Parallel Computing 19 (1993) 281–302
North-Holland

PARCO 752

Convergence analysis of asynchronous linear iterations with stochastic delays

Bassem F. Beidas and George P. Papavassilopoulos

Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089-2563, USA

Received 7 March 1991
Revised 18 October 1991, 11 May 1992, 26 June 1992

Abstract

Beidas, B.F. and G.P. Papavassilopoulos, Convergence analysis of asynchronous linear iterations with stochastic delays, Parallel Computing 19 (1993) 281–302.

We present a general linear model of asynchronous iterations, the communication delays of which are stochastic with Markovian character. This model allows static or dynamic allocation of the iterate vector components to processors. It also allows simultaneous updating of the same vector component by multiple processors. Sufficient conditions under which the model of asynchronous iterations converges in the second moment (and in the mean) to the sought solution are provided. For the specialization of the Markov case when the communication delays are independent, identically distributed (i.i.d.), we provide sufficient conditions for convergence in the second moment and necessary and sufficient conditions for convergence in the mean.

Keywords. Linear algebra; iterative algorithms; asynchronous iterations; convergence analysis; multiprocessor systems; stochastic delays.

1. Introduction

Asynchronous implementation of iterative algorithms has recently witnessed extensive attention [3,18,19]. Asynchronous iterations allow processors not to fall idle but further compute after performing an update without waiting for an a priori fixed set of data. Hence, processors compute with available information at full speed. Asynchronous iterations are shown [8] to enhance efficiency in a parallel processing environment.

The model we present permits static allocation of components to different processors in which each component gets assigned at the start of the algorithm to be updated by the same processor. For better load balancing among processors, dynamic allocation can be employed in which the assignment of components may change with time. In an effort to elevate the level of tolerance to processor failure, the model also allows concurrent updating of the same component by several processors. It is allowed that the processors use their individual updates with a possible bounded delay.

In an early paper on asynchronous iterations, Chazan and Miranaker [5] discussed linear system of equations with nonnegative coefficients. Kung [8] considered issues that pertain to

Correspondence to: G.P. Papavassilopoulos, Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089-2563, USA.

the implementation of asynchronous iterations in multiprocessors. Baudet [1] proved a convergence condition for problems where no bound exists on the communication delays. Some applications are discussed in [11,13]. Tsitkilis, Bertsekas and Athans [15] studied an asynchronous model that seems to be the first of its class to treat optimization problems. In spite of assuming that the delays were unknown, they were dealt with in a deterministic manner. In Kushner and Yin [10], a model related to the one studied here is considered. The differences, however, are twofold. First, our model takes no regard to the structure of the matrix sequence which is associated with the augmented system model defined in Section 2. In [10] they discussed a model whose matrix sequence possesses the special form which has nonegative entries and the sum of its rows equals one. Second, in [10] the matrix sequence is assumed to have a unique stationary distribution. In our model, this property does not necessarily hold. A scheme can be implemented (see [17] for the deterministic case) to ensure that the receipt of updates conforms to the order with which they are produced. In particular, we exploit a proper labelling process to impose that if the arrival of $x_j(k)$ to processor i is attained, future computations will be prevented from using iterates prior to $x_j(k-1)$ inclusive. It is shown (see [16] for the case of i.i.d. communication delays) that the aforementioned scenario produces a good communication complexity in which managability of the number of transmitted messages is preserved.

A key assumption in the algorithms under study is that the communication delays among collaborating processors are assumed to be stochastic. Cases with stochastic delays appear when the load of the communication network that the processors use varies in an unpredictable manner, or some of its links are temporarily incapacitated. Therefore, the random delay framework addresses the varying load conditions of the communication network as well as the reliability issue of temporary link or processor failure. In this paper, we derive sufficiency conditions that guarantee the convergence of the asynchronous linear iterations with stochastic delays. The virtue of these conditions is that they are developed not in the usual sense of the nonexpansiveness property available in [3,14]. Instead, we pass the iterations to a certain augmented space which is specified by the delay bound and utilize the Kronecker product notation [4] to facilitate the computation of the second moment of linear systems, the coefficients of which are Markov chains.

In Section 2, we introduce the main model of asynchronous iterations in which specialized and overlapped computations are used. Motivations behind such classification of the main model are presented. State space formulae are given for the purpose of obtaining a generic description of the model. Convergence analysis is discussed in Section 3. Our main result provides sufficient conditions under which the main model converges in the mean and second moment to the origin for every initial condition. This is partly achieved by utilizing the generalized formula, represented by Equation (15), to the equation found in [4]. We discuss and provide a weaker set of conditions for the situation in which the Markov chain associated with the communication delays takes a special property. In particular, we discuss the case when the initial distribution of the Markov chain is a stationary one. Furthermore, we study the models of ordered scheduling and independent delays. Finally, in Section 4 we present an example, the significance of which is to gain insight into the conditions furnished by our main result. Lengthy formula derivations and properties of Kronecker products and sums are relegated in the Appendix.

2. System model

In this section, we introduce the basic problem that will be studied, see Equations (9a)–(9c). We start by considering several cases that give rise to this model.

As processor j finishes computing, it propagates its value to other processors. With a possible delay, processor i combines the value of its own computation with the information received from processor j multiplied by the constant a_{ij}. We also assume that processor i picks up random measurement noise with every computation it completes.

Case (i). Specialized computations

We discuss a model of specialized computations in which each component of the vector is updated by at most one processor at the same time. We let $d_{ji}(k)$ be the delay incurred by transmitting a message from processor j to processor i at the kth iteration. We assume that the communication delays are bounded. We allow the bound of these communication delays to be different depending on the source and destination of the transmitted messages. This can be attributed to the location of processors in a certain network where some processors are placed farther apart than other processors. Consequently, we expect the delay bound of the processor transmitting messages to one closer to it to be smaller in comparison with the bound of the messages sent to a distant processor.

The processors perform the updating according to the recursive scheme

$$x_i(k+1) = \sum_{j=1}^{q} a_{ij} x_j\big(k+1-d_{ji}(k)\big) + \gamma_i(k) r_i(k). \tag{1}$$

We let $\{d_{ji}(k)\}$, for all j and i, be a stationary Markov chain with state space

$$S_{ji} = \{1, 2, \ldots, B_{ji}\},$$

where B_{ji} is the maximum allowable communication delay for messages sent from processor j to processor i. We let te probability transition matrix corresponding to $d_{ji}(k)$ be $\tilde{P}_{ji} = (\tilde{p}_{ji}(l, m))$, where

$$\tilde{p}_{ji}(l, m) = \text{Prob}\{d_{ji}(k) = m \mid d_{ji}(k-1) = l\}, \quad \text{for } l, m = 1, 2, \ldots, B_{ji}. \tag{2}$$

The initial distribution of $d_{ji}(0)$ can be arbitrary. In some instances, it is required that the knowledge of the initial conditions for $t \le 0$ be available. We, however, let the sequence generated by Equation (1) be defined only for nonnegative times and assume that the initializations $\{x_1(0), \ldots, x_q(0)\}$ of the algorithm are random with finite mean and finite variance. We also let $\{r_i(k)\}$ be a sequence of zero-mean independent identically distributed random variables with finite variance and $\gamma_i(k)$ be a nonnegative number.

