A vector almost-supermartingale convergence theorem and its applications

Aditya Mahajan, Silviu-Iulian Niculescu, and Mathukumalli Vidyasagar

***Abstract*— The almost-supermartingale convergence theorem of Robbins and Siegmund (1971) is a fundamental tool for establishing the convergence of various stochastic iterative algorithms including system identification, adaptive control, and reinforcement learning. The theorem is stated for non-negative scalar valued stochastic processes. In this paper, we generalize the theorem to non-negative vector valued stochastic processes and provide two set of sufficient conditions for such processes to converge almost surely. We present several applications of vector almost-supermartingale convergence theorem, including convergence of autoregressive supermartingales, delayed supermartingales, and stochastic approximation with delayed updates.**

I. Introduction

Stochastic iterative algorithms, where one starts with an initial guess $\theta_0 \in \mathbb{R}^p$ and recursively updates it based on the outcome of a stochastic experiment, arise in a variety of applications. Examples include recursive least square methods in system identification [1]–[3], certainty equivalent methods in adaptive control [1], [4], [5], stochastic gradient descent methods in machine learning [6], [7], various learning algorithms such as Q-learning in reinforcement learning [8], [9], and distributed consensus [10], [11]. Broadly speaking, two methodologies are used to analyze the convergence of such algorithms: the martingale approach [12]–[15] and the ODE (ordinary differential equation) approach [16]–[20]. We refer the reader to [4], [21]–[23] for an overview.

The focus of this paper is a fundamental result used in the martingale approach. For the sake of completeness, we start with a brief overview of supermartingales. See [24], [25] for a detailed treatment and see [26] for an introduction of the use of martingale theory in Systems and Control as well as some of the earliest results in the area. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A filtration $\{\mathcal{F}_t\}_{t\ge1}$ is an increasing family of sub-sigma-algebras of $\mathcal{F}$, i.e., for any $s < t$, $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$. A family of integrable random variables $\{X_t\}_{t\ge1}$ is said to be adapted with respect to the filtration $\{\mathcal{F}_t\}_{t\ge1}$ if X_t if $\mathcal{F}_t$-measurable for each t. Such a family is called a **supermartingale** if

$$\mathbb{E}_t[X_{t+1}] \le X_t, \quad \forall t \ge 1$$

where the notation $\mathbb{E}_t[\cdot]$ is a shorthand for $\mathbb{E}[\,\cdot \mid \mathcal{F}_t]$. Roughly speaking, supermartingales generalize the notion of monotonically decreasing sequences to stochastic processes, see, e.g., [27]. A fundamental result in martingale theory is the **supermartingale convergence theorem** [24], [25], which states that non-negative supermartingales converge almost surely.

When analyzing convergence of stochastic iterative algorithms, it is not always convenient to directly apply the supermartingale convergence theorem and a slight generalization, initially proposed by Robins and Siegmund [28], is much more useful:

Theorem 1 (Almost supermartingale convergence) *Suppose $\{X_t\}_{t\ge1}$, $\{\beta_t\}_{t\ge1}$, $\{Y_t\}_{t\ge1}$, $\{Z_t\}_{t\ge1}$ are $\mathbb{R}_+$-valued stochastic processes adapted to some filtration $\{\mathcal{F}_t\}_{t\ge1}$ that satisfy*

$$\mathbb{E}_t[X_{t+1}] \le (1+\beta_t)X_t + Y_t - Z_t, \quad t \ge 1 \tag{1}$$

Define the set Ω_0 by

$$\Omega_0 = \Big\{\omega \in \Omega : \sum_{t\ge1} \beta_t(\omega) < \infty\Big\} \cap \Big\{\omega : \sum_{t\ge1} Y_t(\omega) < \infty\Big\}.$$

Then, for all $\omega \in \Omega_0$, we have that

1) *$\lim_{t\to\infty} X_t(\omega)$ exists and is finite.*
2) *$\sum_{t\ge1} Z_t(\omega) < \infty$.*

Note that all processes in Theorem 1 are $\mathbb{R}_+$-valued. In this paper, we investigate variation of Theorem 1 for general $\mathbb{R}^p_+$-valued processes. Such a generalization is trivial for the standard supermartingale convergence theorem (which is the special case of Theorem 1 when the $\{Y_t\}_{t\ge1}$ and $\{Z_t\}_{t\ge1}$ processes are identically zero) because we can use the scalar supermartingale convergence theorem to argue that each component of the vector-valued process $\{X_t\}_{t\ge1}$ converges. However, such simple arguments do not work in the almost-supermartingale case.

In this paper, we present two sufficient conditions under which the vector-valued almost supermartingales converge. Our conditions rely on the existence of the limit of infinite product of matrices. So, we review some basic results of infinite product of matrices in Sec. II and then present the two main convergence theorems in Sec. III. We then present application of these theorems in Sec. IV to provide sufficient conditions for convergence of auto-regressive supermartingales, delayed supermartingales, and delayed temporal difference learning.

Aditya Mahajan is with the department of Electrical and Computer Engineering at McGill University, Montreal, QC H3A 0E9 Canada. Email: aditya.mahajan@mcgill.ca

Silviu-Iulian Niculescu is with University of Paris-Saclay, CNRS, CentraleSupelec, Inria Saclay, Laboratory of Signals and Systems (L2S), 91190 Gif-sur-Yvette, France. Email: silviu.niculescu@centralesupelec.fr

Mathukumalli Vidyasagar is with the department of Artificial Intelligence at the Indian Institute of Technology Hyderabad, Kandi, Telangana 502285, India. Email: m.vidyasagar@ee.iith.ac.in

The work of Aditya Mahajan was supported by DATAIA Fellowship and Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN-2021-0351. The work of Mathukumalli Vidyasagar was supported by the Science and Engineering Research Board of India.

