Numer. Math. (1999) 82: 1–20

Numerische
Mathematik

Electronic Edition

Block and asynchronous two-stage methods for mildly nonlinear systems

Zhong-Zhi Bai[1,⋆], **Violeta Migallón**[2,⋆⋆], **José Penadés**[2,⋆⋆], **Daniel B. Szyld**[3,⋆⋆⋆]

[1] State Key Laboratory of Scientific/Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, P.O. Box 2719, Beijing 100080, P.R. China; e-mail: bzz@lsec.cc.ac.cn
[2] Departamento de Ciencia de la Computación e Inteligencia Artificial, Universidad de Alicante, E-03071 Alicante, Spain; e-mail: violeta@dtic.ua.es, jpenades@dtic.ua.es
[3] Department of Mathematics, Temple University, Philadelphia, Pennsylvania 19122-6094, USA; e-mail: szyld@math.temple.edu

Received June 5, 1997 / Revised version received December 29, 1997

Summary. Block parallel iterative methods for the solution of mildly nonlinear systems of equations of the form $Ax = \Phi(x)$ are studied. Two-stage methods, where the solution of each block is approximated by an inner iteration, are treated. Both synchronous and asynchronous versions are analyzed, and both pointwise and blockwise convergence theorems provided. The case where there are overlapping blocks is also considered. The analysis of the asynchronous method when applied to linear systems includes cases not treated before in the literature.

Mathematics Subject Classification (1991): 65H10, 65F10

1. Introduction

In this paper we consider the solution of the mildly nonlinear system

(1) $$Ax = \Phi(x),$$

⋆ Supported by the Chinese National Natural Science Foundation Project 19601036
⋆⋆ Supported by Spanish CICYT grant numbers TIC96-0718-C02-02 and PB96-1054-CV02-01
⋆⋆⋆ Supported by National Science Foundation grants INT-9521226 and DMS-9625865
Correspondence to: D.B. Szyld

where $A \in \mathbb{R}^{n\times n}$ is nonsingular and $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear function with certain local smoothness properties; see Sects. 3 and 4. This equation arises in many problems of science and engineering, and in particular in discretizations of certain nonlinear differential equations, e.g., of the form $\Delta u = \sigma(u)$; see e.g., [36].

We are interested in solution methods of (1) in which the matrix A is partitioned into L by L blocks $A_{ij} \in \mathbb{R}^{n_i\times n_j}$, with $\sum_{\ell=1}^{L} n_\ell = n$, i.e.,

$$\begin{aligned} &A \in \mathbb{L}_n(n_1, n_2, \ldots, n_L) \\ &\quad = \{A \in \mathbb{R}^{n\times n} \mid A = (A_{\ell k}), A_{\ell k} \in \mathbb{R}^{n_\ell \times n_k}, 1 \le \ell, k \le L\}, \end{aligned}$$

and $A_{\ell\ell}$ are nonsingular for $\ell = 1, \ldots, L$. When the context is clear we will simply use $\mathbb{L}_n$ for $\mathbb{L}_n(n_1, n_2, \ldots, n_L)$. This partition may correspond to a partition of the underlying grid, or of the domain of the differential equation being studied, or it may originate from a partitioning algorithm of the sparse matrix A, as done, e.g., in [6], [35]. In particular, we are interested in several parallel generalizations of the following block Jacobi algorithm. Let x_i be the vector at the ith iteration. Assume that the vectors x_i and $\Phi(x_i)$ are partitioned in subvectors $x_i^{(\ell)}, \Phi(x_i)^{(\ell)} \in \mathbb{R}^{n_\ell}$, $\ell = 1, \ldots, L$, in a way conformally with the partition of A, i.e.,

$$\begin{aligned} &x_i \in V_n(n_1, n_2, \ldots, n_L) \\ &\quad = \{x \in \mathbb{R}^n \mid x = (x^{(1)^{\mathrm{T}}}, \ldots, x^{(L)^{\mathrm{T}}})^{\mathrm{T}}, x^{(\ell)} \in \mathbb{R}^{n_\ell}, 1 \le \ell \le L\}. \end{aligned}$$

Algorithm 1 (Block Jacobi). Given an initial vector x_0,

For $i = 1, 2, \ldots,$ until convergence.

For $\ell = 1$ to L

(2) Solve $A_{\ell\ell} x_i^{(\ell)} = -\sum_{k\neq\ell,\ k=1}^{L} A_{\ell k} x_{i-1}^{(k)} + \Phi(x_{i-1})^{(\ell)}.$

The block methods considered in this paper include, in particular, those in which (2) is not solved exactly, but instead approximated using an (inner) iterative method. These are block two-stage methods, also called inner/outer iterations, and have been studied extensively for linear and nonlinear systems; see e.g., [12], [13], [14], [25], [32], and the references given therein. A point two-stage method for the solution of (1), i.e., when $L = 1$, was recently studied in [1]. Algorithm 1 as well as its two-stage generalizations are ideal for parallel processing, since up to L different processors can each solve or approximate one of the problems (2). These algorithms are synchronous in the sense that to begin the calculation of the ith iterate, each processor has to wait until all processors have completed their calculation of the $(i-1)$th iterate.

In this paper, we also study asynchronous block methods, i.e., methods in which each processor begins a new calculation without waiting for the others to complete their respective tasks. Asynchronous methods have the potential of converging much faster than synchronous methods, especially when there is load imbalance, e.g., when one of the systems (2) takes much longer to solve than all the others; see e.g., [13], [24], [27], [31]. A block asynchronous method for the solution of (1) was analyzed in [9] using a different approach, and without considering the two-stage case.

In the following section we present some definitions and preliminary results used in the paper. In Sect. 3, we present a general framework to study the block methods and prove their convergence, while in Sect. 4 we analyze the convergence of the asynchronous methods. Our results apply to a rather general class of nonsingular matrices, including block H-matrices; see e.g., [2], [38], [40]. We introduce a very general computational model for these asynchronous iterations. Thus, our convergence proofs include a large class of methods, including those with overlap.

2. Preliminaries

Given a vector $x \in \mathbb{R}^n$, we say that it is nonnegative (positive), denoted $x \geq 0$ ($x > 0$), if all components of x are nonnegative (positive). Similarly, if $x, y \in \mathbb{R}^n$, $x \geq y$ ($x > y$) means that $x - y \geq 0$ ($x - y > 0$). For a vector $x \in \mathbb{R}^n$, $|x|$ denotes the vector whose components are the absolute values of the corresponding components of x. These definitions carry over immediately to matrices.

Let $x > 0$, we consider the vector norm

$$\|y\|_x = \max_{1 \leq j \leq n} \left| \frac{1}{x_j} y_j \right| . \tag{3}$$

This vector norm is monotonic and for every matrix $B \in \mathbb{R}^{n \times n}$ it satisfies $\| \, |B|x\|_x = \|B\|_x$, where $\|B\|_x$ denotes the matrix norm of B induced by the vector norm defined in (3); see e.g., [30]. A nonsingular matrix A is called M-matrix if it has non-positive off-diagonal entries and it is monotone (i.e., $A^{-1} \geq O$); see e.g., [7], [42]. By $\rho(A)$ we denote the spectral radius of the square matrix A.

We define the following subset of $\mathbb{L}_n$ used in the analysis of iterative methods for block H-matrices; see, e.g., [3], [4], [17], [40]. Again, we do not write the parameters $(n_1, n_2, \ldots, n_L)$, when they are clear from the context.

$\mathbb{L}_{n,\mathrm{I}}(n_1, n_2, \ldots, n_L)$

$$= \{A = (A_{\ell k}) \in \mathbb{L}_n \mid \quad A_{\ell\ell} \in \mathbb{R}^{n_\ell \times n_\ell} \text{ nonsingular}, \ \ell = 1, \ldots, L\},$$

For a matrix $A \in \mathbb{L}_n$, let $D(A) = \mathrm{Diag}(A_{11}, A_{22}, \ldots, A_{LL})$, i.e., its block-diagonal part. Thus, $A \in \mathbb{L}_{n,\mathrm{I}}$ if and only if $A \in \mathbb{L}_n$ and $D(A)$ is nonsingular.

