i}}(z^{(i-1)}), \tag{2.2}$$

where the F_l's are the operators given in (2.1) and where the $\mu_{l,i}$ are positive integers which can depend on both l, the processor, and i, the index of the iteration step.

We remark that we have permitted the number of compositions $\mu_{l,i}$ of F_l to depend not only on the processor, but also on the index of the present global step of the iteration, to allow for more generality. In practice we may expect the number of local iterations which each processor performs between two major steps of the algorithm to be fixed and to depend only upon A and the relative amont of work which is involved in computing the vectors (1.6) for $l = 1,\ldots,k$.

The simplest way of tackling the question of the convergence of the algorithm in (2.2) is to first develop an error analysis. For that purpose, for $i = 1,2,\ldots$ let us associate with the affine operator on the right hand side of (2.2) the linear operator

$$B_i := \sum_{l=1}^{k} E_l\left(M_l^{-1}N_l\right)^{\mu_{l,i}}, \tag{2.3}$$

in which case for some fixed vector $r^{(i)} \in R^n$,

$$\sum_{l=1}^{k} E_l F_l^{\mu_{l,i}}(z) = B_i z + r^{(i)} \qquad \forall z \in R^n. \tag{2.4}$$

Now let ξ be the exact solution to the system (1.1). Then, as $\xi = F_l(\xi)$ for each $l = 1,\ldots,k$, we have that

$$\xi = \sum_{l=1}^{k} E_l F_l^{\mu_{l,i}}(\xi) = B_i\xi + r^{(i)}, \qquad i = 1,2,\ldots. \tag{2.5}$$

Thus on defining the error vectors

$$e^{(i)} = z^{(i)} - \xi, \qquad i = 0,1,\ldots, \tag{2.6}$$

we see from (2.2), (2.4), and (2.5) that

$$e^{(i)} = B_i e^{(i-1)}. \tag{2.7}$$

It is this relation which permits us to develop, subsequent to a brief introduction of monotonic norms, sufficient conditions for the convergence of the algorithm given in (2.2).

For a vector $x \in R^n$, $x \gg 0$ ($x \geqslant 0$) will denote that all its components are positive (nonnegative). Similarly, for $x, y \in R^n$, $x \gg y$ ($x \geqslant y$) will mean that $x - y \gg 0$ ($x - y \geqslant 0$). For $x \in R^n$, $|x|$ will denote the vector whose components are the absolute value of the corresponding components of x. We shall employ similar notation for matrices.

Let $x \gg 0$. Then the functional

$$\|y\|_x = \inf\{\alpha > 0 \mid -\alpha x \leqslant y \leqslant \alpha x\} = \max_{1 \leqslant i \leqslant n} \left| \frac{y_i}{x_i} \right|$$

is a vector norm on R^n which is *monotonic* in the sense that $|u| \leqslant |w|$ implies that $\|u\|_x \leqslant \|w\|_x$. It is well known (see Rheinboldt and Vandergraft, 1973) that $\||B|x\|_x = \|B\|_x$, where $\|B\|_x$ denotes the matrix norm of B induced by the monotonic vector norm $\|\cdot\|_x$. It easily follows that if $x \gg 0$ is a vector and $\beta \geqslant 0$ is a scalar for which $|B|x \leqslant \beta x$, then $\|B\|_x \leqslant \beta$.

In order to introduce our first main result we require the following lemma.

LEMMA 2.1. *Let $A = M_l - N_l$, $l = 1, \ldots, k$, be k splittings of the $n \times n$ matrix A satisfying* (i.2), *and let E_l, $l = 1, \ldots, k$, be k nonnegative diagonal matrices satisfying* (1.3). *If there exists a vector $x \gg 0$ and a scalar $0 \leqslant \beta < 1$ such that*

$$|M_l^{-1} N_l| x \leqslant \beta x, \qquad l = 1, 2, \ldots, k, \tag{2.8}$$

then for each i, 1,2,...,

$$\|B_i\|_x \leqslant \beta, \tag{2.9}$$

where B_i is given in (2.3).