In our model, the communication delays, $d_{ji}(k)$, can be stochastic in nature due to several factors: unknown load conditions resulting from other users, possible failures of some processors, etc. Their stochasticity can also be deliberately imposed in an effort to localize communication around processors of interest. Assume that it is possible to group the processors involved in our model into different clusters of processors according to their level of computational efficiency. For the purpose of maintaining a low communication overhead, we reduce the frequency of message propagation from the cluster consisting of less computationally efficient processors to the others. Let F_l be the set of indices that correspond to the processors of this cluster, where

$$F_l \subset \{1, 2, \ldots, q\}.$$

We choose f_l such that

$$1 < f_l \le B_{ji}$$

and impose that

$$\text{Prob}\{d_{ji}(k) < f_l\} = 0, \quad \text{for every } k,\ j \in F_l \text{ and } i \notin F_l. \tag{3}$$

Case (ii). Overlapped computations

We introduce the model of overlapped computations where each component can be simultaneously updated by several processors through which less vulnerability to processor failure is achieved. If some processors cease updating at a certain time instant and on, asynchronous iteration will rely on the rest for generating the values of these components of $x(k)$. We use superscripts to indicate processors involved, i.e. $x_j^i(k)$ is the jth component produced by processor i at the kth time instant. As a consequence of the unpredictable behavior of communication delays, various processors may render different updates of the same component. An averaging process among these updates is shown [15] to bring them closer to parity. A different technique can be exploited in which a processor that is faced with different updates of the same component selects its input with a certain probability. Assume that we have q processors to iterate on the vector $x(k)$ with Q components. For every component index i, we let $F_i \subset \{1, 2, \ldots, q\}$ be the set of processors updating the ith component x_i. Also for every processor i, we let $D_i \subset \{1, 2, \ldots, Q\}$ be the set of components updated by processor i. We let $d_{ji}^{lm}(k)$ be the communication delay resulting from sending component j calculated by processor l to be used in the updating of component i by processor m at the kth time instant and only define $d_{ji}^{lm}(k)$ for $j \in F_l$ and $i \in D_m$. The Markovian character of the $d_{ji}^{lm}(k)$ can also be specified as was done in the previous case. Therefore, the processors perform the updating as follows

$$x_i^m(k+1) = \sum_{j=1}^{Q} a_{ij} s_{ji}^m(k) + \gamma_i^m(k) r_i^m(k), \quad \forall i \in D_m, \quad m = 1, \ldots, q, \tag{4}$$

where

$$s_{ji}^m(k) = x_j^l\big(k+1-d_{ji}^{lm}(k)\big), \quad \forall l \in F_j, \quad j, i = 1, \ldots, Q \quad \text{and } m = 1, \ldots, q$$

and x_j^l are allowed to occur with some probability for different l. It is assumed that $s_{ji}^m(k)$ are zero when $d_{ji}^{lm}(k)$ are undefined.

Example 1. We consider a system involving four processors that are responsible for updating the vector $x(k)$ of two components. In particular, the first two processors update the first component $x_1(k)$ while the other two processors update the second component $x_2(k)$. In this case, we have $Q = 2$, $q = 4$ and

$$\begin{aligned} &F_1 = \{1, 2\}, \quad &&F_2 = \{3, 4\} \\ &D_1 = \{1\}, &&D_2 = \{1\} \\ &D_3 = \{2\}, &&D_4 = \{2\}. \end{aligned}$$

Therefore, the processors update as follows

$$x_1^1(k+1) = a_{11} s_{11}^1(k) + a_{12} s_{21}^1(k) + \gamma_1^1(k) r_1^1(k)$$

$$x_1^2(k+1) = a_{11} s_{11}^2(k) + a_{12} s_{21}^2(k) + \gamma_1^2(k) r_1^2(k)$$

$$x_2^3(k+1) = a_{21} s_{12}^3(k) + a_{22} s_{22}^3(k) + \gamma_2^3(k) r_2^3(k)$$

$$x_2^4(k+1) = a_{21} s_{12}^4(k) + a_{22} s_{22}^4(k) + \gamma_2^4(k) r_2^4(k),$$

where

$$s_{11}^1(k)=\begin{cases}x_1^1(k+1-d_{11}^{11}(k))\\x_1^2(k+1-d_{11}^{21}(k))\end{cases};\quad s_{21}^1(k)=\begin{cases}x_2^3(k+1-d_{21}^{31}(k))\\x_2^4(k+1-d_{21}^{41}(k))\end{cases}$$

$$s_{11}^2(k)=\begin{cases}x_1^1(k+1-d_{11}^{12}(k))\\x_1^2(k+1-d_{11}^{22}(k))\end{cases};\quad s_{21}^2(k)=\begin{cases}x_2^3(k+1-d_{21}^{32}(k))\\x_2^4(k+1-d_{21}^{42}(k))\end{cases}$$

$$s_{12}^3(k)=\begin{cases}x_1^1(k+1-d_{12}^{13}(k))\\x_1^2(k+1-d_{12}^{23}(k))\end{cases};\quad s_{22}^3(k)=\begin{cases}x_2^3(k+1-d_{22}^{33}(k))\\x_2^4(k+1-d_{22}^{43}(k))\end{cases}$$

$$s_{12}^4(k)=\begin{cases}x_1^1(k+1-d_{12}^{14}(k))\\x_1^2(k+1-d_{12}^{24}(k))\end{cases};\quad s_{22}^4(k)=\begin{cases}x_2^3(k+1-d_{22}^{34}(k))\\x_2^4(k+1-d_{22}^{44}(k))\end{cases}$$

Assuming that $d_{11}^{11}(k)$ and $d_{11}^{21}(k)$ take values from {1, 2} and {1, 2, 3}, respectively, we can have several schemes for $s_{11}^1(k)$. For example,

$$s_{11}^1(k)=x_1^1(k),\ x_1^1(k-1). \tag{5}$$

Equation (5) represents the case as processor 1 only considers the iterates provided by its own computations to update $x_1(k)$. Another case is to allow $s_{11}^1(k)$ to use the following choices

$$s_{11}^1(k)=x_1^1(k),\quad x_1^1(k-1),\quad x_1^2(k),\quad x_1^2(k-1),\quad x_1^2(k-2). \tag{6}$$

We assume that the choices in Equation (6) occur with some probabilities.

2.1 State augmentation

For the purpose of easily observing the behavior of the models generated by Equations (1) and (4), we utilize state augmentation to formulate a suitable state-space representation in which a generic description of the models is provided. We apply the state augmentation procedure to the model of specialized computation of Case (i). We let

$$B_i^*=\max_j B_{ij}, \tag{7}$$

and introduce the vectors $y(k)$ and $w(k)$ of dimension $n=\sum_{i=1}^q B_i^*$ defined in the following manner:

$$y(k)\triangleq\begin{bmatrix}x_1(k)\\x_1(k-1)\\\vdots\\x_1(k-B_1^*+1)\\x_2(k)\\x_2(k-1)\\\vdots\\x_2(k-B_2^*+1)\\\vdots\\\vdots\\x_q(k)\\x_q(k-1)\\\vdots\\x_q(k-B_q^*+1)\end{bmatrix};\quad w(k)\triangleq\begin{bmatrix}r_1(k)\\0\\\vdots\\0\\r_2(k)\\0\\\vdots\\0\\\vdots\\\vdots\\r_q(k)\\0\\\vdots\\0\end{bmatrix}. \tag{8}$$

Similarly, we can use state augmentation for the model of overlapped computation of Case (ii).

This suggests that the above models can be rewritten in the following form

$$y(k+1) = C(k)y(k) + \Gamma(k)w(k), \tag{9a}$$

where $\{C(k)\}$ is a sequence of square constant sparse matrices containing a_{ij} as the only nonvanishing elements and $\{w(k)\}$ is a sequence of zero-mean i.i.d. random vectors with finite variance, independent of $\{C(k)\}$ and $\{y(k)\}$. Furthermore, $\{\Gamma(k)\}$ is a sequence of diagonal matrices where the only nonzero entries of this diagonal are $\gamma_i(k)$ and the initial condition $y(0)$ is random with finite mean and finite variance. The matrix sequence $\{C(k)\}$ is a stationary Markov chain of finite state space

$$\Omega = \{C_1, C_2, \ldots, C_l\},$$

with probability of transition $P = (p_{ij})$, where the probability of evolving to the matrix C_j at the kth iteration given that we started with C_i at the $(k-1)$th iteration is

$$p_{ij} = \text{Prob}\{C(k) = C_j \mid C(k-1) = C_i\}. \tag{9b}$$

Let the row vector p_0 define the probability distribution of the initial state, where

$$p_{0i} = \text{Prob}\{C(0) = C_i\}, \quad p_{0i} \geq 0 \quad \text{and} \quad \sum_{i=1}^{l} p_{0i} = 1. \tag{9c}$$

Example 2. We consider a system representing a model of specialized computation of Case (i) involving two processors where $B_{11} = B_{22} = 1$, $B_{12} = 2$, $B_{21} = 3$ and $\{d_{ji}(k)\}$ is a Markov chain with transition probability $\tilde{P}_{ji}$ and state space $S_{ji} = \{1, 2, \ldots, B_{ji}\}$. The algorithm iterates on the vector $x(k)$ whose ith component is iterated upon by processor i according to

$$\begin{aligned} x_1(k+1) &= a_{11}x_1(k) + a_{12}x_2(k+1-d_{21}(k)) + \gamma_1(k)r_1(k) \\ x_2(k+1) &= a_{21}x_1(k+1-d_{12}(k)) + a_{22}x_2(k) + \gamma_2(k)r_2(k). \end{aligned} \tag{10}$$