For any matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$, we define its comparison matrix $< A > = (\alpha_{ij})$ by $\alpha_{ii} = |a_{ii}|$, $\alpha_{ij} = -|a_{ij}|$, $i \neq j$. Similarly, for $A \in \mathbb{L}_{n,\mathrm{I}}$ we define its type-I and type-II comparison matrices $\langle A \rangle = (\langle A \rangle_{ij}) \in \mathbb{R}^{L \times L}$ and $\langle\langle A \rangle\rangle = (\langle\langle A \rangle\rangle_{ij}) \in \mathbb{R}^{L \times L}$ as $\langle A \rangle_{ii} = \|A_{ii}^{-1}\|^{-1}$, $\langle A \rangle_{ij} = -\|A_{ij}\|$, $i \neq j$, and $\langle\langle A \rangle\rangle_{ii} = 1$, $\langle\langle A \rangle\rangle_{ij} = -\|A_{ii}^{-1} A_{ij}\|$, $i \neq j$, $i, j = 1, 2, \ldots, L$, respectively; see [2], [18], [38]. We also define, for $A \in \mathbb{L}_n$, the block absolute value $[A] = (\|A_{ij}\|) \in \mathbb{R}^{L \times L}$. The definition for a vector $v \in V_n = V_n(n_1, n_2, \ldots, n_L)$ is analogous. Here $\| \cdot \|$ is any consistent matrix norm satisfying $\|I\| = 1$. This block absolute value has the following properties.

Lemma 2.1. [2] *Let $A, B \in \mathbb{L}_n$, $x, y \in V_n$ and $\gamma \in \mathbb{R}$. Then,*
(a) $|[A] - [B]| \leq [A + B] \leq [A] + [B], \ |[x] - [y]| \leq [x + y] \leq [x] + [y]$,
(b) $[AB] \leq [A][B], \ [Ax] \leq [A][x]$, *and*
(c) $[\gamma A] \leq |\gamma|[A], \ [\gamma x] \leq |\gamma|[x]$.

Following [37], A is said to be an H-matrix if $< A >$ is an M-matrix. We say that $A \in \mathbb{L}_{n,\mathrm{I}}$ is a Type-I (Type-II) block H-matrix if $\langle A \rangle$ ($\langle\langle A \rangle\rangle$) is an M-matrix. We denote this by $A \in H_B^{\mathrm{I}}$ ($A \in H_B^{\mathrm{II}}$). It follows that $H_B^{\mathrm{I}} \subset H_B^{\mathrm{II}}$ with the inclusion being strict.

Lemma 2.2. *(a) If $A \in \mathbb{R}^{n \times n}$ is an H-matrix, then $|A^{-1}| \leq < A >^{-1}$* [33], [37].
(b) If $A \in H_B^{\mathrm{I}} \subset \mathbb{L}_{n,\mathrm{I}}$, then $[A^{-1}] \leq \langle A \rangle^{-1}$ [2].
(c) If $A \in H_B^{\mathrm{II}} \subset \mathbb{L}_{n,\mathrm{I}}$, then $[A^{-1}] \leq \langle\langle A \rangle\rangle^{-1}[D(A)^{-1}]$ [2].

Definition 2.3. 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 $M^{-1} \geq O$ and $N \geq O$ [42],
(b) weak regular if $M^{-1} \geq O$ and $M^{-1}N \geq O$ [7], [36],
(c) H-splitting if $< M > - |N|$ is an M-matrix [25],
(d) H-compatible splitting if $< A > = < M > - |N|$ [25],
(e) H_B^{I}-compatible splitting if $\langle A \rangle = \langle M \rangle - [N]$, and
(f) H_B^{II}-compatible splitting if $\langle\langle A \rangle\rangle = \langle\langle M \rangle\rangle - [D(M)^{-1}N]$.

Lemma 2.4. *Let $A = M - N$ be a splitting.*
(a) If the splitting is an H-splitting, then A and M are H-matrices and $\rho(M^{-1}N) \leq \rho(< M >^{-1}|N|) < 1$.
(b) If the splitting is H-compatible and A is an H-matrix, then it is an H-splitting and thus convergent.
(c) If the splitting is H_B^{I}-compatible, then both A and $M \in H_B^{\mathrm{I}}$.
(d) If the splitting is H_B^{II}-compatible, then both A and $M \in H_B^{\mathrm{II}}$.

Proof. Parts (a) and (b) are shown in [25]. Parts (c) and (d) follow from the definitions and some simple bounds; see [2]. □

Lemma 2.5. [41] *Let $H_1, H_2, \ldots, H_i, \ldots$ 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*

$$H_i v \le \theta v,\ i = 1, 2, \ldots,$$

then $\rho(K_j) \le \theta^j < 1$, where $K_j = H_j \cdots H_2 \cdot H_1$, and therefore $\lim_{j\to\infty} K_j = O$.

Definition 2.6. A mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is called P-bounded (block P-bounded) if there exists a nonnegative matrix $P \in \mathbb{R}^{n\times n}$ ($P \in \mathbb{R}^{L\times L}$) such that

$$|\Phi(x) - \Phi(y)| \le P|x - y| \ \ ([\Phi(x) - \Phi(y)] \le P[x - y]),\ \text{for all } x, y \in \mathbb{R}^n.$$

Furthermore, if $\rho(P) < 1$, Φ is said to be a (block) P-contracting mapping.

Lemma 2.7. *Let $A \in \mathbb{R}^{n\times n}$ be nonsingular. Then the mildly nonlinear system (1) has a unique solution provided that anyone of the following conditions hold.*
(a) A is a monotone matrix, Φ is P-bounded, and $\rho(A^{-1}P) < 1$.
(b) A is an H-matrix, Φ is P-bounded, and $\rho(< A >^{-1}P) < 1$.
(c) $A \in H_B^{\mathrm{I}}$, Φ is block P-bounded, and $\rho(\langle A\rangle^{-1}P) < 1$.
(d) $A \in H_B^{\mathrm{II}}$, Φ is block P-bounded, and $\rho(\langle\langle A\rangle\rangle^{-1}[D(A)^{-1}]P) < 1$.

Proof. Parts (a) and (b) can be found in [1]. Parts (c) and (d) are shown in a similar way. □

3. Block methods

We present a general framework which includes, as a particular case, the block two-stage method described in Sect. 1. To that end, consider the (outer and inner) splittings $A = B_\ell - C_\ell$, $B_\ell = M_\ell - N_\ell$, $\ell = 1, \ldots, L$, and a set of diagonal nonnegative matrices E_ℓ, such that

$$\sum_{\ell=1}^{L} E_\ell = I. \tag{4}$$

The sequence $q(\ell, i)$ indicates, e.g., the number of inner iterations an iterative method uses to approximate the solution of the ℓth system (2), at the ith iteration.

Algorithm 2 (Nonlinear Two-stage Multisplitting). Given an initial vector x_0, and a sequence of numbers of inner iterations $q(\ell, i)$, $\ell = 1, \ldots,$ L, $i = 1, 2, \ldots$

For $i = 1, 2, \ldots,$ until convergence.

For $\ell = 1$ to L

$y_{\ell,0} = x_{i-1}$

For $j = 1$ to $q(\ell, i)$

$$M_\ell y_{\ell,j} = N_\ell y_{\ell,j-1} + C_\ell x_{i-1} + \Phi(x_{i-1}) \tag{5}$$

$$x_i = \sum_{\ell=1}^{L} E_\ell y_{\ell,q(\ell,i)} \,. \tag{6}$$

The concept of multisplittings, first introduced in [34], provides a very general setting to study parallel block methods; see, e.g., [2], [3], [4], [13], [17], [19], [21], [28], [31], [39]. This general setting encompasses cases, e.g., where there is overlap, i.e., where more than one processor computes approximations to the same variable, and the *weighting matrices* E_ℓ have positive entries smaller than 1, see e.g., [22], [29]. The two-stage generalization of Algorithm 1 can be recovered from Algorithm 2 by the appropriate choice of diagonal matrices E_ℓ, and by choosing, e.g., B_ℓ to be block diagonal $D(A)$, cf. [13], [24], and the comments after Theorem 3.1. We emphasize that only the components of $y_{\ell,j}$ for which the diagonal matrix E_ℓ is nonzero is used in (6). Thus, (5) needs to be interpreted more as a representation of the work in the ℓth processor, usually involving of the order of n_ℓ variables, than as a global operation involving all n variables.