II. Review of infinite product of matrices

We start with a review of infinite product of matrices. If $\{B_n\}_{n\geq 1}$ are $p\times p$ real matrices, we define

$$\prod_{k=n}^{m} B_k = \begin{cases} B_m B_{m-1}\cdots B_n & \text{if } n\leq m,\\ I & \text{if } n>m;\end{cases}$$

to be the product where successive terms multiply on the left.

We say that an infinite product $\prod_{n=1}^{\infty} B_n$ of $p\times p$ matrices converges if there exists an integer N such that

$$Q=\lim_{m\to\infty}\prod_{n=N}^{m} B_n$$

exists. In this case, we define $\prod_{n=1}^{\infty} B_n = Q\prod_{n=1}^{N-1} B_n$.

Following [29], we say that the product $\prod_{n=1}^{\infty} B_n$ converges *invertibly* if for all $n\geq N$, B_n is invertible and the product Q defined above is invertible as well. It was argued in [29] that the above definition has the following consequences:

(P1) An invertibly convergent infinite product is singular if and only if at least one of its factors is singular.

(P2) If $\prod_{n=1}^{\infty} B_n$ converges invertibly then $\lim_{n\to\infty} B_n = I$.

We now present some sufficient conditions for convergence of infinite product of matrices. If the matrices being multiplied are invertible, then the product converges to an invertible limit. Let $\|\cdot\|$ denote any sub-multiplicative matrix norm on $\mathbb{R}^{p\times p}$

(C1) If

$$\sum_{n=1}^{\infty}\|B_n\|<\infty$$

then $\prod_{n=1}^{\infty}(I+B_n)$ converges.

(C2) Let $\{R_n\}_{n\geq 1}$ be a sequence of $p\times p$ matrices such that

$$\lim_{n\to\infty} R_n = I$$

and

$$\sum_{n=1}^{\infty}\|(I+B_n)R_n - R_{n+1}\|<\infty$$

then $\prod_{n=1}^{\infty}(I+B_n)$ converges.

(C3) Let $\{U_n\}_{n\geq 1}$ be a sequence of *invertible* $p\times p$ matrices such that $\|U_n\|=1$ for all n, there exists a N such that $\prod_{n=N}^{m} U_n$ converges for $m\to\infty$, and

$$\sum_{n=1}^{\infty}\|B_n\|<\infty$$

then $\prod_{n=1}^{\infty}(U_n+B_n)$ converges.

Condition (C1) is a standard result and stated as Theorem 1 in [29]. Condition (C2) is Theorem 5 of [29]. Condition (C3) is Theorem 2.1 combined with the remark on page 15 of [30].

III. A vector-valued generalization of the almost supermartingale convergence theorem

Suppose $\{X_t\}_{t\geq 1}$, $\{Y_t\}_{t\geq 1}$, and $\{Z_t\}_{t\geq 1}$ are $\mathbb{R}_+^p$-valued stochastic processes and $\{A_t\}_{t\geq 1}$ is a $\mathbb{R}_+^{p\times p}$-valued stochastic process, all adapted to some filtration $\{\mathcal{F}_t\}_{t\geq 1}$ that satisfy

$$\mathbb{E}_t[X_{t+1}]\leq A_t X_t + Y_t - Z_t,\quad t\geq 1. \tag{2}$$

We assume that there exists a deterministic sequence $\{\bar{A}_t\}_{t\geq 1}$, $\bar{A}_t\in\mathbb{R}_+^{p\times p}$ that satisfies the following properties:

(A1) For each t, $A_t\leq \bar{A}_t$ almost surely.

(A2) For each t, $\bar{A}_t$ is invertible

(A3) The infinite product

$$M := \prod_{t=1}^{\infty}\bar{A}_t$$

converges invertibly. Clearly, $M\in\mathbb{R}_+^{p\times p}$.

Remark 1 The sufficient conditions (C1)–(C3) mentioned above for existence of infinite product of matrices can be used to verify (A3). In particular, if we assume that $A_t = I + B_t$ where $\sum_{t\geq 1}\|B_t\|<\infty$, we have a natural generalization of the conditions in Theorem 1 to the vector case. However, condition (A3) is more general. For example, as shown in [29], $\prod_{t\geq}(I+B_t)$ converges even if $\sum_{t\geq 1}\|B_t\|=\infty$ if condition (C2) holds.

Define

$$\Psi_t := \prod_{s\geq t}\bar{A}_s. \tag{3}$$

It follows from (A2) and (A3) that Ψ_t is well defined for all t because

$$\Psi_t = M\Bigg[\prod_{t=1}^{\tau-1}\bar{A}_\tau\Bigg]^{-1}. \tag{4}$$

It is also clear that for all $t\geq 1$, $\Psi_t\in\mathbb{R}_+^{p\times p}$, and

$$\Psi_{t+1}\bar{A}_t = \Psi_t. \tag{5}$$

Theorem 2 *Suppose (A1)–(A3) hold. Define the set*

$$\Omega_0 = \Bigg\{\omega\in\Omega : \sum_{t\geq 1} Y_t(\omega)<\infty\Bigg\}.$$

Then, for all $\omega\in\Omega_0$, we have that

1) $\lim_{t\to\infty} X_t(\omega)$ *exists and is finite.*
2) $\sum_{t\geq 1}\Psi_{t+1}Z_t(\omega)<\infty$.

