2009 American Control Conference
Hyatt Regency Riverfront, St. Louis, MO, USA
June 10-12, 2009

On the Adaptive Control of a Class of Partially Observed Markov Decision Processes

Shun-Pin Hsu
National Chung-Hsing University
Electrical Engineering
Taichung, Taiwan 402

Dong-Ming Chuang
The University of Texas
Electrical & Computer Engineering
Austin, TX 78712-0240

Ari Arapostathis
The University of Texas
Electrical & Computer Engineering
Austin, TX 78712-0240

***Abstract*— We study the adaptive control problems of a class of discrete-time partially observed Markov decision processes whose transition kernels are parameterized by a unknown vector. Given a sequence of parameter estimates converging to the true value with probability 1, we propose an adaptive control policy and show that under some conditions this policy is self-optimizing in the long-run average sense.**

I. INTRODUCTION

During the past decade considerable effort has been invested in the study of stochastic adaptive control. Special attentions have been paid to the system with incomplete or noisy state observation, in particular, the discrete-time partially observable Markov decision processes (POMDPs) with the transition probability matrix depending on some unknown parameter vector $\theta \in \Theta$ where $\Theta \in \mathbb{R}^{N_\theta}$ is the parameter space with cardinality N_θ. The purpose of the adaptive control of POMDPs is to optimally regulate the modelled system in the presence of parameter uncertainty. To achieve this, a control strategy widely used in the linear filtering was proposed in [1], where the parameter uncertainty in the partially observed system was handled by deriving the filter for the conditional probabilities and estimating the parameters at the same time, then plugging into the filter with the most updated parameter estimates. This adaptive estimation algorithm can be analyzed by the ordinary differential equation method and the existence of a convergent sequence $\{\hat{\theta}_t\}_{t=0}^{\infty} \to \theta$ in appropriate probability measure can be shown (see [21]). Along with this result, a methodology involving a set of assumptions was proposed in [14] to show the self-optimality (defined later) of the adaptive policy. The major assumptions mentioned there included that *a.* for each $\theta \in \Theta$ the optimal policy exists, and *b.* the sequence of estimation errors between the true and the estimated information states converges to 0. For the long-run average cost model, positive transition matrices in [11] and *renewability* condition in [16] are assumed to guarantee *a*. Recently (see [19]) we propose several sufficient conditions for the same topic and show that our results solve a wider class of practical problems. In this paper we present one of the condition (which leads to *a.*) from [19] on the structure of the transition matrices and show that a little modification of this condition leads to *b.*, if the convergence $\{\hat{\theta}_t\}_{t=0}^{\infty} \to \theta$ is fast enough. Hence, the self-optimizing property of the adaptive policy is obtained according to the methodology in [14]. We note that our conditions are either weaker or more justifiable than those of other work in the literature.

This paper is organized as follows. Section 2 reviews the technical preliminaries, including the ergodic and adaptive control problem of discrete-time POMDPs. Section 3 presents the main result of the paper. Several structural properties of the product of non-negative matrices are proved and applied to derive the self-optimization, in the long-run average sense, of the proposed adaptive policy.

II. PRELIMINARIES

A. Partially Observed Markov Decision Process

A discrete-time partially observable Markov decision process is governed by a five-tuple $(\boldsymbol{S},\boldsymbol{U},\mathcal{U},Q,c)$ with the following meanings: $\boldsymbol{S} = \boldsymbol{X} \times \boldsymbol{Y}$ is the process's state space where $\boldsymbol{X} = \{1, 2, \cdots, N_x\}$ is the finite system space and $\boldsymbol{Y} = \{1, 2, \cdots, N_y\}$ is the finite observable space. $\boldsymbol{U}$ is the finite action space. Let $\mathcal{B}(\boldsymbol{V})$ denote the σ-algebra for a given topological space $\boldsymbol{V}$, then $\mathcal{U}$: $\boldsymbol{X} \to \mathcal{B}(\boldsymbol{U})$ means a set-valued map with compact non-empty value and $\mathcal{U}(x)$ is the set of feasible actions when the system is in state $x \in \boldsymbol{X}$. Q is the transition matrix of the process and c is the cost function. Specifically, when the system state at time t is X_t and a control U_t is taken, a cost $c(X_t, U_t)$ is incurred and the system moves to next state X_{t+1} with observation Y_{t+1} according to the transition matrix Q defined by

