

Chapter 2

Products of Stochastic Matrices and Averaging Dynamics

In this chapter, we introduce and review some of the results on products of stochastic matrices and averaging dynamics. Throughout this thesis, by a product of stochastic matrices, we mean *left product* of stochastic matrices. More precisely, let $\{A(k)\}$ be a stochastic chain. By left product of stochastic matrices, we mean the product of the form $A(k) \cdots A(t_0)$ where $k \geq t_0 \geq 0$ and k often approaches to infinity.

Generally, there are two alternative viewpoints for product of stochastic matrices. The first viewpoint is directly involved with the product itself. The second viewpoint is the dynamic system viewpoint, which is based on the study of the dynamics driven by such a product. In this section, we first discuss the two viewpoints and we show their equivalency. Then, we will present some results that are used in the thesis.

2.1 Averaging Dynamics: Two Viewpoints

Here, we discuss two viewpoints for averaging dynamics and show their equivalency.

2.1.1 Left Product of Stochastic Matrices

The first viewpoint builds on the convergence properties of the left product of stochastic matrices. Suppose that we have a sequence of stochastic matrices $\{A(k)\}$. Let

$$A(k : s) = A(k - 1) \cdots A(s) \quad \text{for } k > s \geq 0,$$

and $A(s : s) = I$. Let us define the concepts of weak and strong ergodicity as appeared in [1].

Definition 2.1 (*Ergodicity* [1]) Let $\{A(k)\}$ be a chain of stochastic matrices. We say that $\{A(k)\}$ is weakly ergodic if $\lim_{k \rightarrow \infty} (A_{i\ell}(k : t_0) - A_{j\ell}(k : t_0)) = 0$ for any $t_0 \geq 0$ and $i, j, \ell \in [m]$.

We say that $\{A(k)\}$ is strongly ergodic, if $\lim_{k \rightarrow \infty} A(k : t_0) = ev^T(t_0)$ for any $t_0 \geq 0$, where $v(t_0)$ is a stochastic vector in $\mathbb{R}^m$.

In words, we say that $\{A(k)\}$ is weakly ergodic if for any *starting time* $t_0 \geq 0$ the difference between any two rows of the product $A(k : t_0)$ goes to zero as k approaches infinity. Likewise, we say that $\{A(k)\}$ is strongly ergodic if for any *starting time* $t_0 \geq 0$, the product $A(k : t_0)$ approaches a stochastic matrix with identical rows, i.e. a rank one stochastic matrix.

Note that in the definition of strong ergodicity the requirement that $v(t_0)$ should be stochastic is redundant. This follows directly from the fact that the ensemble of $m \times m$ stochastic matrices is a closed set in $\mathbb{R}^{m \times m}$, which can be verified by noting that

$$\mathbb{S}_m = \{A \in \mathbb{R}^{m \times m} \mid A \geq 0, Ae = e\}. \quad (2.1)$$

Thus, $\mathbb{S}_m$ is the intersection of two closed sets $\{A \in \mathbb{R}^{m \times m} \mid A \geq 0\}$ and $\{A \in \mathbb{R}^{m \times m} \mid Ae = e\}$, implying that $\mathbb{S}_m$ is a closed subset of $\mathbb{R}^{m \times m}$. Hence, $\lim_{k \rightarrow \infty} A(k : t_0) = ev^T(t_0)$ automatically implies that $v(t_0)$ is a stochastic vector.