Proof Observe that since all processes are non-negative, Assumption (A1) implies that we can replace (2) by the following:

$$\mathbb{E}_t[X_{t+1}]\leq \bar{A}_t X_t + Y_t - Z_t,\quad t\geq 1. \tag{6}$$

Define $X'_t = \Psi_t X_t$, $Y'_t = \Psi_{t+1} Y_t$, and $Z'_t = \Psi_{t+1} Z_t$. Now consider

$$\begin{aligned}
\mathbb{E}_t[X'_{t+1}] &= \mathbb{E}_t[\Psi_{t+1} X_{t+1}] \\
&\overset{(a)}{=} \Psi_{t+1} \mathbb{E}_t[X_{t+1}] \\
&\overset{(b)}{\leq} \Psi_{t+1}\big(\bar{A}_t X_t + Y_t - Z_t\big) \\
&\overset{(c)}{=} X'_t + Y'_t - Z'_t
\end{aligned} \tag{7}$$

where (a) follows because Ψ_{t+1} is deterministic, (b) follows from (6), and (c) follows from (5) and the definition of X'_t, Y'_t, and Z'_t. Note that (7) holds for every component.

Define

$$\Omega'_0 = \left\{ \omega \in \Omega : \sum_{t \geq 1} \|\Psi_{t+1} Y_t(\omega)\| < \infty \right\}$$

(where $\|\cdot\|$ is any norm on $\mathbb{R}^p$). Since all norms are equivalent in a finite-dimensional space, we have that for any $\omega \in \Omega'_0$, $\sum_{t \geq 1} [Y'_t]_i < \infty$, for all $i \in \{1, \dots, p\}$, where $[Y'_t]_i$ denotes the i-th component of Y'_t. Therefore, by applying Theorem 1 to every component of (7), we get that for any $\omega \in \Omega'_0$, $\lim_{t\to\infty} X'_t(\omega)$ exists and $\sum_{t \geq 1} Z'_t(\omega) < \infty$.

Now we will show that convergence of X'_t implies convergence of X_t. For $\omega \in \Omega'_0$, define

$$\underline{X}_\infty(\omega) := \liminf_{t\to\infty} X_t(\omega) \quad \text{and} \quad \overline{X}_\infty(\omega) := \limsup_{t\to\infty} X_t(\omega).$$

Let $\{i_k\}_{k \geq 0}$ be a subsequence such that $X_{i_k}(\omega)$ converges to $\underline{X}_\infty(\omega)$ and let $\{j_k\}_{k \geq 0}$ be a subsequence such that $X_{j_k}(\omega)$ converges to $\overline{X}_\infty(\omega)$. Then,

$$\lim_{k\to\infty} \Psi_{i_k} X_{i_k}(\omega) = \lim_{k\to\infty} \Psi_{i_k} \lim_{k\to\infty} X_{i_k} = \underline{X}_\infty(\omega) \tag{8}$$

and

$$\lim_{k\to\infty} \Psi_{j_k} X_{j_k}(\omega) = \lim_{k\to\infty} \Psi_{j_k} \lim_{k\to\infty} X_{j_k} = \overline{X}_\infty(\omega) \tag{9}$$

because $\lim_{t\to\infty} \Psi_t = I$ (which follows from (3)). However, the left hand sides of (8) and (9) are the same because we have shown that $\Psi_t X_t = X'_t$ converges to a limit. Therefore, $\underline{X}_\infty(\omega) = \overline{X}_\infty(\omega)$. Consequently, $\{X_t(\omega)\}_{t \geq 1}$ converges to a limit and that limit is the same as the limit of $\{X'_t(\omega)\}_{t \geq 1}$.

So, we have shown that the result holds for all $\omega \in \Omega'_0$. We will now show that $\Omega_0 \subseteq \Omega'_0$. For any $\omega \in \Omega_0$, we have that $\sum_{t \geq 1} \|Y_t(\omega)\|_\infty < \infty$. Due to the equivalence of norms in a finite dimensional space, this implies that

$$\sum_{t \geq 1} \|Y_t(\omega)\| < \infty. \tag{10}$$

From (A3), it follows that $\Psi^* := \sup_{t \geq 1} \|\Psi_t\| < \infty$. Then, for any $\omega \in \Omega_0$,

$$\sum_{t \geq 1} \|\Psi_{t+1} Y_t(\omega)\| \leq \Psi^* \sum_{t \geq 1} \|Y_t(\omega)\| < \infty$$

where the first inequality follows from the definition of Ψ^* and the second inequality follows from (10). Thus, $\Omega_0 \subseteq \Omega'_0$. Hence, the result holds for all $\omega \in \Omega_0$. ■

We now present a slightly different set of assumptions under which (2) converges. In particular, assume that the sequence $\{A_t\}_{t \geq 1}$ satisfies the following assumptions:

(A4) For each t, A_t is inverible.
(A5) For each t, $A_t^{-1} \in \mathbb{R}_+^{p \times p}$.

Remark 2 A necessary and sufficient condition for (A5) is that each A_t is *monomial*, i.e., for every t there exists a permutation matrix P_t and a diagonal matrix D_t with strictly positive diagonal elements such that $A_t = P_t D_t$.

Define

$$\Phi_t := \prod_{s \leq t} A_t.$$

It follows from (A4) that Φ_t is invertible and from (A5) that $\Phi_t \in \mathbb{R}_+^{p \times p}$. It is also clear that for all $t \geq 1$,

$$\Phi_{t+1}^{-1} A_t = \Phi_t^{-1}. \tag{11}$$

Theorem 3 *Suppose assumptions (A4) and (A5) hold. Define*

$$\Omega_0 = \left\{ \omega : \lim_{t\to\infty} \Phi_t(\omega) < \infty \right\} \cap \left\{ \omega : \sum_{t \geq 1} Y_t(\omega) < \infty \right\}. \tag{12}$$

Then, for all $\omega \in \Omega_0$, we have that

1) $\lim_{t\to\infty} X_t(\omega)$ *exists and is finite.*
2) $\sum_{t \geq 1} \Phi_{t+1}^{-1}(\omega) Z_t(\omega) < \infty$.