For this case, we write

$$y(k) = \begin{bmatrix} x_1(k) \\ x_1(k-1) \\ x_2(k) \\ x_2(k-1) \\ x_2(k-2) \end{bmatrix}.$$

The $C_1, C_2, \ldots, C_6$ have the following forms

$$C_1 = \begin{bmatrix} a_{11} & 0 & a_{12} & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ a_{21} & 0 & a_{22} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}; \quad C_2 = \begin{bmatrix} a_{11} & 0 & 0 & a_{12} & 0 \\ 1 & 0 & 0 & 0 & 0 \\ a_{21} & 0 & a_{22} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$C_3 = \begin{bmatrix} a_{11} & 0 & 0 & 0 & a_{12} \\ 1 & 0 & 0 & 0 & 0 \\ a_{21} & 0 & a_{22} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}; \quad C_4 = \begin{bmatrix} a_{11} & 0 & a_{12} & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & a_{21} & a_{22} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$C_5 = \begin{bmatrix} a_{11} & 0 & 0 & a_{12} & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & a_{21} & a_{22} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}; \quad C_6 = \begin{bmatrix} a_{11} & 0 & 0 & 0 & a_{12} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & a_{21} & a_{22} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

$\Gamma(k)$ and $w(k)$ can be written as follows

$$\Gamma(k) = \begin{bmatrix} \gamma_1(k) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_2(k) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad w(k) = \begin{bmatrix} r_1(k) \\ 0 \\ r_2(k) \\ 0 \\ 0 \end{bmatrix}.$$

In order to relate the probability of motion among the various states of the state space Ω to the probability transition matrix corresponding to $\{d_{ij}(k)\}$, we evaluate

$$\begin{aligned} &\mathrm{Prob}\{C(k) = C_1 \mid C(k-1) = C_1\} \\ &\quad = \mathrm{Prob}\{d_{12}(k) = 1,\, d_{21}(k) = 1 \mid d_{12}(k-1) = 1,\, d_{21}(k-1) = 1\} \\ &\quad = \mathrm{Prob}\{d_{12}(k) = 1 \mid d_{12}(k-1) = 1\}\, \mathrm{Prob}\{d_{21}(k) = 1 \mid d_{21}(k-1) = 1\} \\ &\quad = \tilde{p}_{12}(1,1)\tilde{p}_{21}(1,1). \end{aligned} \tag{11}$$

Performing a similar procedure to the remaining entries of the transition matrix yields

$$P = \tilde{P}_{12} \otimes \tilde{P}_{21}, \tag{12}$$

where $\otimes$ denotes the Kronecker product.

3. Convergence analysis

In this section, we study the convergence of $\{y(k)\}$ generated according to Equations (9a)–(9c). We will first derive sufficient conditions under which $\{y(k)\}$ is convergent in the mean and second moment.

An inductive argument shows the solution of Equation (9a) to be

$$y(k+1) = \prod_{i=0}^{k} C(k-i)y(0) + \sum_{i=0}^{k} \prod_{j=i+1}^{k} C(k-j+i+1)\Gamma(i)w(i). \tag{13}$$

We note that $A \otimes A = A_{[2]}$ and $A_{[1]} = A$. Evaluating the Kronecker product of $y(k+1)$ with itself yields

$$\begin{aligned} &y_{[2]}(k+1) \\ &\quad = \prod_{i=0}^{k} C_{[2]}(k-i)y_{[2]}(0) + \sum_{i=0}^{k} \prod_{j=i+1}^{k} C_{[2]}(k-j+i+1)\Gamma_{[2]}(i)w_{[2]}(i) \\ &\quad + \prod_{l=0}^{k} C(k-l)y(0) \otimes \sum_{i=0}^{k} \prod_{j=i+1}^{k} C(k-j+i+1)\Gamma(i)w(i) \\ &\quad + \sum_{i=0}^{k} \prod_{j=i+1}^{k} C(k-j+1)\Gamma(i)w(i) \otimes \prod_{l=0}^{k} C(k-l)y(0) \end{aligned}$$

$$+\sum_{i=0}^{k}\prod_{j=i+1}^{k}C(k-j+i+1)\Gamma(i)w(i)$$
$$\otimes\underbrace{\sum_{m=0}^{k}}_{m\neq i}\prod_{n=m+1}^{k}C(k-n+m+1)\Gamma(m)w(m). \tag{14}$$

In Appendix III, the following formula is derived

$$E\prod_{j=m}^{k}C_{[i]}(k-j+m)$$
$$=\mathscr{I}_i\left(\mathscr{C}_{[i]}(P'\otimes I_{n^i})\right)^{k-m}\mathscr{C}_{[i]}(P'\otimes I_{n^i})^{m}(p_0'\otimes I_{n^i}),\quad \text{for } i=1,2, \tag{15}$$

where

$$\mathscr{I}_i=\underbrace{[I_{n^i}\ I_{n^i}\dots I_{n^i}]}_{l-\text{times}}$$
$$\mathscr{C}_{[i]}=C_{1_{[i]}}\oplus C_{2_{[i]}}\oplus\cdots\oplus C_{l[i]},$$

I_{n^i} is the identity matrix of dimension $n^i\times n^i$, $'$ denotes the transpose and $\oplus$ is the Kronecker sum. Equation (15) is a generalized form of the one contained in Bharucha [4]. The latter evaluates the expected value of the above product of matrices for $m=0$ only.

Let $\|\cdot\|$ be an arbitrary matrix norm and $\rho(A)$ denote the spectral radius of A, defined as

$$\rho(A)\triangleq\max_i|\lambda_i(A)|, \tag{16}$$

where the λ_i's are the eigenvalues of A.

Theorem 1. *Consider the sequence $\{y(k)\}$ generated by Equations (9a)–(9c).*
(a) *Assume that*

(i) $\rho(\mathscr{C}(P'\otimes I_n))<1.$

Then the sequence $\{y(k)\}$ converges in the mean to the origin for every initial condition.
(b) *Assume that*

(i) $\rho\left(\mathscr{C}_{[2]}(P'\otimes I_{n^2})\right)<1,$

(ii) $\sum_{i=0}^{\infty}\|P'\otimes I_{n^2}\|^{i}\|\Gamma_{[2]}(i)\|<\infty.$

Then the sequence $\{y(k)\}$ converges in the second moment to the origin for every initial condition.

Proof of (a). We take expectations on Equation (13). Recalling that $\{w(k)\}$ is a sequence of zero-mean i.i.d. random vectors, independent of $\{C(k)\}$ and that $y(0)$ is independent of $\{C(k)\}$, we obtain

$$Ey(k+1)$$
$$=E\left[\prod_{j=0}^{k}C(k-j)\right]Ey(0)+E\sum_{i=0}^{k}\prod_{j=i+1}^{k}C(k-j+i+1)\Gamma(i)w(i)$$
$$=E\left[\prod_{j=0}^{k}C(k-j)\right]Ey(0)+\sum_{i=0}^{k}E\left[\prod_{j=i+1}^{k}C(k-j+i+1)\Gamma(i)\right]Ew(i). \tag{17}$$

Taking limits on Equation (17) and using the formula of Equation (15) where $i = 1$ and $m = 0$ leads to

$$\begin{aligned}
&\lim_{k\to\infty} Ey(k+1)\\
&\quad = \lim_{k\to\infty} E\prod_{j=0}^{k} C(j)\,Ey(0)\\
&\quad = \lim_{k\to\infty}\mathscr{I}_1(\mathscr{C}(P'\otimes I_n))^k\mathscr{C}(p_0'\otimes I_n)\,Ey(0).
\end{aligned}\tag{18}$$

By virtue of condition a(i), we get the required result as

$$\begin{aligned}
&\lim_{k\to\infty} Ey(k+1)\\
&\quad = \mathscr{I}_1\left[\lim_{k\to\infty}(\mathscr{C}(P'\otimes I_n))^k\right]\mathscr{C}(p_0'\otimes I_n)\,Ey(0)\\
&\quad = 0.
\end{aligned}\tag{19}$$

Proof of (b). We take expectations on Equation (14). Utilizing the final result of Appendix II and Equation (15) yields

$$\begin{aligned}
&Ey_{[2]}(k+1)\\
&\quad = E\left[\prod_{i=0}^{k} C_{[2]}(k-i)\right]Ey_{[2]}(0) + \sum_{i=0}^{k} E\left[\prod_{j=i+1}^{k} C_{[2]}(k-j+i+1)\Gamma_{[2]}(i)\right]Ew_{[2]}\\
&\quad = \mathscr{I}_2(\mathscr{C}_{[2]}(P'\otimes I_{n^2}))^k\mathscr{C}_{[2]}(p_0'\otimes I_2)\,Ey_{[2]}(0)\\
&\quad\quad + \mathscr{I}_2\sum_{i=0}^{k-1}(\mathscr{C}_{[2]}(P'\otimes I_{n^2}))^{k-i}(P'\otimes I_{n^2})^i(p_0'\otimes I_{n^2})\Gamma_{[2]}(i)\,Ew_{[2]}\\
&\quad\quad + \Gamma_{[2]}(k)\,Ew_{[2]}.
\end{aligned}\tag{20}$$

We consider the behavior of the three expressions summed in Equation (20) individually.