Proof. From (2.3) it follows that

$$|B_i| x = \left| \sum_{l=1}^{k} E_l \left(M_l^{-1} N_l \right)^{\mu_{l,i}} \right| x \leqslant \sum_{l=1}^{k} |E_l| |M_l^{-1} N_l|^{\mu_{l,i}} x$$

$$\leqslant \sum_{l=1}^{n} E_l \beta^{\mu_{l,i}} x \leqslant \beta \sum_{l=1}^{n} E_l x = \beta x,$$

showing that (2.9) is valid. ∎

We are now in a position to state our main result concerning the chaotic iteration scheme introduced in Model A.

THEOREM 2.1. *Suppose that A is an $n \times n$ monotone matrix and that all the splittings in* (1.2) *are weak regular. If E_l, $l=1,\dots,k$ are k nonnegative diagonal matrices satisfying* (1.4), *then the iteration* (2.2) *converges from any initial vector $z^{(0)}$ whenever*

$$\mu_{l,t} \geqslant 1, \qquad t=1,2,\dots, \quad l=1,\dots,k. \tag{2.10}$$

Proof. Since A is monotone, there exists a vector $x \gg 0$ [e.g. $x = A^{-1}(1 \ \cdots \ 1)^T$] such that $Ax \gg 0$. Thus, as $A = M_l - N_l$ is a regular splitting for $1 \leqslant l \leqslant k$, we see that for each such l,

$$x - M_l^{-1}N_l x = M_l^{-1}Ax \gg 0,$$

yielding

$$|M_l^{-1}N_l|x = M_l^{-1}N_l z \ll x.$$

Thus for a suitable constant $0 \leqslant \beta \leqslant 1$,

$$\|B_i\|_x \leqslant \beta, \qquad i=1,2,\dots.$$

But then, from (2.9), we have that

$$\begin{aligned} \lim_{i\to\infty} \|e^{(i)}\|_x &= \lim_{i\to\infty} \|B_i \cdots B_1 e^{(0)}\|_x \\ &\leqslant \lim_{i\to\infty} \|B_i\|_x \cdots \|B_1\|_x \|e^{(0)}\|_x \\ &\leqslant \lim_{i\to\infty} \beta^i \|e^{(0)}\|_x = 0 \end{aligned}$$

for all $e^{(0)} \in R^n$, proving our claim. ∎

We remark that the assumption in (2.10) can be weakened as follows:

$$\mu_{l,t} \geqslant 0, \qquad t=1,2,\dots, \quad l=1,2,\dots,k, \tag{2.11a}$$

and for infinitely many t's,

$$\mu_{l,t} \geqslant 1 \qquad \text{for all} \quad l=1,\dots,k. \tag{2.11b}$$

This is because under (2.11), $\|B_t\|_x \leqslant 1$ for all $t \geqslant 1$ with $\|B_t\|_x \leqslant \beta$ for infinitely many t's. The difference between the conditions (2.10) and (2.11) is that (2.11) permits, if necessary, for any processor to skip its contribution to any major step of the iteration provided that infinitely often all processors contribute simultaneously towards a global iteration.

Although we described Model A as a model for chaotic iterations, a certain degree of synchronization between the processors is still present in the model. This is evidently so when we consider that all the processors must complete their local iterations, however their number may vary between processor and processor, before the new global iterate can be computed in (2.2). In our second model any processor can update the global iterate $z^{(i)}$, which resides in the central processor, at any time.

In Model B the typical task of the lth processor can be divided into three subtasks which are cyclically implemented. These subtasks can be described as follows:

(i) If $\tilde{w}$ is the approximation to the global solution residing in the processor at the beginning of the cycle, the processor updates some or all of the components of $\tilde{w}$ by the corresponding components of the current approximation to the solution residing in the central processor and sets the revised approximation to w.

(ii) The processor computes the local iteration

$$u = E_l(F_l(w)) = E_l M_l^{-1} N_l w + E_l M_l^{-1} b. \tag{2.12}$$

(iii) If y is the approximation to the solution present in the central processor when the processor completes the computation of (2.11), the processor updates the host processor as follows:

$$z = u + (I - E_l)y. \tag{2.13}$$

To simplify our model we shall assume that no two processors update the central processor at the same instant; otherwise some queueing priorities are introduced into the algorithm or a more sophisticated processor ring is exploited. Suppose now that $t_1, t_2, \ldots$ are the times at which the central processor is updated by processors $i_1, i_2, \ldots$, where $1 \leqslant i_j \leqslant k$, $j = 1, 2, \ldots$.

