Numer. Math. 69: 141–153 (1994)

Numerische Mathematik

Asynchronous two-stage iterative methods

Andreas Frommer[1,⋆], **Daniel B. Szyld**[2,⋆⋆]

[1] Fachbereich Mathematik, Bergische Universität GH Wuppertal, Gauss-Strasse 20, D-42097 Wuppertal, Germany
e-mail: frommer@math.uni-wuppertal.de
[2] Department of Mathematics, Temple University, Philadelphia, Pennsylvania 19122-2585, USA
e-mail: szyld@math.temple.edu

Received September 7, 1993 / Revised version received April 21, 1994

Dedicated to Prof. Wilhelm Niethammer on the occasion of his sixtieth birthday

Summary. Parallel block two-stage iterative methods for the solution of linear systems of algebraic equations are studied. Convergence is shown for monotone matrices and for H-matrices. Two different asynchronous versions of these methods are considered and their convergence investigated.

Mathematics Subject Classification (1991): 65F10

1. Introduction and preliminaries

Let us consider the solution of the algebraic linear system of equations

$$Ax = b, \tag{1}$$

where x and b are n-vectors, and the square nonsingular matrix A is partitioned into $q \times q$ blocks

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1q} \\ A_{21} & A_{22} & \cdots & A_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ A_{q1} & A_{q2} & \cdots & A_{qq} \end{bmatrix} \tag{2}$$

with the diagonal blocks A_{ii} being square of order n_i, $i = 1, \cdots, q$, and $\sum_{i=1}^{q} n_i = n$.

Block versions of classical iterative methods are well known; see e.g. Varga [26] or Young [28]. Parallel computation makes Block Jacobi type methods particularly attractive. In such methods, a splitting $A = M - N$ is used, where M is block diagonal, denoted $M = \text{diag}(M_i)$, with the blocks M_i nonsingular of order n_i, $i = 1, \cdots, q$. The vectors x, b, and other intermediate vectors are partitioned in a way consistent with (2). The block iterative method is then

⋆ This work originated during a visit of this author at Temple University supported by the Deutsche Forschungsgemeinschaft
⋆⋆ This work was supported by National Science Foundation grants DMS-9201728 and INT-9123273

Algorithm 1.1. (Block Jacobi)
Given an initial vector $x_0^{\mathrm{T}} = [x_0^{(1)}, \cdots, x_0^{(q)}]$
    for $k = 1, 2, \cdots$
        for $i = 1$ **to** q
(3) $M_i x_k^{(i)} = (Nx_{k-1} + b)^{(i)}$

In the standard Block Jacobi method, the blocks in the diagonal are chosen as $M_i = A_{ii}$. Other choices of the blocks M_i, such as changing some entries in A_{ii}, can also be considered without any change in the analysis; cf. Varga [25], [26].

To implement Algorithm 1.1 on a parallel computer, suppose for example that we have q processors. During iteration k the processors can solve the q equations (3) simultaneously, since they are independent from each other. Before starting the next iteration $k+1$, however, all results from the preceding iteration must be made available to all processors because they might be needed to calculate Nx_k. We remark that the computational model underlying our algorithms is more general, e.g. the number of processors does not need to be equal to the number of diagonal blocks; cf. Elsner, Neumann and Vemmer [10].

In practice, it may be too expensive to solve (3) directly. Instead, another (inner) iterative method can be used to approximate the solution of (3) yielding the type of two-stage methods which are considered in this paper; see [12], [16], and the references given therein. We point out that, since the number of inner iterations may vary from block to block, the convergence results in Baudet [1] or Chazan and Miranker [5] cannot be applied to our situation.

In Sect. 2, we derive two new convergence results for block two-stage iterative methods. We then investigate asynchronous variations of our two-stage methods. These asynchronous methods arise naturally in parallel computations if one tries to reduce idle times of the processors. In Sects. 3 and 4, we introduce two different asynchronous models for block two-stage methods and investigate their convergence. As our major result, we establish convergence for both asynchronous models under the same conditions as those for the synchronous method.

In the rest of this section we present some notation, definitions and preliminary results which we refer to later.

We say that a vector x is nonnegative (positive), denoted $x \geq 0$ ($x > 0$), if all its entries are nonnegative (positive). Similarly, a matrix B is said to be nonnegative, denoted $B \geq O$, if all its entries are nonnegative or, equivalently, if it leaves invariant the set of all nonnegative vectors. We compare two matrices $A \geq B$, when $A-B \geq O$, and two vectors $x \geq y$ ($x > y$) when $x-y \geq 0$ ($x-y > 0$). Given a matrix $A = (a_{ij})$, we define the matrix $|A| = (|a_{ij}|)$. It follows that $|A| \geq O$ and that $|AB| \leq |A|\ |B|$ for any two matrices A and B of compatible size.

Let $Z^{n\times n}$ denote the set of all real $n \times n$ matrices which have all non–positive off–diagonal entries. A nonsingular matrix $A \in Z^{n\times n}$ is called M-matrix if $A^{-1} \geq O$, i.e. if A is a monotone matrix; see e.g. Berman and Plemmons [2] or Varga [26]. By $\rho(A)$ we denote the spectral radius of the square matrix A.

For any matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$, we define its comparison matrix $\langle A\rangle = (\alpha_{ij})$ by $\alpha_{ii} = |a_{ii}|, \quad \alpha_{ij} = -|a_{ij}|,\ i \neq j$. Following Ostrowski [21], [22], A is said to be an H-matrix if $\langle A\rangle$ is an M-matrix. Of course, M-matrices are special cases of H-matrices. H-matrices arise in many applications and were studied by a number of authors in connection with iterative solutions of linear systems; see the classical paper by Varga [27], or Frommer and Szyld [12] for an extensive bibliography and for an example that shows that H-matrices need not be monotone.