(1) *Expression 1*, i.e.

$$\lim_{k\to\infty}\mathscr{I}_2(\mathscr{C}_{[2]}(P'\otimes I_{n^2}))^k\mathscr{C}_{[2]}(p_0'\otimes I_{n^2})\,Ey_{[2]}(0).\tag{21}$$

By invoking a similar line of argument which appears in the proof of part (a), it is shown that using condition b(i) yields

$$\lim_{k\to\infty}\mathscr{I}_2(\mathscr{C}_{[2]}(P'\otimes I_{n^2}))^k\mathscr{C}_{[2]}(p_0'\otimes I_{n^2})\,Ey_{[2]}(0) = 0.\tag{22}$$

(2) *Expression 2*, i.e.

$$\lim_{k\to\infty}\mathscr{I}_2\sum_{i=0}^{k-1}(\mathscr{C}_{[2]}(P'\otimes I_{n^2}))^{k-i}(P'\otimes I_{n^2})^i(p_0'\otimes I_{n^2})\Gamma_{[2]}(i)\,Ew_{[2]}.\tag{23}$$

First, we show that using Conditions b(i) and b(ii) establishes the absolute convergence of

$$\lim_{k\to\infty}\sum_{i=0}^{k-1}(\mathscr{C}_{[2]}(P'\otimes I_{n^2}))^{k-i}(P'\otimes I_{n^2})^i(p_0'\otimes I_{n^2})\Gamma_{[2]}(i).\tag{24}$$

Condition b(i) suggests that

$$\exists M>0 \quad\text{such that}\quad \|(\mathscr{C}_{[2]}(P'\otimes I_{n^2}))^k\| < M,\quad \forall k.\tag{25}$$

Using this fact, we obtain

$$\lim_{k\to\infty}\sum_{i=0}^{k-1}\left\|\left(\mathscr{C}_{[2]}(P'\otimes I_{n^2})\right)^{k-i}(P'\otimes I_{n^2})^i(p_0'\otimes I_{n^2})\Gamma_{[2]}(i)\right\|$$

$$\leq \lim_{k\to\infty}\sum_{i=0}^{k-1}\left\|\left(\mathscr{C}_{[2]}(p'\otimes I_{n^2})\right)^{k-i}\right\|\left\|(P'\otimes I_{n^2})^i(p_0'\otimes I_{n^2})\Gamma_{[2]}(i)\right\|$$

$$< M\|p_0'\otimes I_{n^2}\|\lim_{k\to\infty}\|(P'\otimes I_{n^2})\|^i\|\Gamma_{[2]}(i)\|$$

$$<\infty. \tag{26}$$

The last inequality follows from condition b(ii). Now, we use the $\mathscr{Z}$-domain analysis as a mathematical artifact to show that these same conditions drive Equation (24) to zero, where

$$\mathscr{Z}\{s(k)\} = \sum_{k=-\infty}^{\infty} s(k)z^{-k} \tag{27}$$

and z is a complex variable. Recalling the fact that if $\lim_{k\to\infty}s(k)$ exists then

$$\lim_{k\to\infty} s(k) = \lim_{z\to 1}(1-z^{-1})\mathscr{Z}\{s(k)\}, \tag{28}$$

and that if $h(k-i)=0$ for $i>k$ then

$$\mathscr{Z}\left\{\sum_{i=0}^{\infty}h(k-i)s(i)\right\} = \mathscr{Z}\{h(k)\}\mathscr{Z}\{s(k)\}. \tag{29}$$

In accord with Equations (28) and (29), we write Equation (24) as

$$\lim_{k\to\infty}\sum_{i=0}^{k-1}\left(\mathscr{C}_{[2]}(P'\otimes I_{n^2})\right)^{k-i}(P'\otimes I_{n^2})^i(p_0'\otimes I_{n^2})\Gamma_{[2]}(i)$$

$$= \lim_{z\to 1}(1-z^{-1})\mathscr{Z}\left\{\left(\mathscr{C}_{[2]}(P'\otimes I_{n^2})\right)^i,\ \text{ for } i\geq 0\right\}$$

$$\mathscr{Z}\left\{(P'\otimes I_{n^2})^i(P_0'\otimes I_{n^2})\Gamma_{[2]}(i)\right\}$$

$$= \lim_{z\to 1}\sum_{i=0}^{\infty}\left(\mathscr{C}_{[2]}(P'\otimes I_{n^2})\right)^i z^{-i}$$

$$\lim_{z\to 1}(1-z^{-1})\mathscr{Z}\left\{(P'\otimes I_{n^2})^i(p_0'\otimes I_{n^2})\Gamma_{[2]}(i)\right\} \tag{30}$$

Condition b(i) allows the interchange of the limit and infinite summation in Equation (30). Condition b(ii) implies that

$$\lim_{k\to\infty}(P'\otimes I_{n^2})^k(p_0'\otimes I_{n^2})\Gamma_{[2]}(k) = 0. \tag{31}$$

Returning back to Equation (30), we obtain

$$\lim_{k\to\infty}\sum_{i=0}^{k-1}\left(\mathscr{C}_{[2]}(P'\otimes I_{n^2})\right)^{k-i}(P'\otimes I_{n^2})^i(p_0'\otimes I_{n^2})\Gamma_{[2]}(i)$$

$$= \sum_{i=0}^{\infty}\left(\mathscr{C}_{[2]}(P'\otimes I_{n^2})\right)^i\lim_{k\to\infty}(P'\otimes I_{n^2})^k(p_0'\otimes I_{n^2})\Gamma_{[2]}(k)$$

$$= 0. \tag{32}$$

Equation (23) can be written as

$$\lim_{k\to\infty}\mathscr{I}_2\sum_{i=0}^{k-1}\left(\mathscr{C}_{[2]}(P'\otimes I_{n^2})\right)^{k-i}(P'\otimes I_{n^2})^i(p_0'\otimes I_{n^2})\Gamma_{[2]}(i)Ew_{[2]}$$

$$=\mathscr{I}_2\lim_{k\to\infty}\left[\sum_{i=0}^{k-1}\left(\mathscr{C}_{[2]}(P'\otimes I_{n^2})\right)^{k-i}(P'\otimes I_{n^2})^i(p_0'\otimes I_{n^2})\Gamma_{[2]}(i)\right]Ew_{[2]}$$

$$=0. \tag{33}$$

(3) *Expression* 3, i.e.

$$\lim_{k\to\infty}\Gamma_{[2]}(k)Ew_{[2]}. \tag{34}$$

From the fact that P is a transition matrix of a Markov chain, $(P'\otimes I_{n^2})$ is a matrix of nonnegative entries whose rows sum to one and

$$\| P'\otimes I_{n^2}\| \geq \rho(P'\otimes I_{n^2}) = 1. \tag{35}$$

We write

$$\|\Gamma_{[2]}(k)\| \leq \|P'\otimes I_{n^2}\|^k \|\Gamma_{[2]}(k)\|, \quad \text{for all } k. \tag{36}$$

Utilizing Equation (36) along with condition b(ii) shows the absolute convergence of the series $\sum_{i=0}^{\infty}\Gamma_{[2]}(i)$ and

$$\lim_{k\to\infty}\Gamma_{[2]}(k)Ew_{[2]}=0. \tag{37}$$

We conclude the proof by combining the results obtained from steps (1), (2) and (3). □