In order to analyze the convergence properties of Algorithm 2, we can write the ith iteration vector as follows, cf. [13],

$$\begin{aligned} x_i = \sum_{\ell=1}^{L} E_\ell \Bigg(& (M_\ell^{-1} N_\ell)^{q(\ell,i)} x_{i-1} \\ & + \sum_{j=0}^{q(\ell,i)-1} (M_\ell^{-1} N_\ell)^j M_\ell^{-1} (C_\ell x_{i-1} + \Phi(x_{i-1})) \Bigg) \,, \end{aligned} \tag{7}$$

or equivalently

$$\begin{aligned} x_i = \sum_{\ell=1}^{L} E_\ell \Big(& (M_\ell^{-1} N_\ell)^{q(\ell,i)} x_{i-1} \\ & + \Big(I - (M_\ell^{-1} N_\ell)^{q(\ell,i)} \Big) B_\ell^{-1} (C_\ell x_{i-1} + \Phi(x_{i-1})) \Big) \,. \end{aligned} \tag{8}$$

The next theorem is our first local convergence result: if the splittings satisfy certain minimum convergence properties, and the initial guess is close to the solution, we have convergence, provided that as the iterations continue, the system is better approximated. This result is similar in spirit to that of [14].

Theorem 3.1. *Let $A \in \mathbb{R}^{n\times n}$ be a nonsingular matrix. Let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable in an open neighborhood of a solution $x_\star$ of the mildly nonlinear system* (1). *Consider that the splittings $A = B_\ell - C_\ell$, $\ell = 1,\ldots,L$, satisfy $\|B_\ell^{-1}(C_\ell + \Phi'(x_\star))\|_\infty < 1$, $\ell = 1,\ldots,L$. Assume further that the splittings $B_\ell = M_\ell - N_\ell$, $\ell = 1,\ldots,L$, are convergent. Then, if $\lim_{i\to\infty} q(\ell,i) = \infty$, $\ell = 1,\ldots,L$, there exists a $\delta > 0$ such that for every initial vector $x_0 \in S_\delta(x_\star) = \{x \in \mathbb{R}^n \mid \|x - x_\star\|_\infty < \delta\}$ the sequence of vectors generated by Algorithm 2 converges to $x_\star$.*

Proof. Since $\rho(M_\ell^{-1}N_\ell) < 1$ and $\lim_{i\to\infty} q(\ell,i) = \infty$, $\ell = 1,\ldots,L$, given any $\epsilon > 0$, there exists an integer i_0 such that $\|(M_\ell^{-1}N_\ell)^{q(\ell,i)}\|_\infty \leq \epsilon$, for all $i \geq i_0$, $\ell = 1,\ldots,L$. Let β be a positive constant satisfying

$$\|B_\ell^{-1}(C_\ell + \Phi'(x_\star))\|_\infty \leq \beta < 1,\ \ell = 1,\ldots,L.$$

Let us consider $\varepsilon_i = x_i - x_\star$ the error vector at the ith iteration. Since $x_\star$ is a fixed point of (7), or equivalently (8), we obtain after some algebraic manipulations that, for $i = 1,2,\ldots,$

$$\begin{aligned}\varepsilon_i = \sum_{\ell=1}^{L} E_\ell \Big(& (M_\ell^{-1}N_\ell)^{q(\ell,i)}\varepsilon_{i-1} \\ & + (I - (M_\ell^{-1}N_\ell)^{q(\ell,i)})(J_\ell(x_\star)\varepsilon_{i-1} + B_\ell^{-1}y(x_\star, x_{i-1}))\Big),\end{aligned}$$

for $i = 1,2,\ldots,$ where $J_\ell(x_\star) = B_\ell^{-1}(C_\ell + \Phi'(x_\star))$, $\ell = 1,\ldots,L$, and

$$y(x_\star, x) = \Phi(x) - \Phi(x_\star) - \Phi'(x_\star)(x - x_\star).$$

Then,

$$\begin{aligned}\|\varepsilon_i\|_\infty \leq \max_{1\leq\ell\leq L} \{ & \|(M_\ell^{-1}N_\ell)^{q(\ell,i)}\|_\infty \|\varepsilon_{i-1}\|_\infty \\ & + (1 + \|(M_\ell^{-1}N_\ell)^{q(\ell,i)}\|_\infty)(\|J_\ell(x_\star)\|_\infty \|\varepsilon_{i-1}\|_\infty \\ & + \|B_\ell^{-1}\|_\infty \|y(x_\star, x_{i-1})\|_\infty)\}, \quad i = 1,2,\ldots. \end{aligned} \tag{9}$$

On the other hand, by the hypotheses we can assume, without loss of generality, that Φ is continuously differentiable on the convex set $S_\epsilon(x_\star) = \{x \in \mathbb{R}^n \mid \|x - x_\star\|_\infty < \epsilon\}$. Hence (see e.g., [36, Exercise 3.2.6]), there exists $\alpha > 0$ such that

$$\|y(x_\star, x)\|_\infty \leq \alpha \|x - x_\star\|_\infty^2 \text{ for all } x \in S_\epsilon(x_\star). \tag{10}$$

From (9), we can consider $0 < \delta < \epsilon$ small enough such that, if $x_0 \in S_\delta(x_\star)$ then

$$\|x_i - x_\star\|_\infty < \epsilon \text{ for all } i < i_0. \tag{11}$$

Furthermore, from (9) again, if $i \geq i_0$ we have

$$\begin{aligned} \|\varepsilon_i\|_\infty &\leq \max_{1\leq\ell\leq L}\{(\epsilon + (1+\epsilon)\beta)\|\varepsilon_{i-1}\|_\infty \\ &\quad + (1+\epsilon)\|B_\ell^{-1}\|_\infty \|y(x_\star, x_{i-1})\|_\infty\} \\ &= \varrho(\epsilon)\|\varepsilon_{i-1}\|_\infty + \gamma_{\ell_0}(\epsilon)\|y(x_\star, x_{i-1})\|_\infty, \end{aligned} \tag{12}$$

where $\varrho(\epsilon) = \epsilon + (1+\epsilon)\beta$ and $\gamma_{\ell_0}(\epsilon) = (1+\epsilon)\max\limits_{1\leq\ell\leq L}\{\|B_\ell^{-1}\|_\infty\}$. Then, for $i = i_0$, using (10), (11) and (12) we obtain

$$\begin{aligned} \|\varepsilon_{i_0}\|_\infty &\leq \varrho(\epsilon)\|\varepsilon_{i_0-1}\|_\infty + \gamma_{\ell_0}(\epsilon)\alpha\|\varepsilon_{i_0-1}\|_\infty^2 \\ &< \varrho(\epsilon)\|\varepsilon_{i_0-1}\|_\infty + \gamma_{\ell_0}(\epsilon)\alpha\epsilon\|\varepsilon_{i_0-1}\|_\infty = \varphi_{\ell_0}(\epsilon)\|\varepsilon_{i_0-1}\|_\infty, \end{aligned}$$

where $\varphi_{\ell_0}(\epsilon) = \varrho(\epsilon) + \gamma_{\ell_0}(\epsilon)\alpha\epsilon$. Without loss of generality we can consider ϵ small enough such that $\varphi_{\ell_0}(\epsilon) < 1$. Then, by induction we easily obtain

$$\|\varepsilon_i\|_\infty < \varphi_{\ell_0}(\epsilon)\|\varepsilon_{i-1}\|_\infty \text{ with } \varphi_{\ell_0}(\epsilon) < 1, \text{ for all } i \geq i_0,$$

and then the proof is complete. □

Several comments on alternative hypotheses for Theorem 3.1 are in order. First, the infinite norm can be replaced by any weighted max-norm associated with a positive vector; see (3) and Sect. 4. Second, when all the outer splittings in Algorithm 2 are the same, i.e., $B_\ell = B,\ C_\ell = C$, $\ell = 1, \ldots, L$, the assumption $\|B^{-1}(C + \Phi'(x_\star))\|_\infty < 1$ can be replaced by the more general $\rho(B^{-1}(C + \Phi'(x_\star))) < 1$. This applies, in particular, for the two-stage version of Algorithm 1. Third, the fact that Φ is continuously differentiable in an open neighborhood of $x_\star$ can be relaxed to the following set of assumptions. The map Φ is Lipschitz continuous, Φ is B-differentiable (i.e., for every $z \in \mathbb{R}^n$ there exists a positively homogeneous function $B\Phi(z) : \mathbb{R}^n \to \mathbb{R}^n$ called the B-derivative of Φ at z such that $\lim\limits_{v\to 0} \dfrac{\Phi(z+v) - \Phi(z) - B\Phi(z)(v)}{\|v\|} = 0$), and the associated function $B\Phi$ is Lipschitz continuous at $x_\star$; see e.g., [1], [36]. We note that B-derivatives are similar to directional derivatives. These hypotheses were used in [1] for the proof of convergence of the point two-stage method with a fixed number of inner iterations.

