Asynchronous Systems and Binary Diagonal Random Matrices: A Proof and Convergence Rate

Syed Amaar Ahmad

Wireless@Virginia Tech,

saahmad@vt.edu

Abstract

In a synchronized network of $n$ nodes, each node will update its parameter based on the system state in a given iteration. It is well-known that the updates can converge to a fixed point if the maximum absolute eigenvalue (spectral radius) of the $n \times n$ iterative matrix $F$ is less than one (i.e. $\rho(F) < 1$). However, if only a subset of the nodes update their parameter in an iteration (due to delays or stale feedback) then this effectively renders the spectral radius of the iterative matrix as one. We consider matrices of unit spectral radii generated from $F$ due to random delays in the updates. We show that if each node updates at least once in every $T$ iterations, then the product of the random matrices (joint spectral radius) corresponding to these iterations is less than one. We then use this property to prove convergence of asynchronous iterative systems. Finally, we show that the convergence rate of such a system is $\rho(F)^{(1-(1-\gamma)^T)^n}$, where assuming ergodicity, $\gamma$ is the lowest bound on the probability that a node will update in any given iteration.

I. INTRODUCTION

Iterative systems naturally arise in wireless networks, parallel processors or in game-theoretic applications when multiple independent agents or nodes update their parameters based on their observations of the system. In a synchronized system, all nodes perform these updates in every iteration. Conversely, we define an asynchronous system, where either (i) only a subset of nodes will update in an iteration, or (ii) all nodes may update but the adaptation of some nodes is based on stale system information based on a previous iteration. Either situation may arise due to random feedback delays on receiving the system state information.

The author is with the Department of Electrical and Computer Engineering at Virginia Tech, Blacksburg, VA, USA.
It is well-known that if the spectral radius of the underlying iterative matrix (which determines the interference or interaction between each pair of nodes) is less than one then the synchronized system converges to a fixed point. In the following discussion, we provide a proof for convergence of an asynchronous iterative system. This proof is based on showing that the product of the effective matrices within any $T$ iterations has a spectral radius is less than one even though the spectral radii of individual matrices may be one. Finally, we also show convergence of the system given estimation error as long as (i) the error is independent of the nodal updates, and (ii) the error projects the iterative matrix into a new matrix which still satisfies the inequality on its spectral radius.

II. System Description

Suppose a linear system comprising $n$ nodes which are all independently updating some parameter. Each node updates its parameter $P_i$ based on its received information regarding the state of the system from its vantage point $E_i$. The nodes update their parameters in iterations $k \in \{1, 2, \ldots, \infty\}$. In iteration $k$, we define:

$$E_i(k) = \sum_{j \neq i} g_{i,j} P_j(k),$$

where $g_{i,j}$ captures the cross-talk or interference effect between node $i$ and node $j$ where in general $g_{j,i} \neq g_{i,j}$. The update of each node $i$ is such that:

$$P_i(k+1) = D_i + E_i(k) = D_i + \sum_{j \neq i} g_{i,j} P_j(k).$$

On the other hand $D_i$ represents some fixed parameter particular to node $i$. We let $D$ be $n \times 1$ vector where element $i$ such that $D(i) = D_i$ and define an $n \times n$ matrix $F$ such that its element in row $i$ and column $j$ given by

$$F(i, j) = g_{i,j}.$$ (3)

Note that depending on the system, the $i^{th}$ diagonal entry in the above matrix may be typically be zero (i.e. $F(i, i) = 0$) when there is no auto-feedback for the node $i$. Each node $i$, which is assumed to know its current state in terms of $E_i(k)$, updates its parameter $P_i$ based on this observation. Note that $P_i(k+1)$ denotes its updated value for iteration $k+1$. In vector notation, we can describe the system state alternatively as

$$E(k) = FP(k)$$ (4)
and the adaptations or updates by the nodes equivalently as
\[ P(k + 1) = D + E(k) = D + FP(k). \] 

III. SYNCHRONOUS CONVERGENCE

Let \( I \) denote the identity matrix and \( \rho(Q) \) be the spectral radius of a \( n \times n \) matrix \( Q \).