PROOF Define $X'_t = \Phi_t^{-1} X_t$, $Y'_t = \Phi_{t+1}^{-1} Y_t$, $Z'_t = \Phi_{t+1}^{-1} Z_t$. Then,

$$\begin{aligned}
\mathbb{E}_t[X'_{t+1}] &\overset{(a)}{=} \Phi_{t+1}^{-1} \mathbb{E}_t[X_{t+1}] \\
&\overset{(b)}{\leq} \Phi_{t+1}^{-1}[A_t X_t + Y_t - Z_t] \\
&\overset{(c)}{=} \Phi_t^{-1} X_t + \Phi_{t+1}^{-1} Y_t + \Phi_{t+1}^{-1} Z_t \\
&= X'_t + Y'_t - Z'_t,
\end{aligned} \tag{13}$$

where (a) follows because Φ_t is $\mathcal{F}_t$-measurable, (b) follows from (2) and $\Phi_t \in \mathbb{R}_+^{p \times p}$ and (c) follows from (11). Inequality (13) holds for each component of the vectors. Therefore, by applying Theorem 1 on each component, we get that $\lim_{t\to\infty} X'_t(\omega)$ and $\sum_{t \geq 1} Z'_t(\omega)$ exist and are finite on the set $\{\omega : \sum_{t \geq 1} Y'_t(\omega) < \infty\}$.

Now, define the set

$$\Omega'_0 = \left\{ \omega : \lim_{t\to\infty} \Phi_t(\omega) < \infty \right\} \cap \left\{ \omega : \sum_{t \geq 1} Y'_t(\omega) < \infty \right\}. \tag{14}$$

We have already shown that on Ω'_0, $\lim_{t\to\infty} X'_t(\omega)$ exists and is finite. Therefore,

$$\lim_{t\to\infty} X_t(\omega) = \lim_{t\to\infty} \Phi_t(\omega) X'_t(\omega)$$

exists and is finite on Ω'_0. Denote this limit by $X_\infty(\omega)$.

Fix an $\varepsilon > 0$. Define

$$\mathcal{E} = \bigcap_{s \geq 1} \bigcup_{t \geq s} \mathbb{1}\{|X_t(\omega) - X_\infty(\omega)| > \varepsilon\}.$$

Since $X_t(\omega)$ converges almost surely to $X_\infty(\omega)$ on Ω_0', we have $\mathbb{P}(\Omega_0' \cap \mathcal{E}) = 0$.

Now for any $c > 0$, define the set $\Omega^{(c)} = \{\omega : \Phi_t(\omega) < cI$ for all $t \geq 1\}$. Since $Y_t = \Phi_{t+1} Y_t'$, we have $\Omega_0 \cap \Omega^{(c)} \subset \Omega_0'$ and therefore

$$\mathbb{P}(\Omega^{(c)} \cap \Omega_0 \cap \mathcal{E}) \leq \mathbb{P}(\Omega_0' \cap \mathcal{E}) = 0. \tag{15}$$

Let $\{c_k\}_{k\geq 1}$ be a countable sequence of positive numbers diverging to infinity. Then, by countable additivity,

$$\begin{aligned}\mathbb{P}(\Omega_0 \cap \mathcal{E}) &= \mathbb{P}(\lim_{k\to\infty} \Omega^{(c_k)} \cap \Omega_0 \cap \mathcal{E}) \\ &= \lim_{k\to\infty} \mathbb{P}(\Omega^{(c_k)} \cap \Omega_0 \cap \mathcal{E}) = 0 \end{aligned} \tag{16}$$

where the last equality follows from (15). Hence $X_t(\omega)$ converges almost surely to $X_\infty(\omega)$ on Ω_0. ■

IV. Some applications

A. Autoregressive almost supermartingales

Consider $\mathbb{R}_+^p$-valued stochastic processes $\{X_t\}_{t\geq 1}$, $\{Y_t\}_{t\geq 1}$, and $\{Z_t\}_{t\geq 1}$ and $\mathbb{R}_+$-valued stochastic processes $\{a_{t,d}\}_{t\geq 1}$, $d \in \{1, \dots, D\}$, all adapted to a filtration $\{\mathcal{F}_t\}_{t\geq 1}$, that satisfy for all $t \geq 1$,

$$\begin{aligned}\mathbb{E}_t[X_{t+1}] \leq a_{t,0} X_t + a_{t,1} X_{t-1} + \cdots + a_{t,D} X_{t-D} \\ + Y_t - Z_t.\end{aligned} \tag{17}$$

We call such a process *autoregressive almost supermartingale*.

Define $\bar{X}_t = \operatorname{vec}(X_{t-D}, \dots, X_t)$. Then, the dynamics (17) may be written as

$$\mathbb{E}_t[\bar{X}_{t+1}] \leq A_t \bar{X}_t + Y_t - Z_t \tag{18}$$

where $A_t \in \mathbb{R}_+^{p(D+1)\times p(D+1)}$ is given by

$$A_t = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 0 & I \\ a_{t,D} I & a_{t,D-1} I & \cdots & \cdots & a_{t,0} I \end{bmatrix} \tag{19}$$

where each block is $p \times p$.