In the next results, we prove the convergence of Algorithm 2 for *any* number of inner iterations in a few general cases, namely, when the matrix A is monotone, or it is an H-matrix or block H-matrix of different types. We impose further conditions on the outer and inner splittings.

Theorem 3.2. *Let $A \in \mathbb{R}^{n\times n}$ be a monotone matrix. Let the splittings $A = B_\ell - C_\ell$, $\ell = 1,\dots,L$, be regular and the splittings $B_\ell = M_\ell - N_\ell$, $\ell = 1,\dots,L$, be weak regular. Assume further that $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is a P-bounded mapping such that $\rho(A^{-1}P) < 1$. Then, the nonlinear two-stage multisplitting Algorithm 2 converges to the unique solution of the mildly nonlinear system* (1)*, for any initial vector x_0 and any sequence of numbers of inner iterations $q(\ell,i) \geq 1$, $\ell = 1,\dots,L$, $i = 1,2,\dots$.*

Proof. From Lemma 2.7 it follows that there exists a unique $x_\star \in \mathbb{R}^n$ such that $Ax_\star = \Phi(x_\star)$. Let $\varepsilon_i = x_i - x_\star$ be the error vector at the ith iteration of Algorithm 2. Since $x_\star$ is a fixed point of (7), it follows that for $i = 1,2,\dots,$

$$\varepsilon_i = \sum_{\ell=1}^{L} E_\ell\Bigg((M_\ell^{-1}N_\ell)^{q(\ell,i)}\varepsilon_{i-1} + \sum_{j=0}^{q(\ell,i)-1}(M_\ell^{-1}N_\ell)^j M_\ell^{-1}(C_\ell\varepsilon_{i-1} + \Phi(x_{i-1}) - \Phi(x_\star))\Bigg). \tag{13}$$

Then, using the inequalities $M_\ell^{-1}N_\ell \geq O$, $\ell = 1,\dots,L$, and

$$M_\ell^{-1}C_\ell \geq O, \quad \ell = 1,\dots,L, \tag{14}$$

and the fact that Φ is P-bounded, we obtain the following bound:

$$|\varepsilon_i| \leq \sum_{\ell=1}^{L} E_\ell T_i^{(\ell)}|\varepsilon_{i-1}|, \quad i = 1,2,\dots,$$

where

$$T_i^{(\ell)} = (M_\ell^{-1}N_\ell)^{q(\ell,i)} + \sum_{j=0}^{q(\ell,i)-1}(M_\ell^{-1}N_\ell)^j M_\ell^{-1}(C_\ell + P) \geq O. \tag{15}$$

On the other hand, some algebraic manipulations yields the equality

$$\begin{aligned}(M_\ell^{-1}N_\ell)^{q(\ell,i)} &+ \sum_{j=0}^{q(\ell,i)-1}(M_\ell^{-1}N_\ell)^j M_\ell^{-1}C_\ell \\ &= I - \sum_{j=0}^{q(\ell,i)-1}(M_\ell^{-1}N_\ell)^j M_\ell^{-1}A,\end{aligned}$$

and we rewrite (15) as

$$T_i^{(\ell)} = I - \sum_{j=0}^{q(\ell,i)-1}(M_\ell^{-1}N_\ell)^j M_\ell^{-1}(A - P). \tag{16}$$

Moreover, since the matrices A^{-1} and P are nonnegative, and $\rho(A^{-1}P) < 1$, the matrix $A - P = A(I - A^{-1}P)$ is monotone.
Consider any fixed vector $e > 0$ (e.g., $e = (1, 1, \ldots, 1)^{\mathrm{T}}$) and $v = (A - P)^{-1}e$. Since $(A - P)^{-1} \geq O$ and no row of $(A - P)^{-1}$ can have all null entries, we get $v > 0$. By the same arguments, $M_\ell^{-1}e > 0$, $\ell = 1, \ldots, L$. Then, we have from (16) that

$$T_i^{(\ell)}v = v - \sum_{j=0}^{q(\ell,i)-1} (M_\ell^{-1}N_\ell)^j M_\ell^{-1}e$$
$$= v - M_\ell^{-1}e - \sum_{j=1}^{q(\ell,i)-1} (M_\ell^{-1}N_\ell)^j M_\ell^{-1}e.$$

Because $e > 0$ and the matrices $M_\ell^{-1}N_\ell$ and M_ℓ^{-1} are nonnegative, it follows that $T_i^{(\ell)}v \leq v - M_\ell^{-1}e$. Moreover, since $T_i^{(\ell)}v \geq 0$ and $v - M_\ell^{-1}e < v$, there exists constants $0 \leq \theta_\ell < 1$, $\ell = 1, \ldots, L$, such that $T_i^{(\ell)}v \leq \theta_\ell v$, $\ell = 1, \ldots, L$, $i = 1, 2, \ldots$. Hence, setting $\theta = \max_{1\leq\ell\leq L}\{\theta_\ell\}$, we get $\sum_{\ell=1}^{L} E_\ell T_i^{(\ell)}v \leq \theta v$. By Lemma 2.5, this implies that the sequence of error vectors tends to zero, and then the proof is complete. □

As it was pointed out in [23], the hypothesis on the outer splittings in Theorem 3.2 can be relaxed to require $A = B_\ell - C_\ell$ to also be weak regular and, in addition, $M_\ell^{-1}C_\ell \geq O$; see (14).

Theorem 3.3. *Let $A \in \mathbb{R}^{n\times n}$ be an H-matrix. Let the splittings $A = B_\ell - C_\ell$, and $B_\ell = M_\ell - N_\ell$, $\ell = 1, \ldots, L$, be H-compatible. Assume further that $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is a P-bounded mapping such that $\rho(< A >^{-1}P) < 1$. Then, the nonlinear two-stage multisplitting Algorithm 2 converges to the unique solution of the mildly nonlinear system* (1)*, for any initial vector x_0 and any sequence of numbers of inner iterations $q(\ell, i) \geq 1$, $\ell = 1, \ldots, L$, $i = 1, 2, \ldots$.*

Proof. By Lemma 2.4, the matrices M_ℓ, $\ell = 1, \ldots, L$, are H-matrices. Therefore, using Lemma 2.2, we have the inequalities $|M_\ell^{-1}| \leq < M_\ell >^{-1}$. Thus, if $\varepsilon_i = x_i - x_\star$ is the error at the ith iteration of Algorithm 2, it follows from (13) that

$$|\varepsilon_i| \leq \sum_{\ell=1}^{L} E_\ell\Bigg(|M_\ell^{-1}N_\ell|^{q(\ell,i)}|\varepsilon_{i-1}| + \sum_{j=0}^{q(\ell,i)-1} |M_\ell^{-1}N_\ell|^j|M_\ell^{-1}|(|C_\ell||\varepsilon_{i-1}| + |\Phi(x_{i-1}) - \Phi(x_\star)|)\Bigg)$$

$$\leq \sum_{\ell=1}^{L} E_\ell \tilde{T}_i^{(\ell)} |\varepsilon_{i-1}|, \quad i = 1, 2, \ldots,$$

where

$$\tilde{T}_i^{(\ell)} = (< M_\ell >^{-1} |N_\ell|)^{q(\ell,i)}$$

$$+ \sum_{j=0}^{q(\ell,i)-1} (< M_\ell >^{-1} |N_\ell|)^j < M_\ell >^{-1} (|C_\ell| + P) \; . \tag{17}$$