$$\begin{aligned} Q(y, U_t)_{X_t j} &:= \text{Prob}(X_{t+1} = j, Y_{t+1} = y | X_t, U_t) \\ &= \text{Prob}(X_{t+1} = j, Y_{t+1} = y | X_k, Y_k, U_k, k \le t) \end{aligned}$$

for all $t \in \mathbb{N}_0$ (the set of nonnegative integers), $j \in \boldsymbol{X}$, and $y \in \boldsymbol{Y}$. Note that by definition the element of Q is nonnegative and satisfies

$$\sum_{y \in \boldsymbol{Y}} \sum_{j \in \boldsymbol{X}} Q(y, u)_{ij} = 1$$

for each $i \in \boldsymbol{X}$ and $u \in \boldsymbol{U}$. Denote $\boldsymbol{\Psi} := \boldsymbol{\mathcal{P}}(\boldsymbol{X})$, the probability (row) vector space on $\boldsymbol{X}$. To transform the partially observed process into its completely observed equivalent, we apply the Bayes rule (see [20, p84]) and construct the information state sequence $\{\psi_t\}$ by recursively calculating

$$\psi_{t+1} := \sum_{y \in \boldsymbol{Y}} \frac{\psi_t Q(y, U_t)}{\psi_t Q(y, U_t) \mathbf{1}} \cdot \mathbf{1}_{\{Y_{t+1} = y\}},$$

for each $\psi_t \in \boldsymbol{\Psi}$, $U_t \in \boldsymbol{U}$, and $t \in \mathbb{N}_0$, where $\mathbf{1}_{\{\cdot\}}$ is the indicator function and $\mathbf{1}$ is a column vector of 1's with size N_x. Write $V(\psi, y, u) = \psi \cdot Q(y, u) \cdot \mathbf{1}$ as the conditional probability and

$$T(\psi, y, u) := \frac{\psi \cdot Q(y, u)}{V(\psi, y, u)} \qquad \text{for} \quad V(\psi, y, u) \neq 0\,,$$

as the posteriori conditional distribution, then the transition kernel for the information state is given by

$$\begin{aligned}\mathcal{K}(B|\psi, u) :&= \text{Prob}\{\psi_{t+1} \in B | \psi_t = \psi, U_t = u\}, \qquad (1)\\ &= \sum_{Y_{t+1} \in \boldsymbol{Y}} V(\psi, Y_{t+1}, u) \cdot \mathbf{1}_{\{T(\psi, Y_{t+1}, u) \in B\}}\end{aligned}$$

for each $B \in \mathcal{B}(\boldsymbol{\Psi})$, $\psi \in \boldsymbol{\Psi}$, and $u \in \boldsymbol{U}$. So we have the transformed five-tuple $(\boldsymbol{\Psi}, \boldsymbol{U}, \tilde{\mathcal{U}}, \mathcal{K}, \tilde{c})$ where $\tilde{\mathcal{U}}$: $\boldsymbol{\Psi} \to \mathcal{B}(\boldsymbol{U})$ and $\tilde{c}(\psi, u) := \sum_{x \in \boldsymbol{X}} c(x, u)\psi(x)$ for all $\psi \in \boldsymbol{\Psi}$ and $u \in \boldsymbol{U}$. For the original history space H_i of the partially observed process up to time i where $H_0 := \boldsymbol{\Psi}$, $H_t := H_{t-1} \times \boldsymbol{U} \times \boldsymbol{Y}$ for all $t \in \mathbb{N}$ (the set of positive integers), we obtain a corresponding completely observed history space: $\hat{H}_0 := \boldsymbol{\Psi}$, $\hat{H}_t := \hat{H}_{t-1} \times \boldsymbol{U} \times \boldsymbol{\Psi}$ for all $t \in \mathbb{N}$.

An *admissible strategy* or *admissible policy* π is a sequence $\{\pi_t\}_{t=0}^{\infty}$ of Borel measurable stochastic kernels π_t on $\boldsymbol{U}$ given $\hat{H}_t$ satisfying $\pi_t(\tilde{\mathcal{U}}(\psi_t)|h_t) = 1$ for all $\psi_t \in \boldsymbol{\Psi}$, $h_t \in \hat{H}_t$, and $t \in \mathbb{N}_0$. An admissible policy is called *deterministic* if there exists a function $f : \boldsymbol{\Psi} \to \boldsymbol{U}$ such that $\pi_t(f(\psi_t)|h_t) = 1$ for all $\psi_t \in \boldsymbol{\Psi}$, $h_t \in \hat{H}_t$ and $t \in \mathbb{N}_0$.