Lemma III.1. If \( \rho(F) < 1 \), then a unique feasible fixed point for the nodal updates is \( [I + F]\^{-1}D \).

**Proof:** It is well-known that the evolution of the updates in (5) is such that:
\[
P(k + 1) = D - F(D + F(D + F(\cdots P(0)))) = [I + F + F^2 + F^3 + \cdots + (-1)^{k-1}F^{(k-1)}]D + (-1)^kF^kP(0)
\]
\[
\lim_{k \to \infty} P(k + 1) = [I + F]\^{-1}D
\]
where \( P(0) \) represents the initial values chosen by the \( n \) nodes. Since the maximum absolute eigenvalue of \( F \) is less than one then, by definition, the term \( \lim_{k \to \infty} F^kP(0) \) will disappear to an all zeros vector [1, pg. 618, 7.10.10]. Finally, as the converged transmit power vector does not depend on the initial transmit powers, the fixed point above is also unique. 

Lemma III.2. If \( \max_F |\lambda_F| < 1 \) then \( \max_{\lambda_{-F}} |\lambda_{-F}| < 1 \).

**Proof:** By definition \( Fx = \lambda x \) where for a given eigenvalue \( \lambda \), \( x \) is the associated eigenvector. Multiplying the equation by \(-1\) would yield \((-F)x = (-\lambda)x\). Since this does not change the absolute value of \(-\lambda\) hence \( |\lambda_{(-F)}| = |\lambda_F| \).

We know from [2, p 184] that when all the eigenvalues of a square matrix \( Q \) satisfy the condition \( |\lambda_Q| < 1 \) then the matrix series \( I + Q + Q^2 + \cdots = [I-Q]^{-1} \). We can therefore substitute \( Q = -F \) when the spectral radius of \( F \) is less than one, to get \( [I - F + F^2 + \cdots] = [I + F]^{-1} \). Thus, convergence is not affected whether the eigenvalues are positive or negative as long as their absolute value remain bounded by 1.

IV. ASYNCHRONOUS SYSTEMS

Thus far, we have assumed that in every iteration all \( n \) nodes update their values. Now suppose that only a subset of the \( n \) nodes will update in an iteration. Moreover, the subset of updating nodes may change from iteration to iteration.
**Binary Diagonal Random Matrices**

A matrix may be considered **random** if its entries consist of random numbers from some specified distribution [3]. We define a binary diagonal random matrix \( A(k) = \text{diag}[a_1(k), \ldots, a_n(k)] \) as an \( n \times n \) diagonal matrix where \( a_i(k) \in \{1, 0\} \). If node \( i \) updates in iteration \( k \) then \( a_i(k) = 1 \), otherwise \( a_i(k) = 0 \). The probability mass function of the values of \( a_i(k) \), \( \forall i, k \) could be arbitrary.

Given \( A(k) \), the asynchronous iterative system can be described as:

\[
P(k + 1) = [I - A(k)]P(k) + A(k)[D + FP(k)]
\]

where

\[
D(k) = A(k)D
\]

and the modified random iterative matrix \( F(k) \) as

\[
F(k) = A(k)F + [I - A(k)].
\]

Note that \( F(k) \) can be considered a **random matrix** since its entries are based on any any arbitrary delay distribution in the nodes’ update.

In (7), a node updates in iteration \( k \) or maintains its value from the previous iteration \( P_i(k+1) = P_i(k) \). If all nodes update in iteration \( k \) then \( A(k) = I \) and we have \( P(k+1) = D + FP(k) \). Conversely, no node updates in the iteration then \( A(k) = 0.I \) and thus \( P(k+1) = P(k) \) (all other cases being intermediate situations).