The matrices $\sum_{\ell=1}^{L} E_\ell \tilde{T}_i^{(\ell)}$ can be considered as the matrices used in the proof of Theorem 3.2, to bound the error vector $\tilde{\varepsilon}_i$ (i.e., $|\tilde{\varepsilon}_i| \leq \sum_{\ell=1}^{L} E_\ell \tilde{T}_i^{(\ell)} |\tilde{\varepsilon}_{i-1}|$) of an iterative process corresponding to Algorithm 2 to solve the mildly nonlinear system $< A > x = \Phi(x)$, with outer splittings $< A > = < B_\ell > - |C_\ell|$ and inner splittings $< B_\ell > = < M_\ell > - |N_\ell|$, $\ell = 1, \ldots, L$. That system and these splittings satisfy the hypotheses of Theorem 3.2, and therefore we can affirm that there exists a positive vector $\tilde{v}$ and a constant $0 \leq \theta < 1$ such that $\sum_{\ell=1}^{L} E_\ell \tilde{T}_i^{(\ell)} \tilde{v} \leq \theta \tilde{v}$. Hence, by Lemma 2.5 the proof is complete.

□

In the case that all the outer splittings in Algorithm 2 are the same, i.e., $B_\ell = B$, $C_\ell = C$, $\ell = 1, \ldots, L$, e.g., for the two-stage version of Algorithm 1, the assumption in Theorem 3.3 that the outer splittings be H-compatible can be replaced by the less restrictive hypothesis of being an H-splitting.

The next convergence result corresponds to the Type-I and Type-II block H-matrices. Its proof is similar to those in Theorems 3.2 and 3.3, but note that many of the inequalities are in $\mathbb{R}^L$ and not in $\mathbb{R}^n$.

Theorem 3.4. *Let* $A \in H_B^{\mathrm{I}}$ $(H_B^{\mathrm{II}}) \subset \mathbb{L}_{n,\mathrm{I}}(n_1, n_2, \ldots, n_L)$. *Let the splittings* $A = B_\ell - C_\ell$ *and* $B_\ell = M_\ell - N_\ell$, $\ell = 1, 2, \ldots, L$, *be* H_B^{I}*-compatible (*H_B^{II}*-compatible and such that* $D(M_\ell) = D(B_\ell) = D(A)$*), and the weighting matrices* E_ℓ, $\ell = 1, 2, \ldots, L$, *satisfying* (4)*, satisfy in addition* $\sum_{\ell=1}^{L} [E_\ell] \leq I$*, the* $L \times L$ *identity matrix. Assume further that* $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ *is a block* P*-bounded mapping such that* $\rho(\langle A \rangle^{-1} P) < 1$ *(*$\rho(\langle\langle A \rangle\rangle^{-1} [D(A)^{-1}] P) < 1$*). Then, the nonlinear two-stage multisplitting Algorithm 2 converges to the unique solution of the mildly nonlinear system*

(1), *for any initial vector* x_0 *and any sequence of numbers of inner iterations* $q(\ell,i) \geq 1$, $\ell = 1,\ldots,L$, $i = 1,2,\ldots$.

Proof. From Lemma 2.7 it follows that (both in the Type-I and Type-II cases) there exists a unique vector $x_\star \in \mathbb{R}^n$ such that $Ax_\star = \Phi(x_\star)$. Thus, if we let $\varepsilon_i = x_i - x_\star$ be the error at the ith iteration of Algorithm 2, then (13) holds.

For the Type-I case, we note that the H_B^{I}-compatibility of the splittings imply that $M_\ell \in H_B^{\mathrm{I}}$, and thus by Lemma 2.2 (b), we have that $[M_\ell^{-1}] \leq \langle M_\ell \rangle^{-1}$, $\ell = 1,2,\ldots,L$. Using this last inequality, by taking block absolute values on both sides of (13), applying Lemma 2.1 and the block P-bounded property of Φ, we obtain the inequality

$$[\varepsilon_i] \leq \sum_{\ell=1}^{L} [E_\ell]\, \tilde{T}_{i,I}^{(\ell)}\, [\varepsilon_{i-1}], \quad i = 1,2,\ldots, \quad \text{where} \tag{18}$$

$$\begin{aligned} \tilde{T}_{i,I}^{(\ell)} &= (\langle M_\ell \rangle^{-1}[N_\ell])^{q(\ell,i)} \\ &\quad + \sum_{j=0}^{q(\ell,i)-1} (\langle M_\ell \rangle^{-1}[N_\ell])^j \langle M_\ell \rangle^{-1}([C_\ell] + P) \end{aligned} \tag{19}$$

is a nonnegative matrix in $\mathbb{R}^{L\times L}$. After some manipulations we obtain the following identity, cf. (16).

$$\tilde{T}_{i,I}^{(\ell)} = I - \sum_{j=0}^{q(\ell,i)-1} (\langle M_\ell \rangle^{-1}[N_\ell])^j \langle M_\ell \rangle^{-1}(\langle A \rangle - P).$$

Furthermore, since $\langle A \rangle^{-1}$ and P are nonnegative matrices and by the hypothesis $\rho(\langle A \rangle^{-1}P) < 1$, the matrix $(\langle A \rangle - P)$ is a monotone matrix. Now, in a way similar to the proof of Theorem 3.2, we can deduce that there exist a positive vector $v \in \mathbb{R}^L$ and a nonnegative constant $\theta \in [0,1)$ such that

$$\tilde{T}_{i,I}^{(\ell)} v \leq \theta v, \qquad \ell = 1,2,\ldots,L, \quad i = 1,2,\ldots.$$

By Lemma 2.5, we immediately get $[\varepsilon_i] \to 0$ when $i \to \infty$. This implies that $\varepsilon_i \to 0$ when $i \to \infty$, completing the proof of the Type-I case.

For the Type-II case, let us denote by $\tilde{P} = [D(A)^{-1}]P$, and for $\ell = 1,2,\ldots,L$, $\tilde{B}_\ell = D(A)^{-1}B_\ell$, $\tilde{C}_\ell = D(A)^{-1}C_\ell$, $\tilde{M}_\ell = D(A)^{-1}M_\ell$, and $\tilde{N}_\ell = D(A)^{-1}N_\ell$. With this notation, observe that $\tilde{M}_\ell \in H_B^{\mathrm{I}}$, and that by Lemma 2.2 (b), we have $[\tilde{M}_\ell^{-1}] \leq \langle \tilde{M}_\ell \rangle^{-1}$, $\ell = 1,2,\ldots,L$. Using these relations, taking block absolute values on both sides of (13) as before, after inserting $D(A)D(A)^{-1}$ in the appropriate places, we obtain

$$[\varepsilon_i] \leq \sum_{\ell=1}^{L} [E_\ell] \left((\langle \tilde{M}_\ell \rangle^{-1} [\tilde{N}_\ell])^{q(\ell,i)} \right.$$
$$\left. + \sum_{j=0}^{q(\ell,i)-1} (\langle \tilde{M}_\ell \rangle^{-1} [\tilde{N}_\ell])^j \langle \tilde{M}_\ell \rangle^{-1} ([\tilde{C}_\ell] + \tilde{P}) \right) [\varepsilon_{i-1}].$$

This expression has the same form as (18), with matrices of the same structure as (19). To complete the proof, we note that the splittings $D(A)^{-1}A = \tilde{B}_\ell - \tilde{C}_\ell$, and $\tilde{B}_\ell = \tilde{M}_\ell - \tilde{N}_\ell$ are H_B^{I}-compatible and correspond to the solution of the system $D(A)^{-1}Ax = D(A)^{-1}\Phi(x)$, which satisfy the hypotheses of the Type-I case. □