It is shown in [7] that for an initial distribution $\psi_0 \in \boldsymbol{\Psi}$ and admissible strategy π, there exists an unique probability measure $\mathbb{P}^{\pi}_{\psi_0}$ induced on the sample path $(\boldsymbol{\Psi} \times \boldsymbol{U})^{\infty}$. We use $\mathbb{E}^{\pi}_{\psi_0}$ to represent the corresponding expectation operator.

B. Ergodic Control

The objective of ergodic control is to decide the optimal strategy $\pi \in \Pi$ (the set of all admissible strategies) to minimize the incurred long-run average cost:

$$J(\psi_0, \pi) := \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}^{\pi}_{\psi_0} \left[\sum_{t=0}^{T-1} \tilde{c}(\psi_t, U_t) \right]. \qquad (2)$$

The classical *vanishing discount limit* method approaches this problem by extending the result from the *β-discounted* cost model:

$$J_\beta(\psi_0, \pi) := \limsup_{T \to \infty} \mathbb{E}^{\pi}_{\psi_0} \left[\sum_{t=0}^{T-1} \beta^t \tilde{c}(\psi_t, U_t) \right]. \qquad (3)$$

where β is in (0,1). If π_β is the minimizing policy in the following sense and results in a value function $h_\beta(\psi)$ where

$$h_\beta(\psi) = J_\beta(\psi, \pi_\beta) = \inf_{\pi \in \Pi} J_\beta(\psi, \pi)\,,\ \psi \in \boldsymbol{\Psi}\,, \qquad (4)$$

then the following assumption and its implication are well known.

Assumption 2.1: $c : \boldsymbol{X} \times \boldsymbol{U} \to \mathbb{R}^+$ is the one-stage cost function that is nonnegative, bounded and continuous. Also, $U \to Q(y, U)$ is continuous for each $y \in \boldsymbol{Y}$.

Lemma 2.1: [17, Chapter 2] Suppose Assumption 2.1 holds. The value function $h_\beta(\psi)$ in (4) corresponding to the β-discounted cost model in (3) can be characterized by Bellman's β-discounted optimality equation:

$$h_\beta(\psi) = \min_{u \in \boldsymbol{U}} \left\{ \tilde{c}(\psi, u) + \beta \int_{\boldsymbol{Y}} h_\beta(\psi') \mathcal{K}(d\psi' | \psi, u) \right\} \qquad (5)$$

for all $\psi \in \boldsymbol{\Psi}$ where $\mathcal{K}$ is defined in (1). Any admissible policy resulting in the value function $h_\beta(\cdot)$ is β-discounted optimal.

It is well known that $h_\beta(\psi)$ is the unique solution in $\boldsymbol{C}(\boldsymbol{\Psi})$ (the space of continuous functions on $\boldsymbol{\Psi}$) for Bellman's β-discounted optimality equation. Also, it can be shown (see [22]) that $h_\beta(\cdot)$ is concave. Suppose β_0 is in (0,1) and $\{\beta_n\}_{n=1}^{\infty} \subset [\beta_0, 1)$ is a sequence with $\beta_n \to 1$. $\psi_* := \arg\min_{\psi \in \boldsymbol{\Psi}} h_\beta(\psi)$, and $\overline{h}_\beta(\psi) := h_\beta(\psi) - h_\beta(\psi_*)$. A well-known major condition (see [15]) that implies the existence of the long-run average optimal policy characterized by *Bellman's ergodic optimality equation* follows.

Assumption 2.2: $\{\overline{h}_{\beta_n}(\cdot)\}_{n=1}^{\infty}$ is uniformly bounded on $\boldsymbol{\Psi}$.

Theorem 2.2: Suppose Assumption 2.1 and Assumption 2.2 hold. Then there exist a constant ρ, which is the optimal ergodic cost, and a bounded, concave and continuous function h: $\boldsymbol{\Psi} \to \mathbb{R}$, such that $(\rho, h(\cdot))$ is a solution of the following dynamic programming equation:

$$\rho + h(\psi) = \min_{u \in \boldsymbol{U}} \left\{ \tilde{c}(\psi, u) + \int_{\boldsymbol{Y}} h(\psi') \mathcal{K}(d\psi' | \psi, u) \right\}. \qquad (6)$$

Also, the following is equivalent.