Example 2.1 As a simple example for ergodicity, let $\{A(k)\}$ be a chain of stochastic matrices such that $A(k)$ is equal to the rank one stochastic matrix $\frac{1}{m}J$ for infinitely many indices $\dots > k_t > \dots > k_2 > k_1$. Note that for any stochastic matrix A , we have $A[\frac{1}{m}J] = \frac{1}{m}J$. Also, we have $[\frac{1}{m}J]A = \frac{1}{m}ee^T A = e[\frac{1}{m}e^T A]$. Note that $[\frac{1}{m}e^T A]$ is a stochastic vector, since $\frac{1}{m}e^T Ae = \frac{1}{m}e^T e$. Thus, for any $t_0 \geq 0$, if t is large enough such that $k_t > t_0$, then we have

$$A(k : t_0) = A(k : k_t)A(k_t : t_0) = \frac{1}{m}JA(k_t : t_0) = e[\frac{1}{m}e^T A(k_t : t_0)],$$

for all $k > k_t$. This implies that $\lim_{k \rightarrow \infty} A(k : t_0) = ev^T(t_0)$ for any $t_0 \geq 0$, where $v(t_0) = \frac{1}{m}A^T(k_t : t_0)e$. Thus such a chain is strongly ergodic.

Note that if $\lim_{k \rightarrow \infty} A(k : t_0) = ev^T(t_0)$, then we have $\lim_{k \rightarrow \infty} (A_{ij}(k : t_0) - A_{\ell j}(k : t_0)) = v_j(t_0) - v_j(t_0) = 0$. Thus, strong ergodicity implies weak ergodicity. In [1], it is proven that the reverse implication also holds.

Theorem 2.1 ([1], Theorem 1) *Weak ergodicity is equivalent to strong ergodicity.*

Since weak and strong ergodicity are the same concepts, we simply refer to this property as *ergodicity*. Thus, there are two equivalent viewpoints to ergodicity: the first viewpoint is that the product $A(k) \cdots A(t_0)$ converges to a matrix with identical stochastic rows, or alternatively, the difference between any two rows of the product $A(k) \cdots A(t_0)$ converges to zero as k goes infinity, for any choice of starting time t_0 . This is the viewpoint to ergodicity and averaging based on the left product of stochastic matrices.

Another related concept that is based on the limiting behavior of the product of stochastic matrices is *consensus* as defined below.

Definition 2.2 (*Consensus*) We say that a chain $\{A(k)\}$ admits consensus if $\lim_{k \rightarrow \infty} A(k : 0) = ev^T$ for some stochastic vector $v \in \mathbb{R}^m$.

The reason that such property is called *consensus* will be clear once we discuss the averaging dynamics from the dynamic system viewpoint.

Note that a chain $\{A(k)\}$ is ergodic if and only if $\{A(k)\}_{k \geq t_0}$ admits consensus for any $t_0 \geq 0$. Thus ergodicity highly depends on the future (tale) of the chain $\{A(k)\}$. However, this may not be the case for consensus as seen from the following example.

Example 2.2 Consider the chain $\{A(k)\}$ defined by $A(0) = \frac{1}{m}J$ and $A(k) = I$ for $k \geq 1$. Then, for any $k > 0$, we have $A(k : 0) = \frac{1}{m}J$ implying that $\{A(k)\}$ admits consensus. However, note that for any starting time $t_0 \geq 1$, and any $k > t_0$, we have $A(k : t_0) = I$ implying that $\{A(k)\}$ is not ergodic.

Although consensus may not be dependent on the future and, in general, it does not imply ergodicity but under a certain condition the reverse holds.

Lemma 2.1 Let $\{A(k)\}$ be a chain of stochastic matrices and suppose that $A(k)$ is invertible for any $k \geq 0$. Then, $\{A(k)\}$ admits consensus if and only if $\{A(k)\}$ is ergodic.

Proof Note that if $A(k)$ is invertible for any $k \geq 0$, then for any $t_0 \geq 0$, the limit $\lim_{k \rightarrow \infty} A(k : t_0)$ exists if and only if $\lim_{k \rightarrow \infty} A(k : 0)$ exists. This follows from the fact that for $k > t_0$, we have $A(k : t_0) = A(k : 0)A^{-1}(t_0 : 0)$.