We should point out that the hypothesis of Theorem 3.4 can be changed to apply to the solution of another nonlinear system

$$Hx = \Phi(x), \tag{20}$$

where the matrix $H \in \mathbb{L}_{n,\mathrm{I}}$ is equimodular to A, i.e., it belongs to the set $\Omega_B^{\mathrm{I}}(A) = \{H = (H_{ij}) \in \mathbb{L}_{n,\mathrm{I}},\ \|H_{ii}^{-1}\| = \|A_{ii}^{-1}\|,\ \|H_{ij}\| = \|A_{ij}\|, i \neq j,\ i,j = 1,\ldots,L\}$, in the Type-I case, and to the set $\Omega_B^{\mathrm{II}}(A) = \{H = (H_{ij}) \in \mathbb{L}_{n,\mathrm{I}},\ \|H_{ii}^{-1}H_{ij}\| = \|A_{ii}^{-1}A_{ij}\|,\ i,j = 1,\ldots,L\}$ in the Type-II case; see [2]. Furthermore, Theorem 3.4 also applies, essentially with the same proof, to matrices A (or H) which may not have a nonsingular $D(A)$, or such that they do not belong to H_B^{I} or H_B^{II} but such that there exists nonsingular matrices R and S, such that RAS has those properties; see [2]. These comments apply to Theorem 4.4 in the next section as well.

We conclude this section by remarking that for the inner iterations one can chose, in addition to the classical splittings corresponding to Jacobi, Gauss-Seidel, etc., relaxed methods such as JOR, SOR, AOR, etc.. In the latter cases, the relaxation parameters need to be chosen in such a way that the splittings satisfy the hypotheses of Theorems 3.2, 3.3 and 3.4. This was done for the point methods in [1], and the same considerations carry through to the block methods studied here.

4. Asynchronous iterations

The computational model we consider for the asynchronous iterations is as follows. Each processor, say the ℓth processor, starts a cycle of computations by collecting the most recent vectors computed by the other processors, say the processors k, $k \neq \ell$. Let us call this cycle the ith iteration (which would be different than the ith iteration in the synchronous case). Thus, the iteration subscript is increased every time a processor starts a new cycle of computations. The vectors computed by the other processors will be from older cycles, some say $i-1$, but many others older than that, and we call these earlier cycles $r(k,i)$, i.e., the cycle in which the processor k computed the vector used at the beginning of the ith cycle (from this definition the condition (21) below follows directly). In other words, in order to get a new vector $x_i^{(\ell)}$, the ℓth processor collects the vectors $x_{r(k,i)}^{(k)}$, and uses the weighting matrices to get the vector $\sum_{k=1}^{L} E_k x_{r(k,i)}^{(k)}$. This sequence of weighted vectors can be considered in practice as the sequence of iterate vectors. Recall that in fact not all the components of the vectors $x_{r(k,i)}^{(k)}$ are needed in the computations, so that the local storage is of order n and not nL. Formally, we define the sets $J_i \subseteq \{1,2,\ldots,L\},\ i=1,2,\ldots$, as $\ell \in J_i$ if the ℓth processor starts its computation of a new iterate at the ith step.

As is customary in the description and analysis of asynchronous algorithms, we assume that the subscripts $r(\ell,i)$ and the sets J_i satisfy the following conditions. They appear as classical conditions in convergence results for asynchronous iterations; see e.g., [5], [8], [15], [20].

(21) $\quad r(\ell,i) < i$ for all $\ell = 1,2,\ldots L,\ i=1,2,\ldots$.

(22) $\quad \lim_{i\to\infty} r(\ell,i) = \infty$ for all $\ell = 1,2,\ldots,L$.

(23) $\quad$ The set $\{i \mid \ell \in J_i\}$ is unbounded for all $\ell = 1,2,\ldots,L$.

With this notation, the asynchronous counterpart of Algorithm 2 can be described by the following algorithm.

Algorithm 3 (Nonlinear Asynchronous Two-stage Multisplitting).
Given the initial vectors $x_0^{(\ell)} = x_0$, $\ell = 1,\ldots,L$.
For $i = 1,2,\ldots$

$$(24)\quad x_i^{(\ell)} = \begin{cases} x_{i-1}^{(\ell)} & \text{if } \ell \notin J_i \\ (M_\ell^{-1}N_\ell)^{q(\ell,i)} \sum_{k=1}^{L} E_k x_{r(k,i)}^{(k)} & \\ + \sum_{j=0}^{q(\ell,i)-1} (M_\ell^{-1}N_\ell)^j M_\ell^{-1}(C_\ell \sum_{k=1}^{L} E_k x_{r(k,i)}^{(k)} & \\ \quad + \Phi(\sum_{k=1}^{L} E_k x_{r(k,i)}^{(k)})) & \text{if } \ell \in J_i. \end{cases}$$

In order to analyze Algorithm 3, we consider the operators $G(i) = (G^{(1)}(i),\ldots,G^{(L)}(i))$, with $G^{(\ell)}(i) : \mathbb{R}^{nL} \to \mathbb{R}^n$ defined for $\hat{y} \in \mathbb{R}^{nL}$ as follows, for $\ell = 1, \ldots, L$ and $i = 1, 2, \ldots$.

$$G^{(\ell)}(i)(\hat{y}) = (M_\ell^{-1}N_\ell)^{q(\ell,i)}Q\hat{y} + \sum_{j=0}^{q(\ell,i)-1} (M_\ell^{-1}N_\ell)^j M_\ell^{-1}(C_\ell Q\hat{y} + \Phi(Q\hat{y})),$$

where $Q = [E_1, \ldots, E_\ell, \ldots, E_L] \in \mathbb{R}^{n \times nL}$. The asynchronous iteration (24) can then be rewritten as the following iteration.

$$(25) \qquad x_i^{(\ell)} = \begin{cases} x_{i-1}^{(\ell)} & \text{if } \ell \notin J_i \\ G^{(\ell)}(i)\left(x_{r(1,i)}^{(1)}, \ldots, x_{r(\ell,i)}^{(\ell)}, \ldots, x_{r(L,i)}^{(L)}\right) & \text{if } \ell \in J_i. \end{cases}$$

The following lemma, which is a special case of Theorem 3.2 in [26], is used in our convergence proofs.

Lemma 4.1. *Let $G(i)$ be a sequence of operators on $\mathbb{R}^{nL}$ having a common fixed point $\hat{x}_\star$. Let $\|\cdot\|_\ell$ be a norm on $\mathbb{R}^n$, $\ell = 1, \ldots, L$. Let $a \in \mathbb{R}^L$, $a > 0$ and denote $\|\cdot\|_a$ the weighted max-norm $\|x\|_a = \max_{1 \le \ell \le L}\{\frac{1}{a_\ell}\|x^{(\ell)}\|_\ell\}$. For all $i = 1, 2, \ldots$, assume that there exists a constant $0 \le \alpha < 1$ such that*

$$\|G(i)\hat{x} - \hat{x}_\star\|_a \le \alpha\|\hat{x} - \hat{x}_\star\|_a, \text{ for all } \hat{x} \in \mathbb{R}^{nL}.$$

Assume further that the sequence $r(\ell, i)$ and the sets J_i, $\ell = 1, \ldots, L$, $i = 1, 2, \ldots$, satisfy conditions (21–23)*. Then the asynchronous iteration* (25) *converges to $\hat{x}_\star$ for any initial vector $\hat{x}_0$.*

Theorem 4.2. *Let $A \in \mathbb{R}^{n\times n}$ be a monotone matrix. Let the splittings $A = B_\ell - C_\ell$, $\ell = 1, \ldots, L$, be regular and the splittings $B_\ell = M_\ell - N_\ell$, $\ell = 1, \ldots, L$, be weak regular. Suppose that $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is a P-bounded mapping such that $\rho(A^{-1}P) < 1$. Assume further that the sequence $r(\ell, i)$ and the sets J_i, $\ell = 1, \ldots, L$, $i = 1, 2, \ldots$, satisfy conditions* (21–23)*. Then, the nonlinear asynchronous two-stage multisplitting Algorithm 3 converges to $(x_\star^{\mathrm{T}}, \ldots, x_\star^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{nL}$, where $x_\star$ is the unique solution of the mildly nonlinear system* (1)*, for any initial vectors $x_0^{(\ell)}$, $\ell = 1, \ldots, L$, and any sequence of numbers of inner iterations $q(\ell, i) \ge 1$, $\ell = 1, \ldots, L$, $i = 1, 2, \ldots$.*