Proposition 1 *Suppose $a_{t,D} \neq 0$ and the matrices A_t defined in (19) satisfy (A3). Define Ψ_t as in (3) and the set Ω_0 as in Theorem 2. Then, for all $\omega \in \Omega_0$, we have*

1) $\lim_{t\to\infty} X_t(\omega)$ *exists and is finite.*
2) $\sum_{t\geq 1} \Psi_t Z_t(\omega) < \infty$.

Proof Eq. (18) may be viewed as a vector almost-supermartingale. The matrices A_t satisfy (A1) and (A2). It is assumed that (A3) is satisfied. Therefore, the result follows from Theorem 2. ■

We now present two examples where we can explicitly verify (A3).

1) Example 1: Suppose that $p = 1$ and there exists a deterministic sequence $\bar{a}_d$, $d \in \{0, \dots, D\}$, such that $a_{t,d} \leq \bar{a}_d$ for all d. Then, $A_t \leq \bar{A}$, where

$$\bar{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 0 & 1 \\ \bar{a}_D & \bar{a}_{D-1} & \cdots & \cdots & \bar{a}_0 \end{bmatrix}.$$

Then, $\lim_{t\to\infty} \bar{A}^t$ exists if all the eigenvalues of $\bar{A}$ lie inside the unit circle. A trivial instance is $\sum_{d=1}^{D} \bar{a}_d < 1$ where $\bar{A}$ is a sub-stochastic matrix and therefore $\bar{A}^t$ converges to zero. Note that in this case, if the inequality is replaced by an equality, then $\bar{A}$ is an irreducible and aperiodic stochastic matrix, and therefore $\bar{A}^t$ still converges to a finite limit. In general, the characteristic polynomial of $\bar{A}$ is[1]

$$f(z) = \det(zI - \bar{A}) = z^{D+1} - \bar{a}_0 z^D - \bar{a}_1 z^{D-1} - \cdots - \bar{a}_D$$

Thus, the stability of $\bar{A}$ may be checked by applying the Jury stability criteria to $f(z)$.

2) Example 2: Suppose $p = 1$, $D = 1$, $a_{t,0} \leq 1 + \beta_{0,t}$ and $a_{t,1} \leq \beta_{1,t}$, where $\beta_{0,t}, \beta_{1,t} \in [0, 1]$. In this case,

$$\bar{A}_t = \begin{bmatrix} 0 & 1+\beta_{1,t} \\ 1 & \beta_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \beta_{0,t} & \beta_{1,t} \end{bmatrix} =: P + B_t,$$

where P is a permutation matrix and $P^2 = I$. To argue that $\prod_{t\geq 1} \bar{A}_t$ converges, we decompose the product as

$$\prod_{t\geq 1} \bar{A}_t = \cdots (\bar{A}_4 \bar{A}_3)(\bar{A}_2 \bar{A}_1).$$

Then, a generic factor of the form $(\bar{A}_{2t} \bar{A}_{2t-1})$ may be simplified as:

$$\begin{aligned}\bar{A}_{2t} \bar{A}_{2t-1} &= (P + B_{2t})(P + B_{2t-1}) \\ &= P^2 + P(B_{2t} + B_{2t-1}) + B_{2t} B_{2t-1} \\ &=: I + C_t.\end{aligned} \tag{20}$$

We now use condition (C1) to show that $\prod_{t\geq 1}(I + C_t)$ converges. In particular, we take $\|\cdot\|$ to be ℓ_1 induced norm given by $\|A\| = \max_j \sum_i |A_{ij}|$. Then,

$$\begin{aligned}\|C_t\| &\overset{(a)}{\leq} \|P(B_{2t} + B_{2t-1})\| + \|B_{2t} B_{2t-1}\| \\ &\overset{(b)}{\leq} \|B_{2t} + B_{2t-1}\| + \|B_{2t} B_{2t-1}\| \\ &= (\beta_{0,2t} + \beta_{0.2t-1} + \beta_{1,2t} + \beta_{1,2t-1}) \\ &\quad + \beta_{1,2t}(\beta_{0,2t-1} + \beta_{1,2t-1}) \\ &\overset{(c)}{\leq} 2(\beta_{0,2t} + \beta_{0.2t-1} + \beta_{1,2t} + \beta_{1,2t-1}) \end{aligned} \tag{21}$$

where (a) follows from triangle inequality, (b) follows from sub-multiplicative property of matrix norm and the fact that $\|P\| = 1$, and (c) follows from the fact that $\beta_{1,t} \in [0, 1]$.

[1]Since $\bar{A}$ is in controllable canonical form, its characteristic polynomial can be written by inspection.

Thus, Eq. (21) implies that a sufficient condition for (A3) (i.e., convergence of $\prod_{t\geq 1}\bar{A}_t = \prod_{t\geq 1}\bar{C}_t$) is

$$\sum_{t\geq 1}(\beta_{0,2t} + \beta_{0.2t-1} + \beta_{1,2t} + \beta_{1,2t-1}) < \infty \tag{22}$$

B. Delayed almost supermartingales

Consider $\mathbb{R}^p_+$-valued stochastic processes $\{X_t\}_{t\geq 1}$, $\{Y_t\}_{t\geq 1}$, and $\{Z_t\}_{t\geq 1}$ and a $\{0,\dots,D\}$-valued stochastic process $\{d_t\}_{t\geq 1}$, all adapted to a filtration $\{\mathcal{F}_t\}_{t\geq 1}$, that satisfy for all $t \geq 1$,

$$\begin{aligned}\mathbb{E}_t[X_{t+1}] &\leq \mathbb{E}_t[X_{t-d_t}] + Y_t - Z_t \\ &\leq \sum_{d=0}^{D}\mathbb{P}_t(d_t = d)X_{t-d} + Y_t - Z_t.\end{aligned} \tag{23}$$

We call $\{X_t\}_{t\geq 1}$ a *delayed almost supermartingale*. Note that when $D = 0$, the process reduces to a standard almost supermartingale. Define $a_{t,d} = \mathbb{P}_t(d_t = d)$. Then (23) is equivalent to

$$\mathbb{E}_t[X_{t+1}] \leq \sum_{d=0}^{D} a_{t,d}X_{t-d} + Y_t - Z_t \tag{24}$$

which is a special case of autoregressive almost supermartingale defined in Sec. IV-A.