Now, if $\{A(k)\}$ admits consensus, then $\lim_{k \rightarrow \infty} A(k : 0)$ is a rank one matrix. Therefore, for any $t_0 \geq 0$, the matrix $[\lim_{k \rightarrow \infty} A(k : t_0)]A(t_0 : 0)$ is a rank one matrix and since $A(t_0 : 0)$ is a full-rank matrix, it implies that $\lim_{k \rightarrow \infty} A(k : t_0)$ is a rank one matrix. But this holds for any $t_0 \geq 0$, implying that $\{A(k)\}$ is ergodic. The reverse implication follows by the definition of ergodicity and consensus. **Q.E.D**

2.1.2 Dynamic System Viewpoint

Here, we discuss the dynamic system viewpoint to the averaging dynamics and we show that it is equivalent to the previously discussed viewpoint.

Let $\{A(k)\}$ be a chain of stochastic matrices. Let $t_0 \geq 0$ and let $x(t_0) \in \mathbb{R}^m$ be arbitrary. Let us define:

$$x(k+1) = A(k)x(k), \quad \text{for } k \geq t_0. \quad (2.2)$$

We say that $\{x(k)\}$ is a dynamics driven by $\{A(k)\}$. We also say that $t_0 \geq 0$ is the *starting time* and $x(t_0) \in \mathbb{R}^m$ is the *starting point* of the dynamics. We refer to $(t_0, x(t_0)) \in \mathbb{Z}^+ \times \mathbb{R}^m$ as an *initial condition* for the dynamics.

A side remark about dynamics (2.2) is that the chain $\{A(k)\}$ may depend on the history of $\{x(k)\}$, i.e. $A(k)$ may be some function of the time k and the history of the dynamics $x(t_0), \dots, x(k)$. In this case, we extract the process $\{B(k)\}$ by letting $B(k) = A(k, x(t_0), \dots, x(k))$ and study the dynamics (2.2) for the chain $\{B(k)\}$ with arbitrary initial condition.

Also, note that any point in the set

$$\mathbf{C} = \{\lambda e \mid \lambda \in \mathbb{R}\},$$

which is the line passing through the all-one vector e , is an equilibrium point for the dynamics (2.2). We refer to this line as the *consensus subspace*.

Now, consider a starting time $t_0 \geq 0$ and a starting point $x(t_0) \in \mathbb{R}^m$, and consider the corresponding dynamics $\{x(k)\}$ driven by a stochastic chain $\{A(k)\}$. At each time $k \geq t_0$, we have $x_i(k+1) = \sum_{j=1}^m A_{ij}(k)x_j(k)$. But $A_i(k)$ is a stochastic vector and hence, $x_i(k+1)$ is simply a *weighted average* of the scalars $\{x_1(k), \dots, x_m(k)\}$. Thus, the m coordinates of the vector $x(k+1)$ are nothing but m weighted averages of the coordinates of $x(k)$. This motivates the name *weighted averaging dynamics* for the dynamics. Based on this observation, an intuitive way of describing dynamics (2.2) is to consider the set $[m]$ as a set of m agents and let $x_i(t_0) \in \mathbb{R}$ to be a scalar representing the initial opinion of the i th agent about an issue. Then, at each time $k \geq t_0$, agents share their opinions and agent i 's opinion will evolve by averaging the observed opinions at time k . Although such a model of opinion dynamics among a set of agents is hypothetical, we refer to this interpretation of the dynamics (2.2) as the *opinion dynamics* viewpoint to dynamics (2.2).

For any dynamics $\{x(k)\}$ and any $k \geq t_0$, we have $x(k) = A(k : t_0)x(t_0)$. Therefore, if $\{A(k)\}$ is an ergodic chain (as defined in Definition 2.1), for any initial condition $(t_0, x(t_0)) \in \mathbb{Z}^+ \times \mathbb{R}^m$, we have:

$$\lim_{k \rightarrow \infty} x(k) = \lim_{k \rightarrow \infty} A(k : t_0)x(t_0) = ev^T(t_0)x(t_0) = c(t_0)e,$$

where $c(t_0) = v^T(t_0)x(t_0)$ is a constant depending on the initial condition. Thus, if $\{A(k)\}$ is an ergodic chain, then for any initial condition $(t_0, x(t_0)) \in \mathbb{Z}^+ \times \mathbb{R}^m$, any dynamics $\{x(k)\}$ will converge to some point in the consensus subspace $\mathbf{C}$. This implies that $\lim_{k \rightarrow \infty} (x_i(k) - x_j(k)) = 0$ for any $i, j \in [m]$ and any initial condition $(t_0, x(t_0)) \in \mathbb{Z}^+ \times \mathbb{R}^m$. In fact, the reverse implication also holds.