Lemma 1.2. *Let $A, B \in \mathbb{R}^{n\times n}$.*
(a) If A is an M-matrix, $B \in Z^{n\times n}$, and $A \le B$, then B is an M-matrix.
(b) If A is an H-matrix, then $|A^{-1}| \le \langle A\rangle^{-1}$.
(c) If $|A| \le B$ then $\rho(A) \le \rho(B)$.

Proof. (a) and (c) can be found, e.g. in [20], 2.4.10 and 2.4.9, respectively. Part (b) goes back to Ostrowski [21]; see also e.g. [19]. □

Definition 1.3. Let $A \in \mathbb{R}^{n\times n}$. The representation $A = M - N$ is called a splitting if M is nonsingular. It is called a convergent splitting if $\rho(M^{-1}N) < 1$. A splitting $A = M - N$ is called
(a) regular if $M^{-1} \ge O$ and $N \ge O$ [25], [26],
(b) weak regular if $M^{-1} \ge O$ and $M^{-1}N \ge O$ [2], [20],
(c) H-splitting if $\langle M\rangle - |N|$ is an M-matrix [12], and
(d) H-compatible splitting if $\langle A\rangle = \langle M\rangle - |N|$ [12].

Lemma 1.4. *Let $A = M - N$ be a splitting.*
(a) If the splitting is weak regular, then $\rho(M^{-1}N) < 1$ if and only if $A^{-1} \ge 0$.
(b) If the splitting is an H-splitting, then A and M are H-matrices and $\rho(M^{-1}N) \le \rho(\langle M\rangle^{-1}|N|) < 1$.
(c) If the splitting is an H-compatible splitting and A is an H-matrix, then it is an H-splitting and thus convergent.

Proof. (a) can be found, e.g. in [2], [20], [26]. The first part of (b) was shown in [17], [19]. The second part as well as (c) is found in [12]. □

Lemma 1.5. [16] *Given a nonsingular matrix $A \in \mathbb{R}^{n\times n}$ and $T \in \mathbb{R}^{n\times n}$ such that $(I - T)^{-1}$ exists, there exists a unique pair of matrices M, N, such that $T = M^{-1}N$ and $A = M - N$, where M is nonsingular. The matrices are $M = A(I - T)^{-1}$ and $N = AT(I - T)^{-1}$.*

Lemma 1.6. [15], [23] *Let $A \in \mathbb{R}^{n\times n}$. Let $v \in \mathbb{R}^n$, $v > 0$, and $\alpha \ge 0$ be such that $|A|v \le \alpha v$. Then for all $y \in \mathbb{R}^n$*

$$\|Ay\|_v \le \alpha\|y\|_v,$$

where $\|\cdot\|_v$ denotes the weighted max-norm $\|y\|_v = \max_{j=1,\cdots,n} |\frac{1}{v_j} y_j|$.

Lemma 1.7. [24] *Let $T_1,\ T_2, \cdots, T_k, \cdots$ be a sequence of nonnegative matrices in $\mathbb{R}^{n\times n}$. If there exist a real number $0 \le \theta < 1$, and a vector $v > 0$ in $\mathbb{R}^n$, such that*

$$T_j v \le \theta v,\ \ j = 1, 2, \cdots,$$

then $\rho(H_k) \le \theta^k < 1$, where $H_k = T_k \cdots T_2 \cdot T_1$, and therefore $\lim_{k\to\infty} H_k = O$.

Lemma 1.7 also follows from Lemma 1.6, since we obtain (with the operator norm induced by $\|\cdot\|_v$) that $\|T_j\|_v \le \theta$ for all j.

2. Block two-stage methods

In this section we consider block two-stage iterative methods, i.e. methods in which the solution of each system in (3) is in turn solved iteratively. To that end let $M_i = F_i - G_i$ be a (convergent) splitting of M_i and let $s(i, k)$ be the number of (inner) iterations used in the approximation to the solution of (3) for the i-th block at the k-th (outer) iteration.

Algorithm 2.1. (Block Two-stage)
Given an initial vector $x_0^{\mathrm{T}} = [x_0^{(1)}, \cdots, x_0^{(q)}]$
for $k = 1, 2, \cdots$
  for $i = 1$ **to** q
  $y_0^{(i)} = x_{k-1}^{(i)}$
    for $j = 1$ **to** $s(i, k)$
    $F_i y_j^{(i)} = G_i y_{j-1}^{(i)} + (Nx_{k-1} + b)^{(i)}$
  $x_k^{(i)} = y_{s(i,k)}^{(i)}$

The number of iterations $s(i, k)$ might be prescribed in advance or it can be determined at each step using some inner convergence criteria as is done e.g. by Elman and Golub [7] or Golub and Overton [13], [14].

As in the case of Block Jacobi, the operations corresponding to each diagonal block can be performed by different processors in parallel. Algorithm 2.1 is synchronous in the sense that the k-th iteration cannot start until all block-components of the $(k-1)$-th iteration have been completed. This synchronous algorithm can be seen as a special case of Algorithm 4.1 of Lanzkron, Rose and Szyld [16], and is more general than Model A of Bru, Elsner and Neumann [4].