3.1 Special case: $p_0 = p_0 P$

In this section, we provide a weaker version of the conditions for convergence of $y(k)$ as the Markov chain $\{C(k)\}$ assumes a special form. In particular, if the initial probability distribution p_0 satisfies the stationarity property in which

$$p_0 = p_0 P, \tag{38}$$

then we write

$$\begin{aligned}\text{Prob}\{C(k)=C_i\} &= (p_0P^k)_i\\ &= p_{0i}\\ &= \text{Prob}\{C(0)=C_i\} \quad \text{for } i=1,\dots,l.\end{aligned} \tag{39}$$

Hence, if the initial probability distribution is a stationary one, then the distribution of the corresponding Markov chain will be invariant as time progresses. Combining Equation (39) with Equation (15) yields

$$E\prod_{j=m}^{k}C_{[i]}(k-j+m)$$

$$=\mathscr{I}_i\left(\mathscr{C}_{[i]}(P'\otimes I_{n^i})\right)^{k-m}\mathscr{C}_{[i]}(p_0'\otimes I_{n^i}), \quad \text{for } i=1,2. \tag{40}$$

Corollary 1. *Consider the sequence $\{y(k)\}$ generated by equations (9a)–(9c). Let $\{C(k)\}$ be a Markov chain whose initial probability distribution is stationary as satisfied by Equation (38). Assume that*

(i) $\rho\left(\mathscr{C}_{[2]}(P'\otimes I_{n^2})\right) < 1,$

(ii) $\sum_{i=0}^{\infty}\|\Gamma_{[2]}(i)\| < \infty.$

Then the sequence $\{y(k)\}$ converges in the second moment to the origin for every initial condition.

Next, we examine the case in which the value of the $\gamma_i(k)$'s is common to all processors. This suggests that Equation (9a) can be written as

$$y(k+1) = C(k)y(k) + \gamma(k)w(k). \tag{41}$$

Corollary 2. *Consider the sequence $\{y(k)\}$ generated by Equation* (41). *Let $\{C(k)\}$ be a Markov chain whose initial probability distribution is stationary as satisfied by Equation* (38). *Assume that*

(i) $\rho\left(\mathscr{C}_{[2]}(P' \otimes I_{n^2})\right) < 1,$

(ii) $\lim_{k\to\infty} \gamma(k) = 0.$

Then the sequence $\{y(k)\}$ converges in the second moment to the origin for every initial condition.

Proof. Using Equation (41) suggests that Equation (20) along with Equation (40) yields

$$\begin{aligned} &Ey_{[2]}(k+1) \\ &= \mathscr{J}_2\left(\mathscr{C}_{[2]}(P' \otimes I_{n^2})\right)^k \mathscr{C}_{[2]}(p_0' \otimes I_{n^2}) Ey_{[2]}(0) \\ &+ \mathscr{J}_2 \sum_{i=1}^{k} \left(\mathscr{C}_{[2]}(P' \otimes I_{n^2})\right)^{k-i} \mathscr{C}_{[2]}(p_0' \otimes I_{n^2}) \gamma^2(i-1) Ew_{[2]} \\ &+ \gamma^2(k) Ew_{[2]}. \end{aligned} \tag{42}$$

The analysis of the behavior of expression 1 summed in Equation (42) remains the same as that of Equation (20). As k gets infinitely large, expression 3 vanishes as a direct consequence of condition (ii). Now, we discuss

$$\lim_{k\to\infty} \mathscr{J}_2 \sum_{i=1}^{k} \left(\mathscr{C}_{[2]}(P' \otimes I_{n^2})\right)^{k-i} \mathscr{C}_{[2]}(p_0' \otimes I_{n^2}) \gamma^2(i-1) Ew_{[2]}.$$

First, we show that using conditions (i) and (ii) establishes the convergence of

$$\lim_{k\to\infty} \sum_{i=1}^{k} \left(\mathscr{C}_{[2]}(P' \otimes I_{n^2})\right)^{k-i} \gamma^2(i). \tag{43}$$

Condition (ii) suggests that

$$\exists M > 0, \quad \text{such that } \gamma^2(k) < M, \quad \forall k. \tag{44}$$

Using Equation (44) and condition (i), we get

$$\begin{aligned} &\lim_{k\to\infty} \sum_{i=1}^{k} \left(\mathscr{C}_{[2]}(P' \otimes I_{n^2})\right)^{k-i} \gamma^2(i) \\ &< M \lim_{k\to\infty} \sum_{i=1}^{k} \left(\mathscr{C}_{[2]}(P' \otimes I_{n^2})\right)^{k-i} \\ &< \infty. \end{aligned} \tag{45}$$

By performing a similar analysis to that used in the proof of part (b) of Theorem 1, we get

$$\lim_{k\to\infty} \mathscr{I}_2 \sum_{i=1}^{k} \left(\mathscr{C}_{[2]}(P' \otimes I_{n^2})\right)^{k-i} \mathscr{C}_{[2]}(p_0' \otimes I_{n^2}) \gamma^2(i-1) E w_{[2]}$$

$$= \mathscr{I}_2 \left[\lim_{k\to\infty} \sum_{i=1}^{k} \left(\mathscr{C}_{[2]}(P' \otimes I_{n^2})\right)^{k-i} \gamma^2(i-1) \right] \mathscr{C}_{[2]}(p_0' \otimes I_{n^2}) E w_{[2]}$$

$$= 0. \quad \square \tag{46}$$

We note that condition (ii) of Corollary 1 is stronger than condition (ii) of Corollary 2.

3.2. Case of ordered scheduling

One of the major advantages of the models with stochastic delays is their ability to maintain the number of transmitted messages under control. This can be achieved by imposing a reasonable restriction on our models where we assume that the information is received in the order it was produced. We assume that each processor i has a local memory where the latest x_i generated at time instant k is kept and when the new information arrives it is labelled using a time stamp as to when it was computed by the other processors. If it happens that this processor acknowledges that the information it receives was generated at a time instant earlier than k, then processor i will discard it. This translates into having the following property

$$\text{Prob}\{d_{ji}(k) > d_{ji}(k-1) + 1\} = 0, \quad \text{for all } j, i \text{ and } k. \tag{47}$$

Consequently, communication delays can be viewed as admitting a particular Markov property with some of the states never visited. Therefore, the transition matrix $\tilde{P}_{ji} = \tilde{p}_{ji}(l, m)$ is such that

$$\tilde{p}_{ji}(l, m) = 0 \quad \text{for } m > l + 1. \tag{48}$$

Convergence conditions of this model can be derived using the results of this section for the particular choice of transition matrix defined in Equation (48).

3.3. Case of independent delays

We let the communication delays be sequences of i.i.d. random variables and provide the corresponding convergence conditions as a specialization of Markov case. For easy reference, we write below Equation (9a) describing this case

$$y(k+1) = C(k) y(k) + \Gamma(k) w(k). \tag{49}$$

Hence, if $\{C(k)\}$ is an i.i.d. matrix sequence and

$$\text{Prob}\{C(k) = C_i\} = p_i, \quad \text{for all } k \text{ and } i = 1, \ldots, l, \tag{50}$$

then it corresponds to a stationary Markov chain of finite state space with transition matrix P of identical rows, i.e. $p_{ij} = p_j$ and whose initial probability distribution is stationary where $p_{0i} = p_i$ for $i = 1, \ldots, l$. Consequently,

$$\mathscr{C}_{[i]}(P' \otimes I_{n^i}) = \begin{bmatrix} C_{1_{[i]}} p_1 & C_{1_{[i]}} p_1 & \cdots & C_{2_{[i]}} p_1 \\ \vdots & & & \vdots \\ C_{l_{[i]}} p_l & C_{l_{[i]}} p_l & \cdots & C_{l_{[i]}} p_l \end{bmatrix}. \tag{51}$$

We define N_i as

$$N_i = Q_i - \mathscr{I}_i, \quad \text{for } i = 1, 2, \tag{52}$$

where

$$Q_i \triangleq \begin{bmatrix} 2I_{n^i} & I_{n^i} & \cdots & I_{n^i} \\ Z & Z & \cdots & Z \\ \vdots & & & \vdots \\ Z & Z & \cdots & Z \end{bmatrix}, \quad \mathscr{I}_i \triangleq I_{n^i} \oplus I_{n^i} \oplus \cdots \oplus I_{n^i}$$

and Z is a matrix of zero entries. It can be verified that $N_i = N_i^{-1}$ by direct multiplication. Using Equation (51), we write

$$N_i^{-1}\left(\mathscr{C}_{[i]}(P' \otimes I_{n^i})\right)N_i = \begin{bmatrix} \sum_{j=1}^{l} C_{j_{[i]}} p_j & Z & \cdots & Z \\ -C_{2_{[i]}} p_2 & Z & \cdots & Z \\ \vdots & & & \vdots \\ -C_{l_{[i]}} p_l & Z & \cdots & Z \end{bmatrix}. \tag{53}$$