1) π^* is an optimal optimal.
2) $\pi^*(\psi)$ assigns a minimizer u for $\{\cdot\}$ in (6) for each $\psi \in \boldsymbol{\Psi}$.
3) $$\lim_{t \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}^{\pi^*}_{\psi_0} \{D(\psi_t, \pi^*(\psi_t))\} = 0$$
where the discrepancy function $D : \boldsymbol{\Psi} \times \boldsymbol{U} \to \mathbb{R}$ is defined by
$$D(\psi, u) := \tilde{c}(\psi, u) + \int_{\boldsymbol{Y}} h(\psi') \mathcal{K}(d\psi' | \psi, u) - \rho - h(\psi)\,.$$

Proof: The results follow from [15], and [18, Proposition 5.5.5]. ∎

C. Adaptive Control

When the transition matrix Q is parameterized by a unknown vector $\theta \in \Theta$, where $\Theta \subseteq \mathbb{R}^{N_\theta}$ is a compact space, an stochastic approximation-type estimation algorithm can be designed to form a sequence $\{\hat{\theta}_t\}_{t=0}^{\infty}$ such that $\hat{\theta}_t \to \theta$ w.p. 1 as $t \to \infty$. For simplicity we denote the parameterized transition matrix $\hat{Q}(y_t, u_{t-1}) = Q(y_t, u_{t-1}, \hat{\theta}_t)$ and sometimes $\overline{Q}(y_t, u_{t-1}) = Q(y_t, u_{t-1}, \theta)$. Suppose for each parameter $\theta \in \Theta$, there exists an associated optimal deterministic policy written as $\pi^*(\cdot, \theta)$. Define the adaptive policy π^a that generates a sequence of actions $\{u_t\}_{t=0}^{\infty}$ where $u_t = \pi^*(\hat{\psi}_t, \hat{\theta}_t)$ and

$$\hat{\psi}_{t+1} := \sum_{y \in \boldsymbol{Y}} \frac{\hat{\psi}_t \hat{Q}(y, u_t)}{\hat{\psi}_t \hat{Q}(y, u_t) \mathbf{1}} \cdot \mathbf{1}_{\{Y_{t+1} = y\}}, \qquad (7)$$

for each $\hat{\psi}_t \in \boldsymbol{\Psi}$, $u_t \in \boldsymbol{U}$, and $t \in \mathbb{N}_0$. According to the methodology in [14], for π^a to be a self-optimizing policy in the sense that $J_\theta(\psi_0, \pi^a) = \inf_{\pi\in\Pi} J_\theta(\psi_0, \pi)$, one important condition is to construct a sequence $\{\hat{\psi}_t\}_{t=0}^{\infty}$ of information state estimates such that under $\mathbb{P}_{\psi_0}^{\pi^a}$ we have $\|\hat{\psi}_t - \psi_t\|_1 \to 0$ as $t \to \infty$ where $\|\cdot\|_1$ is the l_1 norm. In the following we study the sufficient conditions for the convergence to hold. Specifically, we would like to show

$$\mathbb{E}_{\psi_0}^{\pi^a}\left[\|\psi_t - \hat{\psi}_t\|_1\right] \longrightarrow 0 \quad \text{as} \quad t \to \infty \quad \forall\, \psi_0 \in \boldsymbol{\Psi}\,. \tag{8}$$

D. Stochastic Matrix

Our main assumption to imply (8) is based on the ideas of the weak ergodicity in product of nonnegative matrices (see [24]). We now review some definitions, notations and concepts to be used later. A *nonnegative matrix* is a square matrix with all of its elements nonnegative. A *row-allowable matrix* is a nonnegative matrix with all of its row sums positive. A *substochastic matrix* is a nonnegative matrix with all of its row sums no greater than 1. If every row sum of a substochastic matrix equals 1, it is called a *stochastic matrix*. It is shown in [1] that if A is a nonnegative matrix, $\psi_1, \psi_2 \in \boldsymbol{\Psi}$, and define

$$\tau_1(A) := \frac{1}{2}\max_{i,j}\|A_{i\cdot} - A_{j\cdot}\|_1\,,$$

then we have

$$\|(\psi_1 - \psi_2)A\|_1 \le \tau_1(A)\,\|\psi_1 - \psi_2\|_1\,.$$

Apparently if A is a stochastic matrix, then $\tau(A) \in [0,1]$. If $\tau_1(A) \in (0,1)$, then A has the property of contraction mapping. Finally, we use the following definitions throughout this paper: $\{B_k\}_{k=1}^{\infty}$ is a sequence of $N_x \times N_x$ matrices; $B_m^{m+n} := B_m B_{m+1}\cdots B_{m+n}$, and $B^k := B_1 B_2 \cdots B_k$. $B_{i\cdot}$, $B_{\cdot j}$ represent the i^{th} row and j^{th} column of B, respectively. e^i is the ith row vector of the identity matrix with size $N_x \times N_x$.