DEFINITION 2.1. A sequence of integers $\mathcal{S}_0 = \{i_j\}_{j=1}^{\infty}$, $1 \leqslant i_j \leqslant k$, k fixed, is *admissible* if every one of the integers $1, \ldots, k$ appears in the sequence infinitely often. An admissible sequence is *regulated* if there exists a

positive integer T such that each of the integers $1,\dots,k$ appears at least once in any T consecutive elements of the sequence.

Algorithmically we can describe the above steps, including the initialization of the starting vectors, as follows:

MODEL B. Let E_l, $l=1,\dots,k$, be k nonnegative diagonal matrices satisfying (1.4). Given an admissible regulated sequence $\mathcal{S}_0=\{i_j\}_{j=1}^{\infty}$ and an initial vector $x^{(0)}$, perform

$$z_i^{(j)}=x_i^{[j+r(i,j)]}, \qquad i=1,\dots,n, \tag{2.14}$$

and

$$x^{(j+r_j)}=\left(I-E_{i_j}\right)x^{(j+r_j-1)}+E_{i_j}M_{i_j}^{-1}\left[N_{i_j}z^{(j)}+b\right],$$

$$j=-k+1,-k+2,\dots,0,1,2,\dots. \tag{2.15}$$

Here, for $j\leqslant 0$, $i_j=-j+1$, while for $j>0$, i_j is the processor that has updated the approximation residing in the central processor with $x^{(j)}$. Finally, r_j is the smallest positive integer such that $i_j=i_{j+r_j}$, $j\geqslant -k+1$. Evidently, for $j>0$, $r_j-1\leqslant T$, and it is equal to the number of times which the approximation residing in the central processor is updated by processors *other* than the i_jth during the time interval in which the i_jth processor executes its subtasks (i) and (ii). $0\leqslant r(i,j)\leqslant r_j-1$. Finally, $x^{(-k+1)}=x^{(-k+2)}=\dots=x^{(-1)}=x^{(0)}$.

We are now ready to prove our second main result of this paper, in which we give sufficient conditions for the convergence of the algorithm stipulated in (2.14) and (2.15).

THEOREM 2.2. *Suppose that the $n\times n$ coefficient matrix A in* (1.1) *is monotone and that each of the k splittings in* (1.2) *is weak regular. If E_l, $l=1,\dots,k$, are nonnegative diagonal matrices satisfying* (1.4), *then for any initial vector $x^{(0)}$ and for any admissible and regulated sequence of integers $\mathcal{S}_0=\{i_j\}_{j=1}^{\infty}$, $1\leqslant i_j\leqslant k$, the sequence of vectors generated by* (2.14) *and* (2.15) *with $0\leqslant r(i,j)\leqslant r_j-1$, $j\geqslant -k+1$, converges to the solution vector $\xi=A^{-1}b$ of the system* (1.1).

Proof. In order to analyze the convergence of (2.14) and (2.15) it will be convenient to embed this iteration procedure in an iteration procedure in

R^{nT}. For that purpose we introduce the following notation:

$$\tilde{\epsilon}^{(j)} = z^{(j)} - \xi, \qquad \epsilon^{(j)} = x^{(j)} - \xi,$$

$$\bar{x} = \begin{bmatrix} x \\ x \\ \vdots \\ x \end{bmatrix} \in R^{nT}, \quad \text{and} \quad \bar{\epsilon}_j = \begin{bmatrix} \epsilon^{(j)} \\ \vdots \\ \epsilon^{(j-T+1)} \end{bmatrix} \in R^{nT}. \tag{2.16}$$

Here $x \gg 0$ is a vector such that $Ax \gg 0$.