The observations extracted from the structure of the matrix on the right-hand side of Equation (53) are twofold. First, the sufficient condition which guarantees convergence in the mean for the case of independent delays to the origin for every initial condition, $\rho(\mathscr{C}(P' \otimes I_n)) < 1$, can be replaced by

$$\rho\left(\sum_{j=1}^{l} C_j p_j\right) < 1.$$

It will be shown that this condition is also necessary for convergence in the mean. Second, the condition $\rho(\mathscr{C}_{[2]}(P' \otimes I_{n^2})) < 1$ of Corollary 1 and Corollary 2 can be replaced by

$$\rho\left(\sum_{j=1}^{l} C_{j_{[2]}} p_j\right) < 1. \tag{54}$$

Corollary 3. *Consider the sequence* $\{y(k)\}$ *generated by Equation* (9*a*). *Let* $\{C(k)\}$ *be an independent, identically distributed matrix sequence. The sequence* $\{y(k)\}$ *converges in the mean to the origin for every initial condition if and only if* $\rho(\Sigma_{j=1}^{l} C_j p_j) < 1$.

Proof. Taking expectations on Equation (9a) and recalling that $\{C(k)\}$ is a matrix sequence independent of $y(k)$ for which it holds that

$$\text{Prob}\{C(k) = C_i\} = p_i, \quad \text{for all } k \text{ and } i = 1, \ldots, l,$$

yields

$$\begin{aligned} Ey(k+1) &= E[C(k)y(k)] \\ &= E[C(k)]Ey(k) \\ &= \left[\sum_{j=1}^{l} C_j p_j\right] Ey(k). \end{aligned} \tag{55}$$

Clearly $Ey(k)$ converges for any $Ey(0)$ if and only if $\rho\Sigma_{j=1}^{l} C_j p_j) < 1$. □

The following are proved using the earlier discussion on the specialization of the Markov chain to the independent case.

Corollary 4. *Consider the sequence* $\{y(k)\}$ *generated by Equation* (9*a*). *Let* $\{C(k)\}$ *be an independent, identically distributed matrix sequence. Assume that*

(i) $$\rho\left(\sum_{j=1}^{l} C_{j_{[2]}} p_j\right) < 1,$$

(ii) $$\sum_{i=0}^{\infty} \| \Gamma_{[2]}(i) \| < \infty.$$

Then the sequence $\{y(k)\}$ *converges in the second moment to the origin for every initial condition.*

Similarly, another corollary is provided below to cover the case of $\{C(k)\}$ being an i.i.d. matrix sequence and Equation (41).

Corollary 5. *Consider the sequence* $\{y(k)\}$ *generated by Equation* (41). *Let* $\{C(k)\}$ *be an independent, identically distributed matrix sequence. Assume that*

(i) $$\rho\left(\sum_{j=1}^{l} C_{j_{[2]}} p_j\right) < 1,$$

(ii) $$\lim_{k \to \infty} \gamma(k) = 0.$$

Then the sequence $\{y(k)\}$ *converges in the second moment to the origin for every initial condition.*

4. Example

Consider a two-processor system where the information from the second processor arrives to the first with a delay such that

$$\begin{aligned} x_1(k+1) &= -\tfrac{2}{3}x_1(k) + \epsilon x_2(k+1-d(k)) \\ x_2(k+1) &= \epsilon x_1(k) + \tfrac{1}{3}x_2(k). \end{aligned} \tag{56}$$

Suppose that

$$d(k) \in \{1, 2\}. \tag{57}$$

We assume that the communication delay $d(k)$ is a stationary Markov chain and let the transition probability matrix P be defined as

$$P = \begin{bmatrix} p & 1-p \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}. \tag{58}$$

Applying state augmentation yields the matrix sequence $\{C(k)\}$ with the two possible matrices

$$C_1 = \begin{bmatrix} -\frac{2}{3} & \epsilon & 0 \\ \epsilon & \frac{1}{3} & 0 \\ 0 & 1 & 0 \end{bmatrix}; \quad C_2 = \begin{bmatrix} -\frac{2}{3} & 0 & \epsilon \\ \epsilon & \frac{1}{3} & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{59}$$

Table 1

p	$\epsilon = 0.6$		$\epsilon = 0.8$		$\epsilon = 1.1$	
	ρ_m	ρ_{sm}	ρ_m	ρ_{sm}	ρ_m	ρ_{sm}
0.0	0.6915	0.4927	0.8299	0.7179	1.0301	1.1098
0.1	0.6909	0.4969	0.8295	0.7261	1.0302	1.1234
0.2	0.6897	0.5050	0.8287	0.7376	1.0304	1.1416
0.3	0.6879	0.5186	0.8275	0.7541	1.0306	1.1669
0.4	0.6855	0.5396	0.8257	0.7780	1.0310	1.2031
0.5	0.6823	0.5710	0.8234	0.8132	1.0316	1.2560
0.6	0.6780	0.6155	0.8203	0.8645	1.0323	1.3327
0.7	0.6721	0.6734	0.8159	0.9349	1.0334	1.4382
0.8	0.7550	0.7421	0.8665	1.0277	1.0671	1.5709
0.9	0.8550	0.8179	0.9938	1.1234	1.2278	1.7241
1.0	0.9447	0.8981	1.1101	1.2322	1.3750	1.8905

Define the spectral radius needed to guarantee mean convergence and the spectral radius needed to guarantee second moment convergence by ρ_m and ρ_{sm}, respectively. We apply Theorem 1 and by examining *Table 1* we observe that for all values of p and for $\epsilon = 0.6$ the system represented by equation (56) converges in the mean and second moment while for $\epsilon = 1.1$ neither type of convergence can be obtained. Nevertheless, for $\epsilon = 0.8$ the variation in p is sufficiently powerful to drive the system in and out of its convergent state. We also note that for $\epsilon = 0.8$ and some values of p, convergence in the mean is the only one of the two types of convergence that is attained.

It is worthy of mention that when the delay $d(k)$ defined by Equation (57) is deterministic and $\epsilon = 0.6$ the system represented by Equation (56) does not fulfill the nonexpansiveness assumption listed in Tseng et al. [14] (Section 3, p. 688), i.e. to asynchronously solve the linear system

$$x = Ax,$$

it must hold that

$$\sum_{j=1} |a_{ij}| \leqslant 1, \quad \text{for all } i. \tag{60}$$

Therefore, while *Table 1* demonstrates the convergence of our asynchronous model with stochastic for $\epsilon = 0.6$, no conclusion can be drawn regarding the convergence of the asynchronous iteration with deterministic delay.

5. Conclusions

We have derived sufficiency conditions for convergence of the general linear model of asynchronous iterations with stochastic delays. These conditions offer an alternative to the usual verification of the nonexpansiveness property contained in [3,14] where the communication delays of the latter works were modeled as unknown but deterministic. The notion of stochastic communication delays appears as a result of the possible failure of some processors, the congestion in a certain network and uneven load conditions. In proving convergence, we utilize the Kronecker product notation that provides a technique by which the computation of the second moments of a linear system with Markov coefficients is achieved. For this purpose, the extended formula (15) is used to accommodate the effect of additive noise. The virtue of this formula is its ability to express the system equation in a proper closed form.

Other types of convergence such as that of almost sure can also be obtained with the aid of the supermartingale machinery. Another issue concerns the extension of the discussed models to the nonlinear case where a nonlinear function appears in the right side of Equations (1) and (4). Towards this goal, the convergence analysis for the nonlinear counterparts of the described models has been carried out in [2].