III. Main Result

In this section we first present a major condition, a weaker version than that in [23, Assumption A.6], to guarantee the existence of long-run average optimal policy for the POMDP. For details on the proof and other weaker sufficient conditions readers are referred to [19].

Assumption 3.1: There exist constants $\varepsilon > 0$, $N_b \in \mathbb{N}$ and $\beta_0 < 1$ such that for each $\beta \in [\beta_0, 1)$ we have

$$\max_{1\le k\le N_b} \mathbb{P}_{e^i}^{\pi_\beta}\{Q(Y^k, U^{k-1})_{i_1 j} \ge \varepsilon Q(Y^k, U^{k-1})_{i_2 j}\} \ge \varepsilon\,,$$

where i, i_1, i_2, j all in $\boldsymbol{X}$ and k-step transition matrix $Q(y^k, u^{k-1}) := Q(y_1, u_0)\cdots Q(y_k, u_{k-1})$.

Lemma 3.1: [19] Assumption 3.1 is a sufficient condition for Assumption 2.2.

Next we present several properties concerning the product of row-allowable matrices. These results can be easily obtained from the definition of the row-allowable matrix mentioned in the previous section.

Lemma 3.2: For a row-allowable matrix B_n and a positive number ε, if $(B_n)_{i_1 j} \ge \varepsilon \cdot (B_n)_{i_2 j}$ for each $n \in \mathbb{N}$ and $i_1, i_2, j \in \boldsymbol{X}$, then we have

$$\frac{1}{(B_m^{m+r})_{i_2\cdot}\mathbf{1}} \ge \varepsilon \frac{1}{(B_m^{m+r})_{i_1\cdot}\mathbf{1}}$$

where $m, r \in \mathbb{N}$.

Corollary 3.3: Under the assumption of Lemma 3.2 we have

$$\frac{(B_m)_{i_2 j}(B_{m+1}^{m+n})_{j\cdot}\mathbf{1}}{(B_m^{m+r})_{i_2\cdot}\mathbf{1}} \ge \varepsilon^2 \frac{(B_m)_{i_1 j}(B_{m+1}^{m+n})_{j\cdot}\mathbf{1}}{(B_m^{m+r})_{i_1\cdot}\mathbf{1}}\,,$$

and for each $\psi_1, \psi_2 \in \boldsymbol{\Psi}$

$$\frac{\psi_1 B^n \mathbf{1}}{\psi_2 B^n \mathbf{1}} \ge \varepsilon\,.$$

Lemma 3.4: For a row-allowable matrix B_n and a positive number ε, if $(B_n)_{i_1\cdot}\mathbf{1} \ge \varepsilon \cdot (B_n)_{i_2\cdot}\mathbf{1}$ for each $n \in \mathbb{N}$ and $i_1, i_2 \in \boldsymbol{X}$, then we have

$$\frac{B^k_{i_1\cdot}\mathbf{1}}{B^k_{i_2\cdot}\mathbf{1}} \ge \varepsilon^k\,.$$

where $k \in \mathbb{N}$.

The next lemma is an improved version of Lemma 2.1 in [1], though it can be similarly derived.

Lemma 3.5: Under the assumption of Lemma 3.2 we have

$$\left|\frac{(B_m^{m+n})_{i_1 j}}{(B_m^{m+n})_{i_1\cdot}\mathbf{1}} - \frac{(B_m^{m+n})_{i_2 j}}{(B_m^{m+n})_{i_2\cdot}\mathbf{1}}\right| \le (1-\varepsilon^2)^n$$

for each $m, n \in \mathbb{N}$.