**Lemma IV.1.** If \( \rho(F) < 1 \) then \( \rho(F(k)) = 1 \) if \( a_i(k) = 0 \) for any node \( i \).

**Proof:** We are given that \( F(k) = A(k)F + [I - A(k)] \) and the spectral radius of the matrix is such that \( \rho(F) < 1 \). Therefore, if \( a_i(k) = 0 \), then row \( i \) in matrix \( F(k) \) will all have zeros elements except for the diagonal element \( F(k)(i, i) = 1 \). As per Gershgorin circle theorem [1], by definition the matrix \( F(k) \) now has an eigenvalue of 1 (i.e. \( \rho(F(k)) = 1 \)).

In other words, if any node does not update in iteration \( k \) then this renders the spectral radius of matrix \( F(k) \) equal to one.

Next assume that each node updates at least once within any \( T \) consecutive iterations. This constraint can be captured as \( \sum_{t=k-T}^{k} A(t) \geq I \) where the diagonal indicates the total number of times each node has updated between the current iteration \( k \) and the preceding \( T \) iterations (i.e. interval \( t \in \{k, k-1, \ldots, k-T+1\} \)). Such a constraint can be considered as a bound on the random delays in the updates.

**Theorem IV.2.** If \( \rho(F) < 1 \), and \( \sum_{t=k-T}^{k} A(t) \geq I \) then \( \rho \left( \prod_{t=k-T+1}^{k} F(t) \right) < 1. \)
Proof: If the spectral radius of $F$ is less than one (i.e. $|\lambda_F| < 1$) then $\lim_{k \to \infty} F^k = 0.1$ [1, pg. 618]. Recall that $F(t) = A(t)F + [I - A(t)]$ where $k - T + 1 \leq t \leq k$. Next, note that column $i$ of the matrix $\lim_{k \to \infty} F_k^k$ will converge to an all zeros vector if $a_i(t) = A(t)(i,i) = 1$. Conversely, if $a_i(t) = A_t(i,i) = 0$, then column $i$ of $\lim_{k \to \infty} F_k^k$ will not converge to a zeros vector (see [1, pg. 630] for a detailed discussion).

Next consider the constraint that link $i$ adapt within $T$ iterations. If the diagonal entry $A(t)(i,i) = 0$, then there must be some other $A(u)(i,i) = 1$ where $u \in \{t, t+1, \ldots t+T\} : u \neq t$. Thus, column $i$ of the corresponding matrix $\lim_{k \to \infty} F_k^u$ would be an all zeros vector. Consequently, over $T$ iterations, we deduce that

$$\lim_{k \to \infty} F^k \lim_{k \to \infty} F(t) \cdots \lim_{k \to \infty} F(t - T) = 0.$$  \hfill (10)

By definition, the spectral radius of a matrix is less than one if its power taken to infinite results an all zeros matrix (i.e. $\lim_{k \to \infty} (F(t)F(t-1) \cdots F(t-T))^k = 0.1$) [1, pg. 618]. Thus, for any consecutive $T$ iterations, the spectral radius of the corresponding matrix $F(t)F(t-1) \cdots F(t-T)$ is strictly less than one. This also implies that as $k \gg T$,

$$\lim_{k \to \infty} F(k)F(k-1) \cdots F(1) = \lim_{k \to \infty} (F(k) \cdots F(k-T)) (F(k-T-1) \cdots F(k-2T-1)) \cdots = 0.1$$  \hfill (11)

In above, if each node updates at least once in every $T$ consecutive iterations then the product of matrices has a spectral radius less than one. Next, we show convergence of the matrix series in (7) based on Theorem IV.2.

**Theorem IV.3.** If $\rho(F) < 1$, and each node updates at least once in every $T$ consecutive iterations, then the system converges.