It follows from (2.14) and (2.16) that

$$\tilde{\epsilon}^{(j)} = S_{j+r_j} \bar{\epsilon}_{j+r_j-1}, \tag{2.17}$$

where S_{j+r_j} is an $n \times nT$ matrix such that for each $i = 1, \ldots, n$, the $r(i, j)$th entry in the ith row is unity and all other remaining entries are zero. Now, by (2.15) we have that

$$\epsilon^{(j+r_j)} = \left(I - E_{r_j}\right)\epsilon^{(j+r_j-1)} + E_{i_j} M_{i_j}^{-1} N_{i_j} \tilde{\epsilon}^{(j)}, \tag{2.18}$$

and so, by (2.16) and (2.17), we can write that

$$\bar{\epsilon}_{j+r_j} = B_{j+r_j} \bar{\epsilon}_{j+r_j-1}, \tag{2.19}$$

where B_{j+r_j} is the $(nT) \times (nT)$ matrix given by

$$B_{j+r_j} = \begin{bmatrix} I - E_{i_j} & 0 & \cdots & 0 \\ I & & & 0 \\ & \ddots & & \vdots \\ 0 & & I & 0 \end{bmatrix} + \begin{bmatrix} E_{i_j} M_{i_j}^{-1} N_{i_j} S_{j+r_j} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

From (2.19) it follows that

$$\bar{\epsilon}_{j+2T-1} = B_{j+2T-1} B_{j+2T-2} \cdots B_{j+1} \bar{\epsilon}_{j+1}.$$

Thus as $B_\nu \geqslant 0$, $\nu \geqslant 1$, to show that $\bar{e}_\nu \to 0$ as $\nu \to \infty$ it clearly suffices to show that a constant $0 < \gamma < 1$ independent of j can be found such that

$$B_{j+2T-1} B_{j+2T-2} \cdots B_{j+1} \bar{x} \leqslant \gamma \bar{x} \qquad \forall j \geqslant 0. \tag{2.20}$$

For this purpose let us define

$$\delta = \min_{\substack{1 \leqslant p \leqslant n \\ 1 \leqslant l \leqslant k}} \{(E_l)_{pp} | (E_l)_{pp} > 0\}$$

and set

$$\gamma = 1 - (1 - \beta)\delta,$$

where, similarly to the proof of Theorem 2.1, $0 \leqslant \beta < 1$ is a number such that

$$|M_l^{-1} N_l x| \leqslant \beta x, \qquad l = 1, \ldots, k.$$

As $\delta \leqslant 1$, it follows that $0 \leqslant \gamma < 1$. Now let

$$w_\nu = B_{j+\nu} B_{j+\nu-1} \cdots B_{j+1} \bar{x} = \begin{bmatrix} w_\nu^1 \\ \vdots \\ w_\nu^T \end{bmatrix}, \tag{2.21}$$

where $w_\nu^s \in R^n$, $1 \leqslant s \leqslant T$, and $1 \leqslant \nu \leqslant 2T - 1$. Note that w_ν depends on j, but that for convenience we have suppressed this subscript. We next show that

$$w_\nu^1 \leqslant \gamma x, \qquad \nu \geqslant T. \tag{2.22}$$

From the obvious inequality

$$B_\nu \bar{x} \leqslant \bar{x}, \qquad \nu \geqslant 1, \tag{2.23}$$

we have that $w_\nu^s \leqslant x$, $1 \leqslant s \leqslant t$, and hence, as $S_j \bar{x} = \bar{x}$ for all $j \geqslant 1$, we can write for some suitable integer p that

$$\begin{aligned} w_\nu^1 &= (I - E_p) w_{\nu-1}^1 + E_p M_p^{-1} N_p S_j w_{\nu-1} \\ &\leqslant (I - E_p) x + E_p M_p^{-1} N_p \bar{x} \\ &\leqslant (I - E_p) x + \beta E_p x = x + (\beta - 1) E_p x. \end{aligned} \tag{2.24}$$

Suppose now that $\tau \in \{1,2,\ldots,n\}$, and observe first that if $(E_p)_{\tau\tau} = 0$, then by the leftmost equality in (2.24), the τth component $(w_p^1)_\tau = (w_{p-1}^1)_\tau$. However, if $(E_p)_{\tau\tau} > 0$ so that $(E_p)_{\tau\tau} \geq \delta$, then according to (2.24),

$$\left(w_p^1\right)_\tau \leq \gamma x_\tau.$$

Next, as the sequence $\{i_j\}_{j=1}^{\infty}$ is regulated all the integers $1,\ldots,k$ appear in its subsequence $\{i_{j+1},\ldots,i_{j+p}\}$, $p \geq T$. Because of (1.4) this means that

$$\left(w_p^1\right)_\tau \leq \gamma x_\tau \qquad \forall \tau = 1,\ldots,n$$

whenever $p \leq T$, which proves (2.22).