Proof. From the proof of Theorem 3.2 we know that there exists a positive vector v and a constant $0 \le \theta < 1$ such that

$$(26) \qquad T_i^{(\ell)} v \le \theta v, \ \ell = 1, \ldots, L, \ i = 1, 2, \ldots,$$

where the matrices $T_i^{(\ell)}$ are defined in (15). Let us define $\hat{v} = (v^{\mathrm{T}}, \ldots, v^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{nL}$ and $\hat{x}_\star = (x_\star^{\mathrm{T}}, \ldots, x_\star^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{nL}$. As $\hat{x}_\star$ is a fixed point of $G(i)$, it is easy to prove that

$$\left|G^{(\ell)}(i)\hat{x} - x_\star\right| \le T_i^{(\ell)} Q|\hat{x} - \hat{x}_\star|, \ \ell = 1, \ldots, L, \ i = 1, 2, \ldots,$$

for all $\hat{x} \in \mathbb{R}^{nL}$. Thus

$$|G(i)\hat{x} - \hat{x}_\star| \leq H(i)|\hat{x} - \hat{x}_\star|, \; i = 1, 2, \ldots, \tag{27}$$

where

$$H(i) = \begin{bmatrix} T_i^{(1)}Q \\ \vdots \\ T_i^{(L)}Q \end{bmatrix} \in \mathbb{R}^{nL \times nL}. \tag{28}$$

From (28) and (26), it follows that $H(i)\hat{v} \leq \theta\hat{v}$. Hence, using the monotonic vector norm $\|\cdot\|_{\hat{v}}$ defined in (3), $\|H(i)\|_{\hat{v}} \leq \theta$, and then by (27),

$$\|G(i)\hat{x} - \hat{x}_\star\|_{\hat{v}} \leq \theta\|\hat{x} - \hat{x}_\star\|_{\hat{v}}, \; i = 1, 2, \ldots.$$

Since the norm $\|\hat{y}\|_{\hat{v}}$ in $\mathbb{R}^{nL}$ can be expressed in the form $\|\hat{y}\|_{\hat{v}} = \max\limits_{1 \leq \ell \leq L} \|\hat{y}^{(\ell)}\|_v$, for any $\hat{y} = (\hat{y}^{(1)^{\mathrm{T}}}, \ldots, \hat{y}^{(L)^{\mathrm{T}}})^{\mathrm{T}}$, using Lemma 4.1, the convergence is shown. □

Again, as with the hypotheses of Theorem 3.2, in Theorem 4.2 the outer splittings need not be regular, but just weak regular with the additional hypothesis (14). We remark that an alternative proof could be obtained using the theory of paracontracting operators; see [16].

Theorem 4.3. *Let $A \in \mathbb{R}^{n\times n}$ be an H-matrix. Let the splittings $A = B_\ell - C_\ell$, and $B_\ell = M_\ell - N_\ell$, $\ell = 1, \ldots, L$, be H-compatible. Suppose that $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is a P-bounded mapping such that $\rho(< A >^{-1} P) < 1$. Assume further that the sequence $r(\ell, i)$ and the sets J_i, $\ell = 1, \ldots, L$, $i = 1, 2, \ldots,$ satisfy conditions* (21–23)*. Then, the nonlinear asynchronous two-stage multisplitting Algorithm 3 converges to $(x_\star^{\mathrm{T}}, \ldots, x_\star^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{nL}$, where $x_\star$ is the unique solution of the mildly nonlinear system* (1)*, for any initial vectors $x_0^{(\ell)}$, $\ell = 1, \ldots, L$, and any sequence of numbers of inner iterations $q(\ell, i) \geq 1$, $\ell = 1, \ldots, L$, $i = 1, 2, \ldots$.*

Proof. Using similar notation as in Theorem 4.2 we can prove, for all $\hat{x} \in \mathbb{R}^{nL}$, that

$$\left|G^{(\ell)}(i)\hat{x} - x_\star\right| \leq \tilde{T}_i^{(\ell)}Q|\hat{x} - \hat{x}_\star|, \; \ell = 1, \ldots, L, \; i = 1, 2, \ldots,$$

where matrices $\tilde{T}_i^{(\ell)}$ are defined in (17). By the proof of Theorem 3.3 we have $\tilde{T}_i^{(\ell)}\tilde{v} \leq \theta\tilde{v}$ for some $\tilde{v} \in \mathbb{R}^n, \; \tilde{v} > 0$ and $\ell = 1, \ldots, L$, and thus the theorem follows in the same manner as Theorem 4.2. □

Analogous to Theorem 4.3, as well as Theorem 3.4, we can establish the following convergence theorem for the systems (1) and (20). Since this new theorem can be demonstrated in similar ways to the proof of Theorem 3.4 with slight modifications, we omit its proof.

Theorem 4.4. *Let $A \in H_B^{\mathrm{I}}$ $(H_B^{\mathrm{II}}) \subset \mathbb{L}_{n,\mathrm{I}}(n_1, n_2, \ldots, n_L)$. Let the splittings $A = B_\ell - C_\ell$ and $B_\ell = M_\ell - N_\ell$, $\ell = 1, 2, \ldots, L$, be H_B^{I}-compatible (H_B^{II}-compatible and such that $D(M_\ell) = D(B_\ell) = D(A)$), and the weighting matrices E_ℓ, $\ell = 1, 2, \ldots, L$, satisfying* (4)*, satisfy in addition $\sum_{\ell=1}^{L}[E_\ell] \leq I$, the $L \times L$ identity matrix. Assume further that $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is a block P-bounded mapping such that $\rho(\langle A\rangle^{-1}P) < 1$ ($\rho(\langle\langle A\rangle\rangle^{-1}[D(A)^{-1}]P) < 1$). Then, the nonlinear asynchronous two-stage multisplitting Algorithm 3 converges to $(x_\star^{\mathrm{T}}, \ldots, x_\star^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{nL}$, where $x_\star$ is the unique solution of the mildly nonlinear system* (1)*, for any initial vectors $x_0^{(\ell)}$, $\ell = 1, \ldots, L$, and any sequence of numbers of inner iterations $q(\ell, i) \geq 1$, $\ell = 1, \ldots, L$, $i = 1, 2, \ldots$.*

We point out that when $\Phi(x) = b \in \mathbb{R}^n$, i.e., when system (1) is linear, Algorithm 3 reduces to Algorithm 6 in [13], and thus Theorems 4.2, 4.3 and 4.4 apply to that case as well. In fact, the results here are more general, since we do not assume that the weighting matrices form a partition of the identity, i.e., when the entries of each E_ℓ are 0 or 1. In particular, Theorems 4.2, 4.3 and 4.4 provide general convergence results for two-stage multisplitting methods with overlap.

We end the paper with a discussion on a different asynchronous computational model, analogous to the one used in [11] for linear systems. Namely, for $i = 0, 1, 2, \ldots,$

$$x_{i+r_i} = (I - E_{j_i})x_{i+r_i-1} + E_{j_i}\Bigg((M_\ell^{-1}N_\ell)^{q(\ell,i)}x_i + \sum_{j=0}^{q(\ell,i)-1}(M_\ell^{-1}N_\ell)^j M_\ell^{-1}(C_\ell x_i + \Phi(x_i))\Bigg), \tag{29}$$

where $\{j_i\}_{i=0}^{\infty}$, $1 \leq j_i \leq L$, is a sequence of integers that indicates the processor which updates the approximation to the solution at the ith iteration and $r_i - 1$ is the number of times that processors other than the j_ith processor update the approximation of the solution during the time interval in which the j_ith processor's calculations are performed.