Appendix

I. Properties of Kronecker products and sums

Let matrix $A = (a_{ij})$ be of dimension $m \times n$ and matrix $B = (b_{ij})$ of dimension $r \times s$. Their Kronecker product is defined as

$$A \otimes B \triangleq (a_{ij}B), \quad \text{for } i = 1, \ldots, m \text{ and } j = 1, \ldots, n$$
$$= \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ \vdots & & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix},$$

of dimension $(mr) \times (ns)$.
The matrix $A \otimes A$ is written as $A_{[2]}$ and $A_{[1]} = A$. The product is associative

$$(A \otimes B) \otimes C = A \otimes (B \otimes C),$$

and obeys the relation

$$A_1B_1 \otimes A_2B_2 \otimes \cdots \otimes A_iB_i = (A_1 \otimes \cdots \otimes A_i)(B_1 \otimes \cdots \otimes B_i)$$

We also define the Kronecker sum $A \oplus B$ as the matrix with A and B diagonal entries and zeros elsewhere, i.e.

$$A \oplus B = \begin{bmatrix} A & Z \\ Z & B \end{bmatrix},$$

where Z is a matrix of all zeros.

The properties of Kronecker products are:
(a) $(A + B) \otimes C = A \otimes C + B \otimes C$
(b) $(A \otimes B)' = A' \otimes B'$, where $'$ denotes transpose.

If A and B are square matrices of dimension a and b respectively,
(c) $A \otimes B = (A \otimes I_b)(I_a \otimes B)$
(d) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
(e) $\text{trace}(A \otimes B) = \text{trace}(A)\,\text{trace}(B)$
(f) $\det(A \otimes B) = \det(A^{a})\det(B^{b})$
(g) $(L_1 \otimes M_1)(A \otimes B)(L_2 \otimes M_2) = L_1AL_2 \otimes M_1BM_2$.

II. Evaluation of $E_{y[2]}(k+1)$

Given

$$y(k+1) = C(k)y(k) + \Gamma(k)w(k), \tag{61}$$

where $\{w(k)\}$ is a sequence of zero-mean i.i.d. random vectors with finite variance, independent of $\{C(k)\}$ and and $\{y(k)\}$. Here we express $Ey_{[2]}(k+1)$. An inductive argument shows

$$y(k+1) = \prod_{i=0}^{k} C(k-i)y(0) + \sum_{i=0}^{k} \prod_{j=i+1}^{k} C(k-j+i+1)\Gamma(i)w(i). \tag{62}$$

Evaluating the Kronecker product of $y(k+1)$ with itself yields

$$\begin{aligned} &y_{[2]}(k+1) \\ &= \prod_{i=0}^{k} C_{[2]}(k-i)\, y_{[2]}(0) + \sum_{i=0}^{k} \prod_{j=i+1}^{k} C_{[2]}(k-j+i+1)\Gamma_{[2]}(i) w_{[2]}(i) \\ &\quad + \prod_{l=0}^{k} C(k-l)\, y(0) \otimes \sum_{i=0}^{k} \prod_{j=i+1}^{k} C(k-j+i+1)\Gamma(i) w(i) \\ &\quad + \sum_{i=0}^{k} \prod_{j=i+1}^{k} C(k-j+i+1)\Gamma(i) w(i) \otimes \prod_{l=0}^{k} C(k-l)\, y(0) \\ &\quad + \sum_{i=0}^{k} \prod_{j=i+1}^{k} C(k-j+i+1)\Gamma(i) w(i) \\ &\qquad \otimes \underbrace{\sum_{m=0}^{k}}_{m \neq i} \prod_{n=m+1}^{k} C(k-n+m+1)\Gamma(m) w(m). \end{aligned} \tag{63}$$

We express Equation (63) in a form which facilitates evaluating expectations.

$$\begin{aligned} &y_{[2]}(k+1) \\ &= \prod_{i=0}^{k} C_{[2]}(k-i)\, y_{[2]}(0) + \sum_{i=0}^{k} \prod_{j=i+1}^{k} C_{[2]}(k-j+i+1)\Gamma_{[2]}(i) w_{[2]}(i) \\ &\quad + \sum_{j=0}^{k} \prod_{i=j+1}^{k} C_{[2]}(k-i+j+1)\left[\prod_{l=0}^{j} C(j-l) \otimes I_n\right](y(0) \otimes \Gamma(j) w(j)) \\ &\quad + \sum_{j=0}^{k} \prod_{i=j+1}^{k} C_{[2]}(k-i+j+1)\left[I_n \otimes \prod_{l=0}^{j} C(j-l)\right](\Gamma(j) w(j) \otimes y(0)) \\ &\quad + \sum_{i=0}^{k-1} \sum_{j=i+1}^{k} \prod_{l=j+1}^{k} C_{[2]}(k-l+j+1)\left[\prod_{m=i+1}^{j} C(j-m+i+1) \otimes I_n\right] \\ &\qquad\qquad (\Gamma(i) w(i) \otimes \Gamma(j) w(j)) \\ &\quad + \sum_{i=1}^{k} \sum_{j=0}^{i-1} \prod_{l=j+1}^{k} C_{[2]}(k-l+i+1)\left[I_n \otimes \prod_{m=j+1}^{i} C(i-m+j+1)\right] \\ &\qquad\qquad (\Gamma(i) w(i) \otimes \Gamma(j) w(j)). \end{aligned} \tag{64}$$

Taking expectations on equation (64), we obtain

$$\begin{aligned} &Ey_{[2]}(k+1) \\ &= E\left[\prod_{i=0}^{k} C_{[2]}(k-i)\right] Ey_{[2]}(0) + \sum_{i=0}^{k} E\left[\prod_{j=i+1}^{k} C_{[2]}(k-j+i+1)\Gamma_{[2]}(i)\right] Ew_{[2]}(i) \\ &\quad + \sum_{j=0}^{k} E \prod_{i=j+1}^{k} C_{[2]}(k-i+j+1)\left[\prod_{l=0}^{j} C(j-l) \otimes I_n\right] E(y(0) \otimes \Gamma(j) w(j)) \end{aligned}$$

$$+\sum_{j=0}^{k} E\prod_{i=j+1}^{k} C_{[2]}(k-i+j+1)\left[I_n \otimes \prod_{l=0}^{j} C(j-l)\right] E(\Gamma(j)w(j)\otimes y(0))$$

$$+\sum_{i=0}^{k-1}\sum_{j=i+1}^{k} E\prod_{l=j+1}^{k} C_{[2]}(k-l+j+1)\left[\prod_{m=i+1}^{j} C(j-m+i+1)\otimes I_n\right]$$

$$E(\Gamma(i)w(i)\otimes\Gamma(j)w(j))$$

$$+\sum_{i=1}^{k}\sum_{j=0}^{i-1} E\prod_{l=j+1}^{k} C_{[2]}(k-l+i+1)\left[I_n\otimes\prod_{m=j+1}^{i} C(i-m+j+1)\right]$$

$$E(\Gamma(i)w(i)\otimes\Gamma(j)w(j)). \tag{65}$$

Recalling that $\{w(k)\}$ is a sequence of i.i.d. zero-mean random vectors, i.e. $Ew(i)=Ew=0$, for all i, we obtain

$$\begin{aligned} &E(y(0)\otimes\Gamma(i)w(i)), \quad \text{for all } i\\ &\quad=[y(0)]\otimes\Gamma(i)Ew(i)\\ &\quad=0 \end{aligned} \tag{66}$$

and

$$\begin{aligned} &E(\Gamma(i)w(i)\otimes\Gamma(j)w(j)), \quad \text{for } i\neq j\\ &\quad=\Gamma(i)Ew(i)\otimes\Gamma(j)Ew(j)\\ &\quad=0. \end{aligned} \tag{67}$$

Combining Equations (66) and (67), we finally obtain

$$\begin{aligned} &Ey_{[2]}(k+1)\\ &\quad=E\left[\prod_{i=0}^{k} C_{[2]}(k-i)\right]Ey_{[2]}(0)+\sum_{i=0}^{k} E\left[\prod_{j=i+1}^{k} C_{[2]}(k-j+i+1)\Gamma_{[2]}(i)\right]Ew_{[2]}. \end{aligned} \tag{68}$$

III. Derivation of $E\prod_{j=m}^{k} C(k-j+m)$

Consider a finite Markov chain with initial distribution row-vector p_0 of nonnegative row entries, where

$$p_{0i}=\text{Prob}\{C(0)=C_i\} \quad \text{and} \sum_{i=1}^{l} p_{0i}=1, \tag{69}$$

and a constant matrix of transition P between the states of constant matrices C_i, $i=1,\ldots,l$. Here we derive

$$E\prod_{j=m}^{k} C(k-j+m). \tag{70}$$

We consider next the case as $m=k-2$.