Finally, from (2.21) and the definition of the B_j's it follows that

$$w_{2T-1}^{1+s} = w_{2T-s-1}^1, \qquad s = 1,\ldots,T-1,$$

which shows that

$$w_{2T-1} \leq \gamma \bar{x}.$$

Thus (2.20) holds and the proof is done. ∎

We remark the following:

(i) The requirement in Theorem 2.2 that the sequence $\mathcal{S}_0 = \{i_j\}_{j=1}^{\infty}$ be regulated can be weakened to the assumption that it is only admissible, provided that we stipulate that there exist an integer T such that

$$r_j - r(i,j) \leq T-1,$$

which, in fact, is an implicit implication of the assumptions in (2.14).

(ii) Other requirements in Theorem 2.2 can also be weakened. For example, the assumption that A is monotone and that each of the k splittings in (1.2) are regular can be replaced by the following: There exists a positive vector x and a constant $\beta < 1$ such that each of the splittings in (1.2) satisfies

$$|M_i^{-1}N_i|x \leq \beta x.$$

Only minor modifications in the proof of Theorem 2.2, such as replacing B_j

by $|B_j|$ and w_j by $|w_j|$, are then necessary to accommodate these more general assumptions in the revised statement of the results.

3. SOME ILLUSTRATIONS: THE OSTROWSKI FREE STEERING METHODS AND THE CHAZAN-MIRANKER CHAOTIC ITERATION

In the introduction we alluded to Ostrowski's (1961) *free steering method* and Chazan and Miranker's *chaotic relaxation method* as having motivated our models here and, in particular, Model B. In this short section we wish to illustrate more specifically the motivating ideas in these works and their relation to our results. We also wish to show that the usual block Jacobi iteration method executed in parallel can be viewed as a special case of Model A.

Suppose that the $n \times n$ coefficient matrix A in (1.1) admits the block partitioning into

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1k} \\ \vdots & \ddots & \vdots \\ A_{k1} & \cdots & A_{kk} \end{bmatrix}, \tag{3.1}$$

where the diagonal blocks are square and nonsingular of dimensions $n_l \times n_l$, respectively, with $\sum_{l=1}^{k} n_l = n$. Partition the solution vector ξ to (1.1) and the constant vector b in conformity with (3.1) as follows:

$$\xi = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_k \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}. \tag{3.2}$$

It is simple to check by substituting (3.1) and (3.2) into (1.1) that for each $l = 1, \ldots, k$,

$$\xi_l = -\sum_{\substack{j=1 \\ j \neq l}}^{k} A_{ll}^{-1} A_{lj} \xi_j + A_{ll}^{-1} b_l. \tag{3.3}$$

Motivated by (3.3), Ostrowski suggested the following iterative scheme for approximating ξ: Beginning with an initial vector $x^{(0)}$, for $p = 1,2,\ldots$ per-

form

$$x_m^{(p)} = \begin{cases} -\sum_{\substack{j=1 \\ j \neq m}}^{k} A_{mm}^{-1} A_{mj} x_j^{(p-1)} + A_{mm} b_m, & m = i_p, \\ x_m^{(p-1)}, & m \neq i_p, \end{cases} \tag{3.4}$$

where $\{i_p\}_{p=1}^{\infty}$ is an admissible sequence. Ostrowski refers to the index i_p [or rather to the indices of the entries of $(x^{(p)})_{i_p}$] as the *steering* or *active* set in the pth stage of the iteration. He refers to the indices of the remaining entries of $x^{(p)}$ as the *passive set* of indices.