Then, if the assumptions of Proposition 1 are satisfied, $\{X_t\}_{t\geq 1}$ converges almost surely. As argued in Example 1, in the special case where $a_{t,d}$ is time-homogeneous, the matrix A_t is irreducible and aperiodic (provided $a_D \neq 0$), and therefore assumption (A3) is satisfied.

It is worth pointing out in the special case when $a_{t,D} = 1$ and $a_{t,d} = 0$ for $d \neq D$, Assumption (A3) is not satisfied. This is because in this case

$$A_t = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 0 & I \\ I & 0 & \cdots & \cdots & 0 \end{bmatrix}$$

where each block is $p \times p$. Since A_t does not depend on time t, we simply denote it by A. Observe that A is a block permutation matrix and $\{A^t\}_{t\geq 1}$ is a periodic process with period D (because $A^D = I$). Therefore the process A^t does not converge. However, note that the matrices A_t satisfy assumptions (A4) and (A5). Therefore, we can use Theorem 3 to argue that in this case $\{X_t\}_{t\geq 1}$ converges almost surely.

C. Delayed temporal difference learning

Let $H\colon \mathbb{R}^p \to \mathbb{R}^p$ be a pseudo-contraction with respect to the Euclidean norm $\|\cdot\|$, i.e., there exist a unique fixed point θ^* and a radius of contraction $\gamma \in (0,1)$ such that

$$\|H\theta - \theta^*\| \leq \gamma\|\theta - \theta^*\|, \quad \forall \theta \in \mathbb{R}^p. \tag{25}$$

We impose the following assumption:

(TD) $\|\theta^*\| \leq M$.

Such an assumption is valid in reinforcement learning where temporal difference learning is used to learn the action-value function (i.e., the Q-function) of a policy. If the per-step reward is bounded by $R_{\max}$, then we know that the action-value function of any policy is bounded by $R_{\max}/(1-\gamma)$.

We assume that there an agent who knows the bound M and wants to find the fixed point θ^*. It has access to an oracle which takes θ as inputs and returns $H\theta + \xi$ *after one step delay*, where ξ is independent noise. We run temporal difference learning to find the fixed point of H, i.e, we start with an initial guess $\theta_0 = \theta_1$ and for $t \geq 1$ update

$$\theta_{t+1} = \Pi_M\Big[(1-\alpha_t)\theta_t + \alpha_t\big[H\theta_{t-1} + \xi_{t+1}\big]\Big], \tag{26}$$

where $\{\alpha_t\}_{t\geq 1}$ are the learning rates and Π_M is the projection from $\mathbb{R}^p$ to $\{\theta\colon \|\theta\| \leq M\}$. Note that the term is the square brackets denotes the output of the oracle, which is delayed. We denote the noise by ξ_{t+1} for consistency of notation. Let $\mathcal{F}_t$ denote the σ-algebra generated by $(\theta_{1:t}, \alpha_{1:t-1}, \xi_{1:t})$. We consider the following assumptions on the noise.

(N1) $\{\xi_t\}_{t\geq 1}$ is a martingale difference sequence with respect to $\{\mathcal{F}_t\}_{t\geq 1}$, i.e.,

$$\mathbb{E}_t[\xi_{t+1}] = 0, \quad \text{a.s.}, \quad \forall t \geq 1.$$

(N2) The noise $\{\xi_t\}_{t\geq 1}$ satisfies $\|\xi_{t+1}\| \leq \sigma$ a.s. for all $t \geq 1$. Therefore,

$$\mathbb{E}_t[\|\xi_{t+1}\|^2] \leq \sigma^2, \quad \text{a.s.}, \quad \forall t \geq 1.$$

Finally, we consider the following assumptions on the learning rates.

(R1) $\sum_{t\geq 1}\alpha_t^2 < \infty$.
(R2) $\sum_{t\geq 1}\alpha_t = \infty$.

Proposition 2 *Under (TD), (N1) and (N2), we have the following:*

1) *If (R1) holds, then θ_t converges to a limit almost surely.*
2) *If, in addition, (R2) holds, then $\theta_t \to \theta^*$ almost surely as $t \to \infty$.*

PROOF The boundedness of the iterates implies that

$$\|\theta_{t+1}-\theta^*\| \leq \min\big\{\|(\theta_t-\theta^*)+\alpha_t\big[H\theta_{t-1}-\theta_t+\xi_{t+1}\big]\|, 2M\big\}. \tag{27}$$

Thus,

$$\begin{aligned}\mathbb{E}_t[\|\theta_{t+1} - \theta^*\|^2] \leq \|\theta_t - \theta^*\|^2 + \alpha_t^2\|H\theta_{t-1} - \theta_t\|^2 + \alpha_t^2\sigma^2 \\ + \alpha_t\langle\theta_t - \theta^*, H\theta_{t-1} - \theta_*\rangle\end{aligned} \tag{28}$$