In order to analyze the convergence of Algorithm 2.1, let $R_i = F_i^{-1} G_i$ and write

$$x_k^{(i)} = R_i^{s(i,k)} x_{k-1}^{(i)} + \sum_{j=0}^{s(i,k)-1} R_i^j F_i^{-1} (Nx_{k-1} + b)^{(i)}. \tag{4}$$

Consider the following two $n \times n$ block diagonal matrices: $R(k) = \mathrm{diag}(R_i^{s(i,k)})$ and

$$Q(k) = (I - R(k))\, M^{-1} = \mathrm{diag}\left(\left(I - R_i^{s(i,k)}\right) M_i^{-1}\right). \tag{5}$$

Since

$$\begin{aligned} &\sum_{j=0}^{s(i,k)-1} R_i^j F_i^{-1} \\ &= \left(I - R_i^{s(i,k)}\right)(I - R_i)^{-1} F_i^{-1} = \left(I - R_i^{s(i,k)}\right) M_i^{-1},\ \ i = 1, \cdots, q, \end{aligned} \tag{6}$$

we can collect the block-components in (4) and write one iteration of Algorithm 2.1 as

$$x_k = R(k)x_{k-1} + Q(k)(Nx_{k-1} + b) = T(k)x_{k-1} + Q(k)b, \tag{7}$$

where

$$T(k) = R(k) + Q(k)N = R(k) + (I - R(k))M^{-1}N. \tag{8}$$

Provided $\rho(R_i) < 1$ for $i = 1, \cdots, q$, we observe that, as we would expect, $T(k)$ tends to $M^{-1}N$, the Block Jacobi iteration matrix, if for all $i = 1, \cdots, q$ we have $\lim_{k\to\infty} s(i,k) = \infty$.

It is not hard to see that the iteration (7) is consistent, i.e. if $x_* = A^{-1}b$, then

$$x_* = T(k)x_* + Q(k)b. \tag{9}$$

Thus, if the error at the k-th iteration is $e_k = x_k - x_*$, we have

$$e_k = T(k)e_{k-1} = T(k)\cdot T(k-1)\cdots T(1)e_0. \tag{10}$$

In the rest of the section we show the convergence of the block two-stage method (7) in two circumstances. The first applies to the case of A monotone and the second to that of A an H-matrix. We point out that only the monotone case was studied in [4] and [16].

Theorem 2.2. *Let A such that $A^{-1} \geq O$ be partitioned as in (2). Let $A = M - N$ be a regular splitting with $M = \mathrm{diag}(M_i)$, and let $M_i = F_i - G_i$ be weak regular splittings, $i = 1, \cdots, q$. Then, the block two-stage iterative method (7) converges to $x_* = A^{-1}b$ for any initial vector x_0 and for any sequence of numbers of inner iterations $s(i,k) \geq 1$, $i = 1, \cdots, q$, $k = 1, 2, \cdots$*

Proof. We can write

$$T(k) = Q(k)[Q(k)^{-1}R(k) + N] \quad \text{and} \quad A = Q(k)^{-1} - [Q(k)^{-1}R(k) + N], \tag{11}$$

cf. Lemma 1.5. The existence of $Q(k)^{-1}$ follows from Lemma 1.4 (a) and the fact that $M^{-1} \geq O$ if and only if $M_i^{-1} \geq O$ for all $i = 1, \cdots, q$. From the hypothesis and the identities (5), (6) and (8) it follows that $Q(k) \geq O$ and $T(k) \geq O$. From (11) it follows that $Q(k)A = I - T(k)$ and thus

$$T(k) = I - Q(k)A. \tag{12}$$

Consider any fixed vector $e > 0$ (e.g. with all components equal to 1), and $v = A^{-1}e$. Since $A^{-1} \geq O$ and no row of A^{-1} can have all null entries, we get $v > 0$. Let $F = \mathrm{diag}(F_i)$. By the same arguments $F^{-1}e > 0$. We have from (12), (5) and (6) that

$$T(k)v = (I - Q(k)A)v = v - Q(k)e = v - F^{-1}e - \mathrm{diag}\left(\sum_{j=1}^{s(i,k)-1} (R_i)^j F_i^{-1}\right) e.$$

Observe that $\mathrm{diag}\left(\sum_{j=1}^{s(i,k)-1}(R_i)^j F_i^{-1}\right) e \geq 0$. Moreover, since $T(k)v \geq 0$ and $v - F^{-1}e < v$, there exists $0 \leq \theta < 1$ such that $v - F^{-1}e \leq \theta v$. Thus

$$T(k)v \leq \theta v, \ \text{for all } k = 1, 2, \cdots \tag{13}$$

By Lemma 1.7 this implies that the product $H(k) = T(k)\cdot T(k-1)\cdots T(1)$ tends to zero as $k \to \infty$, and thus $\lim\limits_{k\to\infty} e_k = 0$. The bounds (13) are independent of the sequence $s(i,k) \geq 1$, $i = 1, \cdots, q$, $k = 1, 2, \cdots$ $\square$

Theorem 2.2 is a slight generalization of a part of Theorem 7.3 in [16], where the splittings $M_i = F_i - G_i$ were assumed to be regular splittings. The proof here is of a different kind but its technique is similar to that used in Theorem 2.1 in [4] and in Theorems 4.3 and 4.4 in [12]. It is needed as part of the proof of the following new result.

Theorem 2.3. *Let A be an H-matrix partitioned as in (2). Let $A = M - N$ be an H-splitting with $M = \operatorname{diag}(M_i)$, and let $M_i = F_i - G_i$ be H-compatible splittings, $i = 1, \cdots, q$. Then, the block two-stage iterative method (7) converges to $x_* = A^{-1}b$ for any initial vector x_0 and for any sequence of numbers of inner iterations $s(i,k) \geq 1$, $i = 1, \cdots, q$, $k = 1, 2, \cdots$*

Proof. We write

$$T(k) = \operatorname{diag}\left((F_i^{-1}G_i)^{s(i,k)}\right) + \operatorname{diag}\left(\sum_{j=0}^{s(i,k)-1}(F_i^{-1}G_i)^j F_i^{-1}\right) N$$

and use Lemma 1.2 (b) to obtain the following bound

$$\begin{aligned} |T(k)| &\leq \operatorname{diag}\left((|F_i^{-1}||G_i|)^{s(i,k)}\right) + \operatorname{diag}\left(\sum_{j=0}^{s(i,k)-1}(|F_i^{-1}||G_i|)^j |F_i^{-1}|\right)|N| \\ &\leq \operatorname{diag}\left((\langle F_i\rangle^{-1}|G_i|)^{s(i,k)}\right) + \operatorname{diag}\left(\sum_{j=0}^{s(i,k)-1}(\langle F_i\rangle^{-1}|G_i|)^j \langle F_i\rangle^{-1}\right)|N|. \qquad (14) \end{aligned}$$