**Proof:** We can expand the series in (7) as follows:

$$P(k+1) = D(k) + F(k-1)D(k-1) + F(k)F(k-1)D(k-2) + F(k)F(k-1)F(k-2)D(k-3) + \cdots$$  \hfill (12)

To show convergence of the series in (12), we take the absolute convergence test [4, pg. 181]. That is, an infinite series of real numbers $\sum_{t=1}^{\infty} f_t$ will converge if the absolute of all its terms
\[ \sum_{t}^{\infty} |f_t| \] converges. We first show the convergence of
\[ D(k) + |F(k-1)|D(k-1) + |F(k)|F(k-1)|D(k-2) + |F(k)|F(k-1)F(k-2)|D(k-3) + \cdots \]
where, the terms \( D(k-i) \) can be separated out. Let us now consider a couple of intermediate steps in our proof. Firstly, a series of non-negative real numbers \( \sum_{t}^{\infty} |q_t| \) will converge if a bounding series \( \sum_{t}^{\infty} |q_t| \) such that \(|q_t| \geq |f_t| \geq 0\) converges [4, pg. 180]. Secondly, a matrix series of the type \( I + Q + Q^2 + \cdots + Q^k \) converges if \( \lim_{k \to \infty} Q^k = 0 \cdot I \) [1, pg. 126]. For some finite \( T : T < k \)
\[ \lim_{k \to \infty} |F(k)F(k-1) \cdots F(1)| = 0 \cdot I. \]
Thus, for some arbitrary non-negative valued \( Q \) and \( R \) of dimensions \( n \times n \) and \( n \times 1 \) respectively we will have
\[ \lim_{k \to \infty} R + QR + Q^2R + \cdots Q^kR \geq \lim_{k \to \infty} D(k) + |F(k-1)|D(k-1) + \cdots |F(k)|F(k-1) \cdots F(1)|D(0). \]
Thus this implies that the series in (13) will converge and so will the series in (12) as per the absolute convergence test.

**Estimation Error**

In practical systems, there may be an estimation error that renders node’s knowledge about the system state as imperfect [5]. The updates will thus be based on inaccurate state information. Essentially, the estimation error will project \( E(k) \) into \( \widehat{E_i}(k) = \sum_{j \neq i} \widehat{g}_{ij} \cdot P_j(k) \). Consequently, the matrix of the whole system can be denoted as \( \widehat{F} \) due to the imperfect or faulty state information remains constant over the time interval if the estimates are independent of the parameter updates. In that case, if we have \( \rho(\widehat{F}) < 1 \) the system still convergences as Theorems IV.2 and IV.3 still hold.

**V. RATE OF CONVERGENCE**

We let the probability mass function of the random variables \( a_i(k) \) be based on an ergodic process and as follows:
\[ a_i(k) = \begin{cases} 
0, & p_{ik}, \\
1, & 1 - p_{ik}
\end{cases} \]
where \( p_{ik} \) denotes the probability that node \( i \) will update in iteration \( k \). We define

\[
\gamma = \min_{i,k} p_{ik}
\]

(16)

that denotes the lowest probability of an update by any node over the ergodic process. Therefore, the lower bound on the probability, denoted as \( \lambda \), that in a \( T \) iteration interval, all nodes will update at least once is

\[
\lambda = \text{Prob}\left(\sum_{t=k-T}^{k} A(t) \geq I\right) \geq (1 - (1 - \gamma)^T)^n
\]

(17)

given that the nodes are independent. The rate of convergence \( R \) may be described as the effective spectral radius over the \( T \) iteration interval (i.e. a measure of \( \frac{P(k) - P(k-T)}{T} \)). Formally, it is defined as follows:

\[
R := \frac{\rho(F) \lambda}{T} = \frac{\rho(F)(1 - (1 - \gamma)^T)^n}{T}.
\]

(18)

Alternatively, if it is certain that each node will update in every iterations \( T \), but the randomness is limited to the exact number of updates that each node will perform (i.e. \( I \leq \sum_{t=k-T}^{k} A(t) \leq TI \)), then the convergence rate is simply

\[
R := \frac{\rho(F)}{T}
\]

(19)

REFERENCES

[1] C. Meyer, *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.
[2] P. Lancaster, *Theory of Matrices*. Academic Press, 1969.
[3] G. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*. Cambridge University Press, 2009.
[4] C. C. Pugh, *Real Mathematical Analysis*. Springer, 2002.
[5] M. Medard, “The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel,” *IEEE Trans. Info. Thery*, vol. 3, pp. 933–946, May 2000.