Observe that

$$\begin{aligned}\|H\theta_{t-1} - \theta_t\| &\leq \|H\theta_{t-1} - \theta^*\| + \|\theta_t - \theta^*\| \\ &\leq \|\theta_{t-1} - \theta^*\| + \|\theta_t - \theta^*\|\end{aligned} \tag{29}$$

where the first inequality follows from triangle inequality and the second follows from (25). For ease of notation, we use $\Delta_t = \|\theta_t - \theta^*\|$ for all t. Therefore, (29) implies

$$\|H\theta_{t-1} - \theta_t\|^2 \leq \Delta_{t-1}^2 + \Delta_t^2 + 2\Delta_t\Delta_{t-1}. \tag{30}$$

Now consider

$$\begin{aligned}
&\langle \theta_t - \theta^*, H\theta_{t-1} - \theta_t \rangle \\
&\quad = \langle \theta_t - \theta^*, H\theta_{t-1} - \theta^* \rangle - \langle \theta_t - \theta^*, \theta_t - \theta^* \rangle \\
&\quad \leq \|\theta_t - \theta_*\| \|H\theta_{t-1} - \theta^*\| - \Delta_t^2 \\
&\quad \leq \gamma \Delta_{t-1} \Delta_t - \Delta_t^2
\end{aligned} \tag{31}$$

Substituting (30) and (31) in (28), we get

$$\begin{aligned}
\mathbb{E}_t[\Delta_{t+1}^2] &\leq \Delta_t^2 + \alpha_t^2[\Delta_{t-1}^2 + \Delta_t^2 + 2\Delta_{t-1}\Delta_t + \sigma^2] \\
&\quad + \alpha_t\big[\gamma\Delta_{t-1}\Delta_t - \Delta_t^2\big]
\end{aligned} \tag{32}$$

In non-delayed TD learning, the last term in the square bracket is $\gamma\Delta_t^2 - \Delta_t^2$, which is negative and therefore can be ignored. However, that is not possible when we have delay, so a different method is needed to bound this term. We do so by exploiting the boundedness of the iterates.

From (27) (evaluated at $t-1$), we have

$$\begin{aligned}
\Delta_t &\leq \Delta_{t-1} + \alpha_t\|H\theta_{t-2} - \theta_*\| + \alpha_t\|\xi_t\| \\
&\leq \Delta_{t-1} + \alpha_t\gamma\Delta_{t-2} + \alpha_t\|\xi_t\|
\end{aligned} \tag{33}$$

Substituting the above, in the last term in (32), we get

$$\begin{aligned}
\alpha_t\big[\gamma\Delta_{t-1}\Delta_t - \Delta_t^2\big] &= \alpha_t\gamma\big[\Delta_{t-1}\Delta_t - \Delta_t^2\big] - \alpha_t\bar{\gamma}\Delta_t^2 \\
&\leq \alpha_t^2\gamma\Delta_t\big[\Delta_{t-2} + \|\xi_t\|\big] - \alpha_t\bar{\gamma}\Delta_t^2
\end{aligned}$$

where we use $\bar{\gamma} = (1-\gamma)$. Substituting back in (32), we get

$$\begin{aligned}
\mathbb{E}_t[\Delta_{t+1}^2] &\leq \Delta_t^2 + \alpha_t^2[\Delta_{t-1}^2 + \Delta_t^2 + 2\Delta_{t-1}\Delta_t + \sigma^2] \\
&\quad + \alpha_t^2\gamma\Delta_t\big[\Delta_{t-2} + \|\xi_t\|\big] - \alpha_2\bar{\gamma}\Delta_t \\
&\overset{(a)}{\leq} (1+\alpha_t^2)\Delta_t^2 + \alpha_t^2\Delta_{t-1}^2 \\
&\quad + \alpha_t^2\big[(8+4\gamma)M^2 + \sigma^2 + \gamma\sigma\big] - \alpha_t\bar{\gamma}\Delta_t
\end{aligned} \tag{34}$$

where (a) uses $\Delta_t \leq 2M$ from (27).

Now observe that (34) is a autoregressive almost supermartingale of the form presented in Example 2 in Sec. IV-A.2 with $X_t = \Delta_t^2$, $Y_t = \alpha_t^2\big[(8+4\gamma)M^2 + \sigma^2 + \gamma\sigma\big]$ and $Z_t = \alpha_t\bar{\gamma}\Delta_t$. Note that (22) is equivalent to (R1), thus (A3) is satisfied. Thus, we can apply the result of Proposition 1.

For the choice of Y_t above, (R1) implies that Ω_0 defined in Theorem 2 is almost surely equal to Ω. Thus, from Proposition 1 and Theorem 2, we get that Δ_t^2 converges almost surely. This proves the first part of the proposition.

For any $\omega \in \Omega$, let $\zeta(\omega) := \lim_{t\to\infty}\Delta_t^2(\omega)$. To prove the second part, we will show that $\zeta(\omega) = 0$. We start with the observation that Proposition 1 and Theorem 2 imply that almost surely.

$$\sum_{t\geq 1}\Psi_{t+1}Z_{t+1} < \infty.$$

Recall that in Sec. IV-A.2, we have argued that $\bar{A}_{2t}\bar{A}_{2t-1}$ is of the form $I + C_t$, where C_t is a non-negative matrix. Therefore, Ψ_{t+1} is element-wise greater than or equal to I. Hence,

$$\sum_{t\geq 1} Z_{t+1} \leq \sum_{t\geq 1}\Psi_{t+1}Z_{t+1} < \infty. \tag{35}$$