Lemma 3.6: If $\psi_1, \psi_2 \in \boldsymbol{\Psi}$,

1) under the assumption of Lemma 3.4 we have

$$\left\|\frac{\psi_1 B^n}{\psi_1 B^n \mathbf{1}} - \frac{\psi_2 B^n}{\psi_2 B^n \mathbf{1}}\right\|_1 \le \frac{2}{\varepsilon^n}\,\|\psi_1 - \psi_2\|_1\,;$$

2) under the assumption of Lemma 3.2 we have

$$\left\|\frac{\psi_1 B^n}{\psi_1 B^n \mathbf{1}} - \frac{\psi_2 B^n}{\psi_2 B^n \mathbf{1}}\right\|_1 \le \frac{N_x}{\varepsilon}(1-\varepsilon^2)^{n-1}\,\|\psi_1 - \psi_2\|_1\,;$$

Proof: See Appendix. ∎

Now we propose the main assumptions based on Lemma 3.6.

Assumption 3.2: For each parameter $\theta \in \Theta$, $Q(y, u, \theta)$ is row-allowable for each $y \in \boldsymbol{Y}$ and $u \in \boldsymbol{U}$. Also, there exist constants $\varepsilon > 0$, $N_b \in \mathbb{N}$ such that for each i_1, i_2, $j \in \boldsymbol{X}$

$$\max_{1\le k\le N_b} \mathbb{P}_{\psi_0}^{\pi^a}\{Q(Y^k, U^{k-1}, \theta)_{i_1 j} \ge \varepsilon Q(Y^k, U^{k-1}, \theta)_{i_2 j}\} = 1$$

for all $\psi_0 \in \boldsymbol{\Psi}$, where π^a is the adaptive strategy described in (7).

Suppose Assumption 3.2 holds, then there exists an increasing sequence of integers $\{m_l\}_{l=0}^{n} \subset \mathbb{N}_0$ satisfying m_0=0, $m_{l+1} - m_l \le N_b$, for $l = 0, 1, \cdots, n-1$, $m_n \le t$ such that

$$\mathbb{P}_{\psi_0}^{\pi^a}\{[\overline{Q}(Y_{m_{l-1}+1}, U_{m_{l-1}})\cdots\overline{Q}(Y_{m_l}, U_{m_l-1})]_{i_1 j} \ge \varepsilon[\overline{Q}(Y_{m_{l-1}+1}, U_{m_{l-1}})\cdots\overline{Q}(Y_{m_l}, U_{m_l-1})]_{i_2 j}\} = 1 \tag{9}$$

for each $i, i_1, i_2, j, j_1, j_2 \in \boldsymbol{X}$. Let $\{\hat{\theta}_t\}_{t=0}^{\infty}$, $\hat{\theta}_t \in \Theta$, be a sequence of estimates of the θ and satisfy $\hat{\theta}_t = \hat{\theta}_{m_l+1}$ for

$m_l + 1 \leq t \leq m_{l+1}$. That is, $\hat{\theta}_t$ is updated only at time $t = m_l + 1$, $l \in \mathbb{N}_0$. The assumption on the properties of the sequence $\{\hat{\theta}_t\}_{t=0}^{\infty}$ is made in the following.

Assumption 3.3: The parameter space Θ is compact and the transition matrix $Q(y, u, \cdot)$ is continuously differentiable on Θ for every $y \in \boldsymbol{Y}$ and $u \in \boldsymbol{U}$.

Assumption 3.4: The sequence $\{\hat{\theta}_t\}_{t=0}^{\infty}$ of estimates of θ satisfies

1) $\hat{\theta}_t$ is $\sigma(Y_0, \cdots, Y_t)-$measurable.
2) $\hat{\theta}_t \to \theta$ as $t \to \infty$ in $\mathbb{P}_{\psi_0}^{\pi^a}$
3) There exists a constant M such that

$$\|\hat{\theta}_{m_{l+1}+1} - \hat{\theta}_{m_l+1}\|_1 \leq \frac{\overline{M}}{l+1}$$

for every $l \in \mathbb{N}_0$.

We are thus ready for the following theorem.

Theorem 3.7: Suppose Assumption 2.1, 3.2, 3.3 and 3.4 are satisfied, then for each $\psi_0 \in \boldsymbol{\Psi}$

$$\mathbb{E}_{\psi_0}^{\pi^a}\left[\|\psi_t - \hat{\psi}_t\|_1\right] \longrightarrow 0 \qquad \text{as} \quad t \to \infty\,.$$