Let us denote by $\bar{T}(k)$ the matrix on the right hand side of (14). This is the matrix corresponding to the k-th iteration of a block two-stage method for the monotone matrix $\langle M\rangle - |N|$ with the regular splitting $\langle M\rangle - |N|$ and $\langle M\rangle = \operatorname{diag}(\langle M_i\rangle)$ with $\langle M_i\rangle = \langle F_i\rangle - |G_i|$. These matrices and splittings satisfy the hypothesis of Theorem 2.2 and we have, as in (13),

$$|T(k)|v \leq \bar{T}(k)v \leq \theta v, \text{ for all } k = 1, 2, \cdots \qquad (15)$$

for some $v \in \mathbb{R}^n$, $v > 0$ and $\theta \in [0, 1)$. Let $H(k) = T(k) \cdot T(k-1) \cdots T(1)$. We can bound $|H(k)| \leq |T(k)| \cdot |T(k-1)| \cdots |T(1)|$. Therefore by (15) and Lemma 1.7, $H(k)$ tends to zero as $k \to \infty$, implying $\lim_{k\to\infty} e_k = 0$. □

3. Outer asynchronous two-stage methods

Consider a parallel implementation of a block two-stage method where, unlike in Algorithm 2.1, the processors are allowed to start the computation of the next iterate of the block-component assigned to it without waiting for the simultaneous completion of the same iterate of the other block-components. Thus the previous iterate is no longer available to all processors. Instead, block-components of x are updated using a vector which is made up of block-components of different previous, not necessarily the latest, iterates.

More generally, in the algorithm to be analyzed in this section we assume that each new iterate block-component is computed by a processor of a parallel computer (e.g.

a processor becomes available and takes the next block-component from a queue of tasks). Before the beginning of the computation of a new iterate, the processor reads the most recent values of the other block-components. No further communication takes place until the computation of the iterate is completed. At that time, the computed value is available for the other processors to access. Whether the new information is written to a shared memory, broadcasted to the other processors, or kept in local memory, this is treated in the same way in the current analysis. We call this algorithm Outer Asynchronous Two-stage to emphasize that the same (outer) block-iterates are used for all inner iterations $j = 1, \cdots, s(i,k)$. This is similar to Model B in [4], and is in contrast to the totally asynchronous Algorithm 4.1 discussed in the next section.

As is customary in the description and analysis of asynchronous algorithms, the iteration subscript is increased every time any (block) component of the iteration vector is computed; see e.g. [1], [3], [4], [5], [6], [18], and the references given therein. We note that as a consequence of this convention, the number of iterations in asynchronous algorithms cannot be compared directly with the number of iterations in synchronous ones. In a formal way, the sets $J_k \subseteq \{1, 2, \cdots, q\}$, $k = 1, 2, \cdots$, are defined by $i \in J_k$ if the i-th block-component of the iteration vector is computed at the k-th step. The subscripts $r(\ell, k)$ are used to denote the iteration number of the ℓ-th block-component being used in the computation of any block-component in the k-th iteration, i.e. the iteration number of the ℓ-th block-component available at the beginning of the computation of $x_k^{(i)}$, if $i \in J_k$.

Each $n \times n$ matrix K partitioned as in (2) can be decomposed into q operators $K^{(i)} : \mathbb{R}^n \to \mathbb{R}^{n_i}$, $i = 1, \cdots, q$, so that $Ku = [K^{(1)}u, \cdots, K^{(q)}u]^{\mathrm{T}}$. With this notation, the asynchronous counterpart of Algorithm 2.1 may be written as follows.

Algorithm 3.1. (Outer Asynchronous Two-stage)
Given an initial vector $x_0^{\mathrm{T}} = [x_0^{(1)}, \cdots, x_0^{(q)}]$
for $k = 1, 2, \cdots$

$$x_k^{(i)} = \begin{cases} x_{k-1}^{(i)} & \text{if } i \notin J_k \\ T^{(i)}(k)\left(x_{r(1,k)}^{(1)}, \ldots, x_{r(q,k)}^{(q)}\right)^{\mathrm{T}} + Q^{(i)}(k)b & \text{if } i \in J_k. \end{cases} \tag{16}$$

with $Q(k)$ and $T(k)$ as defined in (5) and (8), respectively.

Let N be partitioned as in (2). For completeness, and for easy comparison with Algorithm 4.1, we rewrite (16) explicitly as

$$x_k^{(i)} = \begin{cases} x_{k-1}^{(i)} & \text{if } i \notin J_k \\ R_i^{s(i,k)} x_{r(i,k)}^{(i)} + \sum_{j=0}^{s(i,k)-1} R_i^j F_i^{-1}\left(\sum_{\ell=1}^{q} N_{i\ell} x_{r(\ell,k)}^{(\ell)} + b^{(i)}\right) & \text{if } i \in J_k. \end{cases} \tag{17}$$

Algorithm 3.1 is a special case of the totally asynchronous Algorithm 4.1 considered in Sect. 4, and it is more general than Model B of [4]. The convergence results in the present section (Theorems 3.3 and 3.4) can be regarded as corollaries of the corresponding Theorems 4.3 and 4.4. Nevertheless, the outer asynchronous model deserves a detailed treatment on its own due to its relevance in practice, cf. Model B of [4]. Also, we would like to emphasize that the proofs of this section do not require the type of more complex technique needed for those of Sect. 4.