Theorem 2.2 *A stochastic chain $\{A(k)\}$ is ergodic if and only if $\lim_{k \rightarrow \infty} (x_i(k) - x_j(k)) = 0$ for every initial condition $(t_0, x(t_0)) \in \mathbb{Z}^+ \times \mathbb{R}^m$ and all $i, j \in [m]$.*

Furthermore, for ergodicity, it suffices that $\lim_{k \rightarrow \infty} (x_i(k) - x_j(k)) = 0$ for any starting time $t_0 \geq 0$ and $x(t_0) = e_\ell$ for all $\ell \in [m]$.

Proof The fact that the ergodicity of $\{A(k)\}$ implies $\lim_{k \rightarrow \infty} (x_i(k) - x_j(k)) = 0$ for any initial condition $(t_0, x(t_0)) \in \mathbb{Z}^+ \times \mathbb{R}^m$ and all $i, j \in [m]$ follows from the above discussion. For the converse implication, suppose that $\lim_{k \rightarrow \infty} (x_i(k) - x_j(k)) = 0$ for all $i, j \in [m]$, and every starting time $t_0 \geq 0$ and any starting point $x(t_0) = e_\ell$

where $\ell \in [m]$. For such a starting time, we have $x(k) = A(k : t_0)e_\ell = A^\ell(k : t_0)$ and hence, $x_i(k) = A_{i\ell}(k : t_0)$ and $x_j(k) = A_{j\ell}(k : t_0)$. Therefore, $\lim_{k \rightarrow \infty} (x_i(k) - x_j(k)) = 0$ for $x(t_0) = e_\ell$ if and only if $\lim_{k \rightarrow \infty} (A_{i\ell}(k : t_0) - A_{j\ell}(k : t_0)) = 0$. Thus, if $\lim_{k \rightarrow \infty} (x_i(k) - x_j(k)) = 0$ for any starting time $t_0 \geq 0$ and any starting point of the form $x(t_0) = e_\ell$ for $\ell \in [m]$, then $\{A(k)\}$ is an ergodic chain. **Q.E.D**

Using a similar argument, the following result follows immediately.

Theorem 2.3 *A stochastic chain $\{A(k)\}$ admits consensus if and only if*

$$\lim_{k \rightarrow \infty} (x_i(k) - x_j(k)) = 0$$

for the starting time $t_0 = 0$ and every starting point $x(0) \in \mathbb{R}^m$.

Furthermore, to show that $\{A(k)\}$ admits consensus, it suffices that

$$\lim_{k \rightarrow \infty} (x_i(k) - x_j(k)) = 0$$

for all starting points $x(0) = e_\ell$ with $\ell \in [m]$.

This result explains why the property defined in Definition 2.2 is referred to as consensus. Basically, admitting consensus means that for any initial opinion of the set of agents at time 0, the dynamics (2.2) will lead to consensus, i.e. the difference $x_i(k) - x_j(k)$ between the opinions of any two agents $i, j \in [m]$ goes to zero as k goes to infinity.

2.1.3 Uniformly Bounded and B-Connected Chains

For the study of the averaging dynamics, it is often assumed that the positive entries of a stochastic chain $\{A(k)\}$ are bounded below by some positive scalar $\gamma > 0$ which makes the study of those chains more convenient. We refer to this property as uniform boundedness property.

Definition 2.3 We say that a stochastic chain $\{A(k)\}$ is uniformly bounded if there exist a scalar $\gamma > 0$ for which $A_{ij}(k) \geq \gamma$ for all $i, j \in [m]$ and $k \geq 0$ such that $A_{ij}(k) > 0$.