Proof: Let $\{m_l\}_{l=0}^{n}$ be the sequence used in (9) and define the following: $B_0^{y^t} = I$ is the identity matrix with size $N_x \times N_x$. For $1 \leq i \leq n$ the multi-step transition matrix: $B_i^{y^t} := \overline{Q}(Y_{m_{i-1}+1}, U_{m_{i-1}}) \cdots \overline{Q}(Y_{m_i}, U_{m_i-1})$. For $i = n+1$ $B_i^{y^t} := \overline{Q}(Y_{m_{i-1}+1}, U_{m_{i-1}}) \cdots \overline{Q}(Y_t, U_{t-1})$. $\hat{B}_i^{y^t}$ is similarly defined with $\overline{Q}(Y, U)$ replaced by $\hat{Q}(Y, U)$ for $i = 1, \cdots n+1$ and $\hat{B}_0^{y^t} = I$. For $l = 1, \cdots n+1$, new information states are denoted by

$$\hat{\psi}_l := \frac{\psi_0 \hat{B}_0^{y^t} \cdots \hat{B}_{l-1}^{y^t}}{\psi_0 \hat{B}_0^{y^t} \cdots \hat{B}_{l-1}^{y^t}\mathbf{1}}\,,\ \hat{\hat{\psi}}_l := \frac{\hat{\psi}_l \hat{B}_l^{y^t}}{\hat{\psi}_l \hat{B}_l^{y^t}\mathbf{1}}\,,\ \overline{\hat{\psi}}_l := \frac{\hat{\psi}_l B_l^{y^t}}{\hat{\psi}_l B_l^{y^t}\mathbf{1}}\,.$$

Then, by triangular inequality we have with probability 1

$$\begin{aligned}
&\left\|\psi_t - \hat{\psi}_t\right\|_1 \\
&\leq \sum_{l=1}^{n+1}\left\|\frac{\psi_0 \hat{B}_0^{y^t} \cdots \hat{B}_{l-1}^{y^t} B_l^{y^t} \cdots B_{n+1}^{y^t}}{\psi_0 \hat{B}_0^{y^t} \cdots \hat{B}_{l-1}^{y^t} B_l^{y^t} \cdots B_{n+1}^{y^t}\mathbf{1}} - \frac{\psi_0 \hat{B}_1^{y^t} \cdots \hat{B}_l^{y^t} B_{l+1}^{y^t} \cdots B_{n+1}^{y^t}}{\psi_0 \hat{B}_1^{y^t} \cdots \hat{B}_l^{y^t} B_{l+1}^{y^t} \cdots B_{n+1}^{y^t}\mathbf{1}}\right\|_1 \\
&= \sum_{l=1}^{n}\left\|\frac{\overline{\hat{\psi}}_l B_{l+1}^{y^t} \cdots B_{n+1}^{y^t}}{\overline{\hat{\psi}}_l B_{l+1}^{y^t} \cdots B_{n+1}^{y^t}\mathbf{1}} - \frac{\hat{\hat{\psi}}_l B_{l+1}^{y^t} \cdots B_{n+1}^{y^t}}{\hat{\hat{\psi}}_l B_{l+1}^{y^t} \cdots B_{n+1}^{y^t}\mathbf{1}}\right\|_1 + \left\|\overline{\hat{\psi}}_{n+1} - \hat{\hat{\psi}}_{n+1}\right\|_1 . \qquad (10)
\end{aligned}$$

If Assumption 3.2 is satisfied, then by Lemma 3.4 and Lemma 3.6-(2) we have

$$\begin{aligned}
(10) &\leq \frac{2}{\varepsilon^{N_b}}\left\{\sum_{l=1}^{n-1}\left\|\frac{\overline{\hat{\psi}}_l B_{l+1}^{y^t} \cdots B_n^{y^t}}{\overline{\hat{\psi}}_l B_{l+1}^{y^t} \cdots B_n^{y^t}\mathbf{1}} - \frac{\hat{\hat{\psi}}_l B_{l+1}^{y^t} \cdots B_n^{y^t}}{\hat{\hat{\psi}}_l B_{l+1}^{y^t} \cdots B_n^{y^t}\mathbf{1}}\right\|_1 + \sum_{l=n}^{n+1}\left\|\overline{\hat{\psi}}_l - \hat{\hat{\psi}}_l\right\|_1\right\} \\
&\leq \frac{2}{\varepsilon^{N_b}}\left\{\sum_{l=1}^{n-1}\frac{N_x}{\varepsilon}(1-\varepsilon^2)^{n-1-l}\left\|\overline{\hat{\psi}}_l - \hat{\hat{\psi}}_l\right\|_1 + \sum_{l=n}^{n+1}\left\|\overline{\hat{\psi}}_l - \hat{\hat{\psi}}_l\right\|_1\right\} \\
&= \frac{2N_x}{\varepsilon^{N_b+1}(1-\varepsilon^2)}\left\{\sum_{l=1}^{n-1}(1-\varepsilon^2)^{n-l}\left\|\overline{\hat{\psi}}_l - \hat{\hat{\psi}}_l\right\|_1 + \sum_{l=n}^{n+1}\left\|\overline{\hat{\psi}}_l - \hat{\hat{\psi}}_l\right\|_1\right\} \\
&= \frac{2N_x}{\varepsilon^{N_b+1}(1-\varepsilon^2)}\left\{\sum_{l=1}^{n}(1-\varepsilon^2)^{n-l}\left\|\overline{\hat{\psi}}_l - \hat{\hat{\psi}}_l\right\|_1 + \left\|\overline{\hat{\psi}}_{n+1} - \hat{\hat{\psi}}_{n+1}\right\|_1\right\}. \qquad (11)
\end{aligned}$$