We always assume that all our asynchronous iterations satisfy the following minimal restrictions:

(18) $$\begin{cases} r(\ell,k) < k \ \text{ for all } \ell = 1,\cdots,q, \quad k = 1,2,\cdots \\ \lim_{k\to\infty} r(\ell,k) = \infty \ \text{ for all } \ell = 1,\ldots,q. \\ \text{The set } \{k | i \in J_k\} \ \text{ is unbounded for all } i = 1,\ldots,q. \end{cases}$$

Conditions (18) are the classical conditions appearing in convergence results for asynchronous iterations. They are minimal in the sense that they are fulfilled in virtually every practical implementation of an asynchronous algorithm; see e.g. [1], [3], [6], [11].

We now formulate a basic convergence theorem for general asynchronous iterations.

Theorem 3.2. *Let $S(k)$ be a sequence of operators on $\mathbb{R}^n$ having a common fixed point x_*. Let $\|\cdot\|_i$ be a norm on $\mathbb{R}^{n_i}$, $i = 1,\ldots,q$. Let $a \in \mathbb{R}^q$, $a > 0$ and denote $\|\cdot\|_a$ the weighted max-norm on $\mathbb{R}^n$ given by*

$$\|x\|_a := \max_{i=1,\cdots,q} \{\frac{1}{a_i}\|x^{(i)}\|_i\}.$$

For all $k = 1,2,\ldots$, assume that there exists a constant $\alpha \in [0,1)$ such that

(19) $$\|S(k)x - x_*\|_a \le \alpha \|x - x_*\|_a \quad \text{for all } x \in \mathbb{R}^n.$$

Assume further that the sequence $r(i,k)$ and the sets J_k, $i = 1,\cdots,q$, $k = 1,2,\cdots$, satisfy conditions (18) Then the asynchronous iteration

(20) $$x_k^{(i)} = \begin{cases} x_{k-1}^{(i)} & \text{if } i \notin J_k \\ S^{(i)}(k)\left(x^{(1)}_{r(1,k)},\ldots,x^{(q)}_{r(q,k)}\right)^{\mathrm{T}} & \text{if } i \in J_k, \end{cases}$$

$k = 1,2,\cdots$, converges to x_* for any initial guess x_0.

Proof. This result is a simple extension of Theorem 3.4 of El Tarazi [6]. El Tarazi actually considers the case of a single operator, i.e. $S(k)$ does not change with each iteration k. Due to the uniformity assumption (19), the proof in [6] immediately carries over to the present situation. □

Similar theorems can be found in recent papers by Elsner, Koltracht, and Neumann [8, Theorem 2], [9, Theorem 2].

We use Theorem 3.2 to prove the following convergence results on outer asynchronous two-stage methods.

Theorem 3.3. *Let A such that $A^{-1} \ge O$ be partitioned as in (2). Let $A = M - N$ be a regular splitting with $M = \mathrm{diag}(M_i)$, and let $M_i = F_i - G_i$ be weak regular splittings, $i = 1,\cdots,q$. Assume that the sequence $r(i,k)$ and the sets J_k, $i = 1,\cdots,q$, $k = 1,2,\cdots$, satisfy conditions (18). Then, the outer asynchronous block two-stage Algorithm 3.1 converges to x_* with $Ax_* = b$ for any initial vector x_0 and for any sequence of numbers of inner iterations $s(i,k) \ge 1$, $i = 1,\cdots,q$, $k = 1,2,\cdots$*

Proof. Denote $S(k)x = T(k)x + Q(k)b$. Let $x_* = A^{-1}b$, then by (9) x_* is a common fixed point of all $S(k)$, $k = 1,2,\cdots$, and

(21) $$S(k)x - x_* = T(k)(x - x_*).$$

According to the proof of Theorem 2.2 we have $T(k) \geq 0$ for all $k = 1, 2, \cdots$, and by (13) there exists $v \in \mathbb{R}^n$, $v > 0$ and $\theta \in [0, 1)$ such that $T(k)v \leq \theta v$ for all $k = 1, 2, \cdots$ By Lemma 1.6 this implies that

$$\|T(k)y\|_v \leq \theta \|y\|_v \tag{22}$$

for all $y \in \mathbb{R}^n$. Obviously, $\|\cdot\|_v$ can be expressed in the form

$$\|y\|_v = \max_{i=1,\cdots,q} \|y^{(i)}\|_{v^{(i)}}.$$

(Each $\|\cdot\|_{v^{(i)}}$ denotes a weighted max-norm on $\mathbb{R}^{n_i}$). Thus, (21) and (22) imply (19) in the hypothesis of Theorem 3.2, which completes the proof. □

Theorem 3.3 is similar to Theorem 2.2 in [4], but our proof uses a different approach.

Theorem 3.4. *Let A be an H-matrix partitioned as in (2). Let $A = M - N$ be an H-splitting with $M = \text{diag}(M_i)$, and let $M_i = F_i - G_i$ be H-compatible splittings, $i = 1, \cdots, q$. Assume that the sequence $r(i, k)$ and the sets J_k, $i = 1, \cdots, q$, $k = 1, 2, \cdots$, satisfy conditions (18). Then, the outer asynchronous block two-stage Algorithm 3.1 converges to x_* with $Ax_* = b$ for any initial vector x_0 and for any sequence of numbers of inner iterations $s(i, k) \geq 1$, $i = 1, \cdots, q$, $k = 1, 2, \cdots$*

Proof. By (15) in the proof of Theorem 2.3 we have $|T(k)|v \leq \theta v$ for some $v \in \mathbb{R}^n$, $v > 0$ and $\theta \in [0, 1)$. By Lemma 1.6 this implies that the inequality (22) holds for all $y \in \mathbb{R}^n$, and thus the theorem follows exactly in the same manner as Theorem 3.3. □

4. Totally asynchronous two-stage methods

We consider in this section an asynchronous two-stage algorithm where, at each inner iteration, the most recent information from the other block-components is used. In other words, the (block) components $x^{(\ell)}_{r(\ell,k)}$ in (17) may differ for different values of j, $j = 0, \cdots, s(i, k) - 1$ ($i \in J_k$). To reflect this, we therefore use indices of the form $r(\ell, j, k)$.