We now show that $\zeta(\omega) = 0$ by contraction. Assume that for some $\omega \in \Omega_0$, $\zeta(\omega) = 2\varepsilon > 0$. Choose a $T(\omega)$ such that for all $t > T(\omega)$, we have $\Delta_t^2(\omega) > \varepsilon$. Therefore,

$$\sum_{t\geq T(\omega)} Z_t(\omega) = \sum_{t\geq T(\omega)}\alpha_t\Delta_t^2(\omega) \geq \sum_{t\geq T(\omega)}\alpha_t\varepsilon = \infty$$

where the last equality follows from (R2). The above limit contradicts (35). Therefore, for almost all $\omega \in \Omega_0$, $\zeta(\omega) = 0$, i.e., $\Delta_t^2(\omega) \to 0$ almost surely. Therefore, $\theta_t \to \theta^*$, almost surely. ∎

V. Conclusion

In this paper, we consider a generalization of almost supermartingale convergence theorem from scalar-valued non-negative processes to vector-valued non-negative processes. We provide two sufficient conditions for convergence which rely on convergence of infinite product of matrices which might be viewed as generalization of the conditions for convergence in the scalar case. We show that the vector-valued almost supermartingales can be used to prove convergence of what we call almost auto-regressive supermartingales. We use convergence of almost auto-regressive supermartingales to establish convergence of delayed almost supermartingales and delayed temporal difference learning. In general, we believe that vector-valued almost supermartingale convergence theorem could be fundamental tool to establish convergence of delayed version of various learning algorithms.

References

[1] H.-F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*. Birkhäuser Boston, 1991.

[2] T. L. Lai and Z. Ying, "Recursive identification and adaptive prediction in linear stochastic systems," *SIAM Journal on Control and Optimization*, vol. 29, no. 5, pp. 1061–1090, 1991.

[3] L. Ljung, "Perspectives on system identification," *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010.

[4] Y. Tsypkin, *Adaptation and learning in automatic systems*. Academic Press: New York, 1971.

[5] P. E. Caines, *Linear stochastic systems*. SIAM, 2018.

[6] P. Mertikopoulos, N. Hallak, A. Kavis, and V. Cevher, "On the almost sure convergence of stochastic gradient descent in non-convex problems," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 1117–1128.

[7] O. Sebbouh, R. M. Gower, and A. Defazio, "Almost sure convergence rates for stochastic gradient descent and stochastic heavy ball," in *Conference on Learning Theory*. PMLR, 2021, pp. 3935–3971.

[8] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279–292, may 1992.

[9] S. Meyn, *Control systems and reinforcement learning*. Cambridge University Press, 2022.

[10] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2014.

[11] A. Nedić and J. Liu, "Distributed optimization for control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 77–103, 2018.

[12] J. R. Blum, "Approximation methods which converge with probability one," *The Annals of Mathematical Statistics*, pp. 382–386, 1954.

[13] A. Dvoretsky, *On stochastic approximation*. Mathematics Division, Office of Scientific Research, US Air Force, 1955.

[14] E. Gladyshev, "On stochastic approximations," *Theory of Probability and its Applications*, vol. 10, no. 2, pp. 275–278, 1965.

[15] M. Vidyasagar, "Convergence of stochastic approximation via martingale and converse Lyapunov methods," *Mathematics of Control, Signals, and Systems*, vol. 35, no. 2, pp. 351–374, Jan. 2023.

[16] D. Derevitskii and A. Fradkov, “Two models for analyzsing dynamics of adaptation algorithms,” *Automation and Remote Control*, vol. 35, no. 1, pp. 59–67, 1974.

[17] L. Ljung, “Analysis of recursive stochastic algorithms,” *IEEE Transactions on Automatic Control*, vol. 22, no. 4, pp. 551–575, 1977.

[18] V. S. Borkar and S. P. Meyn, “The o.d.e. method for convergence of stochastic approximation and reinforcement learning,” *SIAM Journal on Control and Optimization*, vol. 38, no. 2, pp. 447–469, jan 2000.

[19] M. Benaïm and M. W. Hirsch, “Stochastic approximation algorithms with constant step size whose average is cooperative,” *The Annals of Applied Probability*, vol. 9, no. 1, Feb. 1999.

[20] V. Borkar, S. Chen, A. Devraj, I. Kontoyiannis, and S. Meyn, “The ode method for asymptotic statistics in stochastic approximation and reinforcement learning,” *arXiv preprint arXiv:2110.14427*, 2021.

[21] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*. Springer New York, 1997.

[22] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive algorithms and stochastic approximations*. Springer Science & Business Media, 2012, vol. 22.

[23] V. S. Borkar, *Stochastic approximation: a dynamical systems viewpoint*. Hindustan Book Agency, 2008.

[24] A. Shiryayev, *Probability*. Springer-Verlag: New York, 1984.

[25] J. Neveu, *Discrete Parameter Martingales*. North Holland, 1975.

[26] M. H. A. Davis, “Martingale methods in stochastic control,” in *Stochastic Control Theory and Stochastic Differential Systems*, M. Kohlmann and W. Vogel, Eds. Berlin, Heidelberg: Springer: Berlin, 1979, pp. 85–117.

[27] B. Polyak, *Introduction to optimization*. Optimization Software Inc: New York, 1987.

[28] H. Robbins and D. Siegmund, “A convergence theorem for non-negative almost supermartingales and some applications,” in *Optimizing Methods in Statistics*. Elsevier, 1971, pp. 233–257.

[29] W. F. Trench, “Invertibly convergent infinite products of matrices,” *Journal of Computational and Applied Mathematics*, vol. 101, no. 1–2, pp. 255–263, Jan. 1999.

[30] M. Artzrouni, “On the convergence of infinite products of matrices,” *Linear Algebra and its Applications*, vol. 74, pp. 11–21, Feb. 1986.