$$\begin{aligned} &EC(k)C(k-1)C(k-2)\\ &\quad=\sum_{q,r,s} C_sC_rC_q\text{Prob}\{C(k)=C_s,\ C(k-1)=C_r,\ C(k-2)=C_q\}. \end{aligned} \tag{71}$$

The properties of Markov chain suggest

$$\begin{aligned}
&\mathrm{Prob}\{C(k)=C_s,\, C(k-1)=C_r,\, C(k-2)=C_q\}\\
&\quad=\mathrm{Prob}\{C(k)=C_s \mid C(k-1)=C_r,\, C(k-2)=C_q\}\\
&\qquad \mathrm{Prob}\{C(k-1)=C_r \mid C(k-2)=C_q\}\mathrm{Prob}\{C(k-2)=C_q\}\\
&\quad=\mathrm{Prob}\{C(k)=C_s \mid C(k-1)=C_r\}\\
&\qquad \mathrm{Prob}\{C(k-1)=C_r \mid C(k-2)=C_q\}\left(p_0P^{k-2}\right)_q\\
&\quad=p_{rs}p_{qr}\left(p_0P^{k-2}\right)_q.
\end{aligned} \tag{72}$$

Combining Equation (71) and Equation (72) leads to

$$\begin{aligned}
&EC(k)C(k-1)C(k-2)\\
&\quad=\sum_{q,r,s} C_sC_rC_q \,\mathrm{Prob}\{C(k)=C_s,\, C(k-1)=C_r,\, C(k-2)=C_q\}\\
&\quad=\sum_{q,r,s} C_sC_rC_q p_{rs}p_{qr}\left(p_0P^{k-2}\right)_q\\
&\quad=\sum_{q=1}^{l}\sum_{s=1}^{l} C_s\left[\sum_{r=1}^{l} p_{rs}C_rp_{qr}\right]C_q\left(p_0P^{k-2}\right)_q.
\end{aligned} \tag{73}$$

We view

$$u_{sq}=\sum_{r=1}^{l} p_{rs}C_rp_{qr} \tag{74}$$

as the matrix entry in row s and column q of

$$U=(P'\otimes I_n)\mathscr{C}(P'\otimes I_n), \tag{75}$$

where

$$\mathscr{C}=C_1\oplus C_2\oplus\cdots\oplus C_l.$$

Using Equation (74) reduces Equation (73) to

$$\begin{aligned}
&EC(k)C(k-1)C(k-2)\\
&\quad=\sum_{s=1}^{l} C_s\left[\sum_{q=1}^{l} u_{sq}C_q\left(p_0P^{k-2}\right)_q\right].
\end{aligned} \tag{76}$$

We also view

$$w_s=\sum_{q=1}^{l} u_{sq}C_q\left(p_0P^{k-2}\right)_q \tag{77}$$

as the matrix entry in row s of

$$W=U\mathscr{C}\left(p_0P^{k-2}\otimes I_n\right)'. \tag{78}$$

Using Equations (77), (75) and (78), we write

$$\begin{aligned}
&EC(k)C(k-1)C(k-2)\\
&\quad=\sum_{s=1}^{l} C_sw_s\\
&\quad=\mathscr{I}_1\mathscr{C}W\\
&\quad=\mathscr{I}_1\left(\mathscr{C}(P'\otimes I_n)\right)^2\mathscr{C}\left(p_0P^{k-2}\otimes I_n\right)',
\end{aligned} \tag{79}$$

where

$$\mathscr{I}_1 = \underbrace{[I_n \; I_n \; \cdots \; I_n]}_{l\text{-times}}$$

In view of the properties of Kronecker products, the term in the last parenthesis of Equation (79) can be rewritten as

$$\begin{aligned}(p_0 P^{k-2} \otimes I_n)' &= (p_0 P^{k-2})' \otimes I_n \\ &= (P'^{k-2} \otimes I_n)(p_0' \otimes I_n) \\ &= (P' \otimes I_n)^{k-2}(p_0' \otimes I_n).\end{aligned} \tag{80}$$

Combining with Equation (80), we finally obtain

$$EC(k)C(k-1)C(k-2) = \mathscr{I}_1(\mathscr{C}(P' \otimes I_n))^2 \mathscr{C}(P' \otimes I_n)^{k-2}(p_0' \otimes I_n) \tag{81}$$

which can be successively generalized to

$$E\sum_{j=m}^{k} C(k-j+m) = \mathscr{I}_1(\mathscr{C}(P' \otimes I_n))^{k-m} \mathscr{C}(P' \otimes I_n)^{m}(p_0' \otimes I_n). \tag{82}$$

This naturally suggests

$$E\prod_{j=m}^{k} C_{[2]}(k-j+m) = \mathscr{I}_2(\mathscr{C}_{[2]}(P' \otimes I_{n^2}))^{k-m} \mathscr{C}_{[2]}(P' \otimes I_{n^2})^{m}(p_0' \otimes I_{n^2}), \tag{83}$$

where

$$\mathscr{I}_2 = \underbrace{[I_{n^2} \; I_{n^2} \; \cdots \; I_{n^2}]}_{l\text{-times}}$$

$$\mathscr{C}_{[2]} = C_{1_{[2]}} \oplus C_{2_{[2]}} \oplus \cdots \oplus C_{l_{[2]}}.$$

References

[1] G.M. Baudet, Asynchronous iterative methods for multiprocessors. *J. ACM* 25 (1978) 226–244.

[2] B.F. Beidas and G.P. Papavassilpoulos, Asynchronous implementation of optimization algorithms with time drift, Report 91-10-01, University of Southern California, 1991.

[3] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation, Numerical Methods* (Prentice-Hall, Englewood Cliffs, NJ, 1989).

[4] B.H. Bharucha, On the stability of randomly varying systems, Ph.D. Thesis, University of California at Berkeley, 1961.

[5] D. Chazan and W. Miranaker, Chaotic Relaxation, *Lin. Algebra Appl.* 2 (1969) 199–222.

[6] G.R. Grimmett and D.R. Stirzaker, *Probability and Random Processes* (Clarendon Press, Oxford, 1982).

[7] K. Hwang and F.A. Briggs, *Computer Architecture and Parallel Processing* (McGraw-Hill, New York, 1984).

[8] H.T. Kung, Synchronized and asynchronous parallel algorithms for multiprocessors, in J.F. Traub, ed., *Algorithms and Complexity: New Directions and Recent Results* (Academic Press, New York, 1976).

[9] B. Kuo, *Digital Control Systems* (Holt, Rinehart and Winston, New York, 1980).

[10] H.J. Kushner and G. Yin, Asymptotic properties of distributed and communicating stochastic approximation algorithms, *SIAM J. Control Optim.* 25 (1987) 1266–1290.

[11] D. Mitra, Asynchronous relaxations for the numerical solution of differential equations by parallel processors, *SIAM J. Sci. Stat. Comput.* 8 (1987) 43–58.

[12] S.M. Ross, *Stochastic Processes* (Wiley, New York, 1983).

[13] P. Spiteri, Parallel asynchronous algorithms for solving boundary value problems, in: M. Cosnard et al., ed. *Parallel Algorithms and Architectures* (North-Holland, Amsterdam, 1986).

[14] P. Tseng, D.P. Bertsekas and J.N. Tsitsiklis, Partially asynchronous, parallel algorithms for network flow and other problems, *SIAM J. Control Optim.* 28 (1990) 678–710.

[15] J.N. Tsitsiklis, D.M. Bertsekas and M. Athans, Distributed asynchronous deterministic and stochastic gradient optimization algorithms, *IEEE Trans. Automat. Control* AC-31 (1986) 803–12.

[16] J.N. Tsitsiklis and G.D. Stamoulis, On the average communication complexity of asynchronous distributed algorithms, Report LIDS-P-1986, Laboratory for Information and Decision Systems, M.I.T., 1990.

[17] A. Üresin, Asynchronous iterative algorithms for problems with discrete data, Ph.D. Thesis, University of Southern California, Los Angeles, CA, 1990.

[18] A. Üresin, Sufficient conditions for the convergence of analysis of asynchronous iterations, *Parallel Comput.* 10 (1989) 83–92.

[19] A. Üresin and M. Dubois, Parallel asynchronous algorithms for discrete data, *J. ACM* 37 (1990) 588–606.