The totally asynchronous two-stage method can then be expressed as

Algorithm 4.1. (Totally Asynchronous Two-stage)
Given an initial vector $x_0^{\mathrm{T}} = [x_0^{(1)}, \cdots, x_0^{(q)}]$
for $k = 1, 2, \cdots$

$$x_k^{(i)} = \begin{cases} x_{k-1}^{(i)} & \text{if } i \notin J_k \\ R_i^{s(i,k)} x^{(i)}_{r(i,0,k)} + \displaystyle\sum_{j=0}^{s(i,k)-1} R_i^j F_i^{-1} \left(\sum_{\ell=1}^{q} N_{i\ell} x^{(\ell)}_{r(\ell,j,k)} + b^{(i)} \right) & \text{if } i \in J_k. \end{cases} \tag{23}$$

Algorithm 4.1 requires more computational work than the outer asynchronous method 3.1, since $\sum_{\ell=1}^{q} N_{i\ell} x^{(\ell)}_{r(\ell,j,k)} + b^{(i)}$ has to be computed anew in each inner iteration. On the other hand, in Algorithm 4.1 processors use the latest information as soon as it is available so that, with a simple heuristic, we can expect it to require less iterations in order to achieve a given accuracy. Whether or not this pays off in a practical implementation also depends on what proportion of the total computation is spent in calculating the additional $\sum_{\ell=1}^{q} N_{i\ell} x^{(\ell)}_{r(\ell,j,k)} + b^{(i)}$. This in turn depends for example on the sparsity of N and A.

Analogous to (18) we now assume

$$(24)\quad \begin{cases} r(i,j,k) < k, \ \text{ for all } i = 1,\cdots,q, \ \ j = 0,\cdots,s(i,k)-1, \ \ k = 1,2,\cdots, \\ \lim\limits_{k\to\infty} \ \min\limits_{j=0,\cdots,s(i,k)-1} r(i,j,k) = \infty, \ \text{ for all } i = 1,\cdots,q, \\ \text{the set } \{k \mid i \in J_k\} \text{ is unbounded for all } i = 1,\cdots,q. \end{cases}$$

Since $s(i,k)$ can become arbitrarily large as k increases, we have a potential for arbitrary many iterates $x^{(\ell)}$ to appear in the last sum in the second row of (23). In order to prove a convergence theorem for the totally asynchronous two-stage-method we will use the following theorem.

Theorem 4.2. *Let $E = \prod_{\gamma\in\Gamma} E^{(\gamma)}$ be a topological product space with possibly infinitely many components where Γ is an appropriate index set. Let $\hat{H}(k)$, $k = 1,2,\cdots,$ be mappings from E to E and consider the asynchronous iteration*

$$(25)\quad \hat{x}^{(\gamma)}_k = \begin{cases} \hat{x}^{(\gamma)}_{k-1} & \textit{if } i \notin \hat{J}_k \\ \hat{H}^{(\gamma)}(k)\left(\prod\limits_{\mu\in\Gamma} \hat{x}^{(\mu)}_{\hat{r}(\mu,k)}\right) & \textit{if } i \in \hat{J}_k, \end{cases}$$

with

$$(26)\quad \begin{cases} \hat{r}(\mu,k) \le k-1 \textit{ for } \mu\in\Gamma,\ k = 1,2,\cdots, \\ \lim_{k\to\infty} \hat{r}(\mu,k) = \infty \ \textit{ uniformly in } \mu, \\ \textit{for all } k = 1,2,\cdots, \textit{ there exists } l(k) \ge k \textit{ such that } \ \Gamma = \hat{J}_k \cup\cdots\cup \hat{J}_{l(k)}. \end{cases}$$

Assume that there is a nonempty subset E_* of E and a sequence $\{E_k\}_{k=0}^{\infty}$ of nonempty subsets of E satisfying

(i) $E_k = \prod\limits_{\gamma\in\Gamma} E^{(\gamma)}_k$ *(with $E^{(\gamma)}_k \subseteq E^{(\gamma)}$, $\gamma\in\Gamma$), $k = 0,1,\cdots$,*
(ii) $\hat{H}(k)E_{k-1} \subseteq E_k \subseteq E_{k-1},\ k = 1,2,\cdots,$
(iii) any limit point of any sequence $\{\hat{z}_k\}_{k=0}^{\infty}$ with $\hat{z}_k \in E_k$ lies in E_*.

Then, provided $\hat{x}_0 \in E_0$, every limit point of the iterates $\hat{x}_k$ of the asynchronous iteration (25) lies in E_*. Particularly, if $E_* = \{\hat{x}_*\}$, the iteration (25) converges to $\hat{x}_*$.

Proof. In Theorem 2.1 of Frommer [11] this result is shown in the case where $H(k)$ does not depend on k. If $H(k)$ does depend on k, the proof is the same, so we do not reproduce it here. □

As opposed to the basic Theorem 3.2, the above theorem allows for spaces with infinitely many components. This is precisely what we need in the proof of the following theorem on the convergence of the totally asynchronous iteration (23).