Ostrowski's algorithm can be represented as a special sequential case of Model B in the sense that the central processor is not updated by another processor while the lth processor executes subtask (ii) of the previous section. To see this, for $l = 1, \ldots, k$ define the splitting $A = M_l - N_l$ as follows:

$$A = \begin{bmatrix} A_{11} & & \\ & \ddots & \\ & & A_{kk} \end{bmatrix} - \begin{bmatrix} 0 & A_{12} & \cdots & A_{1,k-1} & A_{1k} \\ A_{21} & 0 & \cdots & A_{2,k-1} & A_{2k} \\ \vdots & & \ddots & \vdots & \vdots \\ A_{k-1,1} & A_{k-1,2} & \cdots & 0 & A_{k-1,k} \\ A_{k1} & A_{k2} & \cdots & A_{k,k-1} & 0 \end{bmatrix}, \tag{3}$$

and define the diagonal matrices

$$E_l = \begin{bmatrix} 0 & & & & & \\ & 0 & & & & \\ & & \ddots & & & \\ & & & I & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}, \tag{3.6}$$

where I occurs in the (l, l) block diagonal position and is the identity matrix of order n_l. Then by (3.5) and (3.6) the vector $x^{(p)}$ in (3.4) can be computed from $x^{(p-1)}$ via

$$x^{(p)} = E_{i_p} M_{i_p}^{-1} N_{i_p} x^{(p-1)} + E_{i_p} M_{i_p}^{-1} b + \left(I - E_{i_p}\right) x^{(p-1)}. \tag{3.7}$$

It is now evident that with $r_p = 1$, $p = 1,2,\ldots$, and with $r(i,p) = 0$ for $p = 1,\ldots,n$, (3.7) is just a special case of (2.14) and (2.15). Hence under the weakened assumptions in the remarks following the proof of Theorem 2.2, Ostrowski's algorithm, namely, the process given in (3.4), converges to the solution ξ to (1.1) from any starting vector $x^{(0)}$.

D. Chazan and W. Miranker (1969) considered the following point iteration scheme for solving the system (1.1), where they assume A to be given in the form $A = I - B$:

$$x_m^{(j+1)} = \begin{cases} x_m^{(j)}, & m \neq i_j, \\ b_m + \sum_{i=1}^{n} b_{mi} x_i^{j-r(i,j)}, & m = i_j, \end{cases} \tag{3.8}$$

where $\{i_j\}_{j=1}^{\infty}$ is assumed to be an admissible sequence. They show that when $\rho(|B|) < 1$ and there exists a fixed integer T such that

$$0 \leqslant r(i,j) \leqslant T - 1,$$

then the iterates generated by (3.8) converge to $\xi = A^{-1}b$ from every initial vector $x^{(0)}$. It is obvious that on taking $k = n$, $M_l = I$, $N_l = B$, and $E_l = (\delta_{lj})$ for $1 \leqslant l \leqslant n$, the algorithm in (3.8) satisfies all the requirements discussed in the remarks following Theorem 2.2, and hence Chazan and Miranker's algorithm is another special case of a convergent Model B.

We have carried out numerous numerical experiments on Ostrowski's sequential free steering method. We have also simulated a parallelized version of Ostrowski's method applying Model B to the solution of (1.1) with $A = M_l - N_l$ and E_l, $l = 1,\ldots,k$, as in (3.5) and (3.6), respectively, but with r_p not equal to zero in general. Our numerical results seem to indicate that the rate of convergence of Ostrowski's sequential algorithm is better than its parallelized version. Of course the computation of the iterates in Ostrowski's algorithm cannot enjoy the speedup due to the parallelism of the machine. We comment that a *vectorized* form of Ostrowski's sequential free steering model was recently considered by Hayes (1984).

We finally consider the "usual" parallel block Jacobi method. With the partitioning of A in (3.1) we can associate k affine operators $G_l: R^n \to R^n$, $l = 1,\ldots,k$, as follows:

$$G_l(x) = H_l x + \tilde{b}_l, \tag{3.9}$$

where

$$H_l = \begin{pmatrix} I & & & & & & 0 \\ & \ddots & & & & & \\ & & I & & & & \\ B_{l1} & \cdots & B_{l,l-1} & 0 & B_{l,l+1} & \cdots & B_{l,k} \\ & & & & I & & \\ & & & & & \ddots & \\ 0 & & & & & & I \end{pmatrix} \tag{3.10}$$

with $B_{lj} = -A_{ll}^{-1}A_{lj}$, $j \neq l$, $1 \leqslant j \leqslant k$, and

$$\bar{b}_l = \begin{pmatrix} 0 \\ \vdots \\ b_l \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{3.11}$$