Some consensus and ergodicity results for deterministic weighted averaging dynamics rely on the existence of a periodical connectivity of the graphs associated with the matrices. We refer to this property as B -connectedness property.

Definition 2.4 For a stochastic chain $\{A(k)\}$, let $G(k) = ([m], \mathcal{E}(k))$ be the graph induced by the positive entries of $A(k)$ for any $k \geq 0$. For $B \geq 1$, we say $\{A(k)\}$ is a B -connected chain if

- (a) $\{A(k)\}$ is uniformly bounded,
- (b) for any time $k \geq 0$ and $i \in [m]$, we have $A_{ii}(k) > 0$, and

(c) the graph:

$$([m], \mathcal{E}(Bk) \cup \mathcal{E}(Bk + 1) \cup \dots \cup \mathcal{E}(B(k + 1) - 1)),$$

is strongly connected for any $k \geq 0$.

The following result shows that a B -connected chain is ergodic. Furthermore, using the result, we can provide some bounds on the limiting matrix $\lim_{k \rightarrow \infty} A(k : t_0)$.

Theorem 2.4 ([2], Lemma 5.2.1) *Let $\{A(k)\}$ be a B -connected chain. Then, the following results hold:*

- (a) $\{A(k)\}$ is ergodic, i.e. for any starting time $t_0 \geq 0$, we have $\lim_{k \rightarrow \infty} A(k : t_0) = ev^T(t_0)$ for a stochastic vector $v(t_0) \in \mathbb{R}^m$.
- (b) There is a constant $\eta \geq \gamma^{(m-1)B}$, which is independent of t_0 , such that $v(t_0) \geq \eta$ for all $t_0 \geq 0$.
- (c) We have

$$\max_{i,j} |A_{ij}(k : t_0) - [ev^T(t_0)]_{ij}| \leq q\mu^{k-t_0},$$

where $\mu \in (0, 1)$ and $q \in \mathbb{R}^+$ are some constants not depending on t_0 .

2.1.4 Birkhoff-von Neumann Theorem

Consider the set of doubly stochastic matrices

$$\mathcal{D} = \{A \in \mathbb{R}^{m \times m} \mid A \geq 0, Ae = A^T e = e\}. \quad (2.3)$$

The given description in Eq. (2.3) clearly shows that the set of doubly stochastic matrices is a polyhedral set in $\mathbb{R}^{m \times m}$. On the other hand, a permutation matrix P is an extreme point of this set, i.e. we cannot write $P = \epsilon A + (1 - \epsilon)B$ for some distinct $A, B \in \mathcal{D}$ and $\epsilon \in (0, 1)$.

The Birkhoff-von Neumann theorem asserts that, in fact, permutation matrices are the only extreme points of $\mathcal{D}$.

Theorem 2.5 (Birkhoff-von Neumann Theorem [3], p. 527) *Let A be a doubly stochastic matrix. Then, A can be written as a convex combination of the permutation matrices, i.e. there exists scalars $q_1, \dots, q_{m!} \in \mathbb{R}^+$ such that*

$$A = \sum_{\xi=1}^{m!} q_{\xi} P^{(\xi)},$$

where $\sum_{\xi \in [m!]} q_{\xi} = 1$.

We use this result in Chap. 6 to provide an alternative characterization of ergodicity for doubly stochastic chains.

References

1. S. Chatterjee, E. Seneta, Towards consensus: Some convergence theorems on re-peated averaging. *J. Appl. Probab.* **14**(1), 89–97 (1977)
2. J. Tsitsiklis, Problems in decentralized decision making and computation, Ph.D. dissertation, Department of Electrical Engineering and Computer Science, MIT, 1984
3. R.A. Horn, C.R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1985)

Product of Random Stochastic Matrices and Distributed
Averaging

Touri, B.

2012, XIV, 142 p., Hardcover

ISBN: 978-3-642-28002-3