Theorem 4.3. *Let A such that $A^{-1} \geq O$ be partitioned as in (2). Let $A = M - N$ be a regular splitting with $M = \mathrm{diag}(M_i)$, and let $M_i = F_i - G_i$ be weak regular splittings, $i = 1, \cdots, q$. Assume that the numbers $r(i,j,k)$ and the sets J_k, $i = 1, \cdots, q$, $k = 1, 2, \cdots$, satisfy conditions (24). Then, the totally asynchronous block two-stage Algorithm 4.1 converges to x_* with $Ax_* = b$ for any initial vector x_0 and for any sequence of numbers of inner iterations $s(i,k) \geq 1$, $i = 1, \cdots, q$, $k = 1, 2, \cdots$*

Proof. We will construct an operator $\hat{H}$ on $E = \prod_{j=0}^{\infty}(\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_q})$ such that the asynchronous iteration (23) can be interpreted in terms of an asynchronous iteration for $\hat{H}$, where $\hat{H}$ satisfies the assumptions of Theorem 4.2. To this purpose, let us use superscripts (j) and (i,j) to denote components in E, i.e. for $u \in E$ we write

$$u = \prod_{j=0}^{\infty} u^{(j)}, \text{ where } u^{(j)} = (u^{(1,j)}, \cdots, u^{(q,j)}) \text{ with } u^{(i,j)} \in \mathbb{R}^{n_i},$$
$$i = 1, \cdots, q,\ j = 0, 1, \cdots$$

Define the i-th block component of the operator $\bar{H}(k) : E \to \mathbb{R}^n$ by

$$\begin{aligned} \bar{H}^{(i)}(k)u &= R_i^{s(i,k)} u^{(i,0)} + \sum_{j=0}^{s(i,k)-1} R_i^j F_i^{-1} \left(\sum_{\ell=1}^{q} N_{i\ell} u^{(\ell,j)} + b^{(i)} \right) \\ &= \bar{T}^{(i)}(k)u + \sum_{j=0}^{s(i,k)-1} R_i^j F_i^{-1} b^{(i)} \end{aligned} \tag{27}$$

with $\bar{T}^{(i)}(k)u = R_i^{s(i,k)} u^{(i,0)} + \sum_{j=0}^{s(i,k)-1} R_i^j F_i^{-1} \left(\sum_{\ell=1}^{q} N_{i\ell} u^{(\ell,j)} \right)$, $i = 1, \cdots, q$. Moreover, let $\hat{H}(k)$, $\hat{T}(k) : E \to E$ be given by

$$\begin{aligned} \hat{H}^{(i,j)}(k)u &= \bar{H}^{(i)}(k)u, \quad i = 1, \cdots, q,\ j = 0, 1, \cdots, \\ \hat{T}^{(i,j)}(k)u &= \bar{T}^{(i)}(k)u, \quad i = 1, \cdots, q,\ j = 0, 1, \cdots, \end{aligned}$$

i.e. $\hat{H}(k)$ and $\hat{T}(k)$ are infinite replications of $\bar{H}$ and $\bar{T}$, respectively. Finally, let

$$\begin{aligned} \hat{J}_k &= \{(i,j) \mid i \in J_k,\ j = 0, 1, \cdots\}, \\ \hat{r}((i,j),k) &= \begin{cases} r(i,j,k) & \text{if } j < s(i,k) \\ k-1 & \text{otherwise.} \end{cases} \end{aligned}$$

Then, because of (24), $\hat{J}_k$ and $\hat{r}((i,j),k)$ satisfy (26). Moreover, if $\hat{x}_k$ denotes the k-th iterate of the asynchronous iteration (25), then $\hat{x}_k = \prod_{j=0}^{\infty} x_k$, where x_k denotes the iterate of the totally asynchronous iteration (23), provided $\hat{x}_0 = \prod_{j=0}^{\infty} x_0$. Thus, we only have to show that the asynchronous iteration (25) (with $\hat{H}$, $\hat{J}_k$, $\hat{r}(i,j,k)$ and $\hat{x}_0$ defined as above) converges to $\hat{x}_* = \prod_{j=0}^{\infty} x_*$, where $x_* = A^{-1}b$.

It follows from the definition of $\hat{H}$ in (27) that $\bar{H}(k)\hat{x}_* = x_*$, cf. (9). Therefore $\hat{H}(k)\hat{x}_* = \hat{x}_*$, and $\hat{H}(k)\hat{x} - \hat{x}_* = \hat{T}(k)(\hat{x} - \hat{x}_*)$, for all $\hat{x} \in E$ and $k = 1, 2, \cdots$. It also follows from (27) that

$$\hat{T}(k)x \geq 0 \text{ for all } x \in E \text{ with } x \geq 0, \tag{28}$$

where here '$\leq$' denotes the partial ordering on E induced by the partial ordering of its components $\mathbb{R}^n$.

Consider any fixed vector $e > 0$ (e.g. with all components equal to 1), and $v = A^{-1}e > 0$. Let $\hat{v} \in E$ be a multiple copy of v. Using exactly the same arguments as in Theorem 2.2 there exists $0 \leq \theta < 1$ such that

$$\hat{T}(k)\hat{v} \leq \theta\hat{v} \text{ for all } k = 1, 2, \cdots \tag{29}$$

Now, let $a > 0$ be sufficiently large such that $-a\hat{v} \leq \hat{x}_0 - \hat{x}_* \leq a\hat{v}$ and define

$$\begin{aligned} E_k &= \{\hat{x} \in E \,|\, -\theta^k a\hat{v} \leq \hat{x} - \hat{x}_* \leq \theta^k a\hat{v}\} \\ &= \prod_{j=0}^{\infty}\{x \in \mathbb{R}^n \,|\, \theta^k av \leq x - x_* \leq \theta^k av\}. \end{aligned}$$

From (28) and (29) it follows that

$$\hat{H}(k)E_{k-1} \subseteq E_k \subseteq E_{k-1},\ k = 1, 2, \cdots,$$

and the only limit point of any sequence $\{\hat{z}_k\}_{k=0}^{\infty}$ with $\hat{z}_k \in E_k$ is $\hat{x}_*$. So we have shown that all assumptions of Theorem 4.2 are met and this concludes our proof. □