The computational model (29) is based on Model B of [10], and conditions (21–23) are replaced by the assumption that the sequence $\{j_i\}_{i=0}^{\infty}$ is regulated; see e.g., [13] for differences and analogies between both sets of conditions. The proof of the convergence of the iteration (29) with the same hypotheses as in Theorems 3.2, 3.3 and 3.4 follows in a similar way as the proof of [31, Theorem 3.2] which in turn is based on [10, Theorem 2.2], and as the proof of [4, Theorem 1].

The two asynchronous models (24) and (29) produce the same sequence of iterate vectors when the weighting matrices E_ℓ, $\ell = 1, \ldots, L$, form a partition of the identity, but when there is overlap, the models are different. In Algorithm 3 the components calculated by the ℓth processor, say, uses only information calculated by the other processors ($k \neq \ell$). However, in model (29) the ℓth processor introduces in its computations some older information computed in the ℓth processor in a previous step.

References

1. Bai, Z.-Z. (1997): A class of two-stage iterative methods for systems of weakly nonlinear equations. Numer. Algorithms **14(4)**, 295–319
2. Bai, Z.-Z. (1995): Parallel matrix multisplitting block relaxation methods. Mathematica Numerica Sinica **17**, 238–252 In Chinese.
3. Bai, Z.-Z. (1998): A class of blockwise matrix multi-splitting multi-parameter block relaxation methods. In: Shi, Z.-C., Mori, M. (eds.) Proceeding of the Third China-Japan Joint Seminar on Numerical Mathematics, Science Press of China, Beijing, pp. 1–10
4. Bai, Z.-Z. (1997): Asynchronous parallel matrix multisplitting block relaxation iteration methods. Numerical Mathematics, A Journal of Chinese Universities (Chinese Series) **19**, 28–39. In Chinese.
5. Baudet, G.M. (1978): Asynchronous iterative methods for multiprocessors. Journal of the Association for Computing Machinery **25**, 226–244
6. Benzi, M., Choi, H., Szyld, D.B. (1997): Threshold Ordering for Preconditioning Nonsymmetric Problems. In: Golub, G., et.al. (eds.) Scientific Computing, Proceedings of the Workshop, 10–12 March 1997, Hong Kong, Springer, Singapore, pp 159–165
7. Berman, A., Plemmons, R.J. (1994): Nonnegative Matrices in the Mathematical Sciences. Academic Press, New York, third edition, 1979. Reprinted by SIAM, Philadelphia
8. Bertsekas, D.P., Tsitsiklis, J.N. (1989): Parallel and Distributed Computation. Prentice-Hall, Englewood Cliffs, New Jersey
9. Bhaya, A., Kaszkurewicz, E., Mota, F. (1991): Asynchronous Block-Iterative Methods for Almost Linear Equations. Linear Algebra and its Applications **154–156**, 487–508
10. Bru, R., Elsner, L., Neumann, M. (1988): Models of parallel chaotic iteration methods. Linear Algebra and its Applications **103**, 175–192
11. Bru, R., Migallón, V., Penadés, J. (1994): Chaotic inner-outer iterative schemes. In: Lewis, J.G. (ed.) Proceedings of the Fifth SIAM Conference on Applied Linear Algebra, pp 434–438. SIAM, Philadelphia
12. Bru, R., Migallón, V., Penadés, J. (1995): Chaotic methods for the parallel solution of linear systems. Computing Systems in Engineering **6**, 385–390
13. Bru, R., Migallón, V., Penadés, J., Szyld, D.B. (1995): Parallel, synchronous and asynchronous two–stage multisplitting methods. Electronic Transactions on Numerical Analysis **3**, 24–38
14. Dembo, R.S., Eisenstat, S.C., Steihaug, T. (1982): Inexact Newton methods. SIAM Journal on Numerical Analysis **19**, 400–408
15. El Tarazi, M.N. (1982): Some convergence results for asynchronous algorithms. Numerische Mathematik **39**, 325–340, in French
16. Elsner, L., Koltracht, I., Neumann, M. (1992): Convergence of sequential and asynchronous nonlinear paracontractions. Numerische Mathematik **62**, 305–319

17. Evans, D.J., Bai, Z.-Z. (1997): Blockwise matrix multi-splitting multi-parameter block relaxation methods. International Journal of Computer Mathematics **64**, 103–118
18. Feingold, D.G., Varga, R.S. (1962): Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem. Pacific Journal of Mathematics **12**, 1241–1250
19. Frommer, A. (1989): Parallel nonlinear multisplitting methods. Numerische Mathematik **56**, 269–282
20. Frommer, A. (1991): On asynchronous iterations in partially ordered spaces. Numerical Functional Analysis and Optimization **12**, 315–325
21. Frommer, A., Mayer, G. (1992): On the theory and practice of multisplitting methods in parallel computation. Computing **49**, 63–74
22. Frommer, A., Pohl, B. (1995): A comparison result for multisplittings and waveform relaxation methods. Numerical Linear Algebra with Applications **2**, 335–346
23. Frommer, A., Schwandt, H. (1997): A unified representation and theory of algebraic additive Schwarz and multisplitting methods, SIAM Journal on Matrix Analysis and Applications **18**, 893–912
24. Frommer, A., Schwandt, H., Szyld, D.B. (1997): Asynchronous weighted additive Schwarz methods. Electronic Transactions on Numerical Analysis **5**, 48–61
25. Frommer, A., Szyld, D.B. (1992): H-splittings and two-stage iterative methods. Numerische Mathematik **63**, 345–356
26. Frommer, A., Szyld, D.B. (1994): Asynchronous two-stage iterative methods. Numerische Mathematik **69**, 141–153
27. Fuster, R., Migallón, V., Penadés, J. (1996): Parallel chaotic extrapolated Jacobi–like methods. Linear Algebra and its Applications **247**, 237–250
28. Huang, C.-M., O'Leary, D.P. (1993): A Krylov multisplitting algorithm for solving linear systems of equations. Linear Algebra and its Applications **194**, 9–29
29. Jones, M.T., Szyld, D.B. (1996): Two-stage multisplitting methods with overlapping blocks. Numerical Linear Algebra with Applications **3**, 113–124
30. Householder, A.S. (1975): The Theory of Matrices in Numerical Analysis. Blaisdell, Waltham, Mass. 1964. Reprinted by Dover, New York
31. Mas, J., Migallón, V., Penadés, J., Szyld, D.B. (1966): Non-stationary parallel relaxed multisplitting methods. Linear Algebra and its Applications **241/243**, 733–748
32. Moré, J.J. (1971): Global convergence of Newton-Gauss-Seidel methods. SIAM Journal on Numerical Analysis **8**, 325–336
33. Neumaier, A. (1984): New techniques for the analysis of linear interval equations. Linear Algebra and its Applications **58**, 273–325
34. O'Leary, D.P., White, R.E. (1985): Multi-splittings of matrices and parallel solution of linear systems. SIAM Journal on Algebraic and Discrete Methods **6**, 630–640
35. O'Neil, J., Szyld, D.B. (1990): A block ordering method for sparse matrices. SIAM Journal on Scientific and Statistical Computing **11**, 811–823
36. Ortega, J.M., Rheinboldt, W.C. (1970): Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, San Diego
37. Ostrowski, A.M. (1937): Über die determinanten mit überwiegender hauptdiagonale. Comentarii Mathematici Helvetici **10**, 69–96
38. Polman, B. (1987): Incomplete blockwise factorization of (block) H-matrices. Linear Algebra and its Applications **90**, 119–132
39. Renaut, R.A., Mittelmann, H.D., He, Q. (1994): Parallel multisplittings: overview and extensions. In: Lewis, J.G. (ed.) Proceedings of the Fifth SIAM Conference on Applied Linear Algebra, pp 34–38. SIAM, Philadelphia
40. Robert, F. (1969): Blocs H-Matrices et Convergence des Méthodes Itératives Classiques par Blocs. Linear Algebra and its Applications **2**, 223–265

41. Robert, F., Charnay, M., Musy, F. (1975): Itérations chaotiques série-parallèle pour des équations non-linéaires de point fixe. Aplikace Matematiky **20**, 1–38
42. Varga, R.S. (1962): Matrix Iterative Analysis. Prentice-Hall, Englewood Cliffs, New Jersey