Now let E_l^i, $l = 1,\ldots,k$, be k nonnegative diagonal matrices satisfying (1.4), and consider a version of Model A applied to the parallel block Jacobi scheme as follows:

$$z^{(i)} = \sum_{l=1}^{k} E_l^i G_l^{\mu_{l,i}} z^{(i-1)}, \tag{3.12}$$

where $\mu_{l,i} \geqslant 1$ $i = 1,2,\ldots$, $l = 1,\ldots,k$. One readily observes for each $l = 1,\ldots,k$, H_l is a projection matrix and that $G_l^t x = x$ $\forall t \geqslant 1$. Thus (3.12) reduces to the parallel block Jacobi method, namely,

$$z^{(i)} = \sum_{l=1}^{k} E_l^i G_l z^{(i-1)}. \tag{3.13}$$

The relation (3.12) and the fact that it reduces to (3.13) suggest the following: On the one hand the parallel block Jacobi method can be best utilized when the diagonal blocks in (3.1) are roughly of the same dimension, since no

advantage can be gained by allowing the processors to perform more than one local iteration until a time is reached at which all the processors are ready to update towards the new global iterate $z^{(t)}$. On the other hand, the error analysis used in and before the proof of Theorem 2.1 [see, for example, (2.14)] indicates that generally the rate of convergence increases with an increase in the number of local iterations carried out by each processor between any two major steps of the algorithm. Thus in terms of the rate of convergence, the parallel block Jacobi method already possesses a certain optimality.

The authors wish to thank the referee and Professor Robert J. Plemmons for numerous suggestions which have led to many improvements in the original manuscript. They also wish to thank several of their colleagues within and without the University of Connecticut for very helpful comments and conversations.

REFERENCES

Axelsson, O. and Barker, V. A. 1984. *Finite Element Solution of Boundary Value Problems*, Academic, New York.

Berman, A. and Plemmons, R. J. 1979. *Nonnegative Matrices in the Mathematical Sciences*, Academic, New York.

Buturla, E. M., Cottrell, P. E., Grossman, B. M., and Salsburg, K. A. 1981. Finite element analysis of semiconductor devices: The FIELDAY program, *IBM J. Res. Develop.* 25:218–231.

Bykat, A. 1983. Design of recursive, shape controlling mesh generator, *Internat. J. Numer. Methods Engrg.* 19:1375–1390.

Chazan, D. and Miranker, W. 1969. Chaotic relaxation, *Linear Algebra Appl.* 2:199–222.

Hayes, L. J. 1984. A vectorized matrix-vector multiply and overlapping block iterative method, in *Supercomputer Applications* (R. W. Numrick, Ed.), Plenum, New York, pp. 91–100.

Neumann, M. and Plemmons R. J. 1987. Convergence of parallel multisplitting iterative methods for *M*-matrices, *Linear Algebra Appl.* 88/89:559–573.

O'Leary, D. P. and White, R. E. 1985. Multisplittings of matrices and parallel solutions of linear systems, *SIAM J. Algebraic Discrete Methods* 6:630–640.

Ortega, J. M. and Rheinboldt, W. C. 1967. Monotone iterations for nonlinear equations with applications to Gauss-Seidel methods, *SIAM J. Numer. Anal.* 4:171–190.

A. M. Ostrowski, 1961. Iterative solution of linear systems of functional equations, *J. Math. Anal. Appl.* 2:351–369.

Przemieniecki, J. S. 1963. Matrix structural analysis of substructures, *AIAA J.* 1:138–147.

Rheinboldt, W. C. and Vandergraft, J. 1973. A simple approach to the Perron-Frobenius theory for positive operators on general partially ordered finite-dimensional spaces, *Math. Comp.* 27:139–145.

Thacker, W. C., Gonzalez, A., and Putland, G. E. 1980. A method for automating the constructional grids for storm surge forecast models, *J. Comput. Phys.* 37:371–387.

Varga, R. S. 1961. *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N.J.

White, R. E. 1986a. A nonlinear parallel algorithm with applications to the Stefan problem, *SIAM J. Numer. Anal.* 23:639–652.

White, R. E. 1986b. Parallel algorithms for nonlinear problems, *SIAM J. Algebraic Discrete Methods* 7:137–149.

Received August 1986; revised 28 July 1987