Theorem 4.4. *Let A be an H-matrix partitioned as in (2). Let $A = M - N$ be an H-splitting with $M = \mathrm{diag}(M_i)$, and let $M_i = F_i - G_i$ be H-compatible splittings, $i = 1, \cdots, q$. Assume that the numbers $r(i, j, k)$ and the sets J_k, $i = 1, \cdots, q$, $k = 1, 2, \cdots$, satisfy conditions (24). Then, the outer asynchronous block two-stage algorithm 4.1 converges to x_* with $Ax_* = b$ for any initial vector x_0 and for any sequence of numbers of inner iterations $s(i, k) \geq 1$, $i = 1, \cdots, q$, $k = 1, 2, \cdots$*

Proof. This theorem follows from Theorem 4.3 in a similar manner as Theorem 3.4 follows from Theorem 3.3. □

References

1. Baudet, G.M. (1978): Asynchronous iterative methods for multiprocessors. J. ACM **25**, 226–244
2. Berman, A., Plemmons, R.J. (1994): Nonnegative Matrices in the Mathematical Sciences, 3rd edn. Academic Press, New York, 1979. Reprinted by SIAM, Philadelphia
3. Bertsekas, D.P., Tsitsiklis, J.N. (1989): Parallel and Distributed Computation. Prentice Hall, Englewood Cliffs, New Jersey
4. Bru, R., Elsner, L., Neumann, M. (1988): Models of parallel chaotic iteration methods. Linear Algebra Appl. **103**, 175–192
5. Chazan, D., Miranker, W. (1969): Chaotic relaxation. Linear Algebra Appl. **2**, 199–222
6. El Tarazi, M.N. (1982): Some convergence results for asynchronous algorithms. Numer. Math. **39**, 325–340
7. Elman, H.C., Golub, G.H.: Inexact and preconditioned Uzawa algorithms for saddle point problems. Technical Report CS-TR-3075 and UMIACS-TR-93-41, Department of Computer Science, University of Maryland, College Park, Md., May 1993. SIAM J. Numer. Anal. (to appear)
8. Elsner, L., Koltracht, I., Neumann, M. (1990): On the convergence of asynchronous paracontractions with application to tomographic reconstruction from incomplete data. Linear Algebra Appl. **130**, 65–82
9. Elsner, L., Koltracht, I., Neumann, M. (1992): Convergence of sequential and asynchronous nonlinear paracontractions. Numer. Math. **62**, 305–319
10. Elsner, L., Neumann, M., Vemmer, B. (1991): The effect of the number of processors on the convergence of the parallel block Jacobi method. Linear Algebra Appl. **154–156**, 311–330

11. Frommer, A. (1991): On asynchronous iterations in partially ordered spaces. Numer. Functional Anal. Optimization **12**, 315–325
12. Frommer, A. Szyld, D.B. (1992): H-splittings and two-stage iterative methods. Numer. Math. **63**, 345–356
13. Golub, G.H., Overton, M.L. (1982): Convergence of a two-stage Richardson iterative procedure for solving systems of linear equations. In: Watson, G.A., ed., Numerical Analysis (Proceedings of the Ninth Biennial Conference, Dundee, Scotland, 1981), Lecture Notes in Mathematics 912, pp. 128–139. Springer, New York
14. Golub, G.H., Overton, M.L. (1988): The convergence of inexact Chebyshev and Richardson iterative methods for solving linear systems. Numer. Math. **53**, 571–593
15. Householder, A.S. (1975): The Theory of Matrices in Numerical Analysis. Blaisdell, Waltham, Mass. 1964. Reprinted by Dover, New York
16. Lanzkron, P.J., Rose, D.J., Szyld, D.B. (1991): Convergence of nested classical iterative methods for linear systems. Numer. Math. **58**, 685–702
17. Mayer, G. (1987): Comparison theorems for iterative methods based on strong splittings. SIAM J. Numer. Anal. **24**, 215–227
18. Miellou, J.-C., Spiteri, P. (1985): Un critère de convergence pour des méthodes générales de point fixe. Math. Modelling Numer. Anal. **19**(4), 645–669
19. Neumaier, A. (1984): New techniques for the analysis of linear interval equations. Linear Algebra Appl. **58**, 273–325
20. Ortega, J.M., Rheinboldt, W.C. (1970): Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York and London
21. Ostrowski, A.M. (1937): Über die Determinanten mit überwiegender Hauptdiagonale. Coment. Math. Helv. **10**, 69–96
22. Ostrowski, A.M. (1956): Determinanten mit überwiegender Hauptdiagonale und die absolute Konvergenz von linearen Iterationsprozessen. Coment. Math. Helv. **30**, 175–210
23. Rheinboldt, W.C., Vandergraft, J.S. (1973): A simple approach to the Perron-Frobenius theory for positive operators on general partially-ordered finite-dimentional linear spaces. Math. Comput. **27**, 139–145
24. Robert, F., Charnay, M., Musy, F. (1975): Itérations chaotiques série-parallèle pour des équations non-linéaires de point fixe. Apl. Mat. **20**, 1–38
25. Varga, R.S. (1960): Factorization and normalized iterative methods. In: Langer, R.E., ed., Boundary Problems in Differential Equations, pp. 121–142. Madison, The University of Wisconsin Press
26. Varga, R.S. (1962): Matrix Iterative Analysis. Prentice-Hall, Englewood Cliffs, New Jersey
27. Varga, R.S. (1976): On recurring theorems on diagonal dominance. Linear Algebra Appl. **13**, 1–9
28. Young, D.M. (1971): Iterative Solution of Large Linear Systems. Academic Press, New York

This article was processed by the author using the LaTeX style file *pljour1* from Springer-Verlag.