Due to the continuous differentiability of $Q(y, u, \cdot)$ on Θ for each $y \in \boldsymbol{Y}$ and $u \in \boldsymbol{U}$, also, for $k = m_l + 1, \cdots, m_{l+1}$ $\hat{\theta}_k = \hat{\theta}_{m_l+1}$, we apply the mean value theorem and write that there exists a finite positive constant M such that

$$\begin{aligned}
\|\overline{\hat{\psi}}_l - \hat{\hat{\psi}}_l\|_1 &\leq M\sum_{k=m_{l-1}+1}^{m_l}\left\|\hat{\theta}_k - \theta\right\|_1 \\
&\leq MN_b\left\|\hat{\theta}_{m_{l-1}+1} - \theta\right\|_1
\end{aligned}$$

for $l = 1, \cdots, n$. With the same reason

$$\begin{aligned}
\left\|\overline{\hat{\psi}}_{n+1} - \hat{\hat{\psi}}_{n+1}\right\|_1 &\leq M\sum_{k=m_n+1}^{t}\left\|\hat{\theta}_k - \theta\right\|_1 \\
&\leq MN_b\left\|\hat{\theta}_{m_{l-1}+1} - \theta\right\|_1 .
\end{aligned}$$

Therefore, following (11) there exist positive and finite numbers M_1, M_2 and $\alpha \in (0, 1)$ such that

$$\begin{aligned}
&\left\|\psi_t - \hat{\psi}_t\right\|_1 \\
&\leq M_1\sum_{l=1}^{n}\alpha^{n-l}\left\|\hat{\theta}_{m_{l-1}+1} - \theta\right\|_1 + M_2\left\|\hat{\theta}_{m_n+1} - \theta\right\|_1 \\
&= M_1\sum_{l=0}^{n-1}\alpha^{n-1-l}\left\|\hat{\theta}_{m_l+1} - \theta\right\|_1 + M_2\left\|\hat{\theta}_{m_n+1} - \theta\right\|_1 . \qquad (12)
\end{aligned}$$

Applying the triangular inequality again we obtain

$$\begin{aligned}
&\left\|\hat{\theta}_{m_l+1} - \theta\right\|_1 \\
&\leq \left\|\hat{\theta}_{m_n+1} - \theta\right\|_1 + \left\|\hat{\theta}_{m_{n-1}+1} - \hat{\theta}_{m_n+1}\right\|_1 + \cdots + \left\|\hat{\theta}_{m_l+1} - \hat{\theta}_{m_{l+1}+1}\right\|_1 \\
&= \left\|\hat{\theta}_{m_n+1} - \theta\right\|_1 + \sum_{i=l}^{n-1}\left\|\hat{\theta}_{m_{i+1}+1} - \hat{\theta}_{m_i+1}\right\|_1 \\
&\leq \left\|\hat{\theta}_{m_n+1} - \theta\right\|_1 + \sum_{i=l}^{n-1}\frac{\overline{M}}{i+1}\,, \qquad (13)
\end{aligned}$$

for $l = 0, 1, \cdots, n-1$, where the last inequality follows from Assumption 3.4-(3). On the other hand,

$$\sum_{l=0}^{n-1}\sum_{i=l}^{n-1}\frac{\alpha^{n-1-l}}{i+1} = \frac{1}{1-\alpha}\left\{\sum_{i=0}^{n-1}\frac{\alpha^i}{n-i} - \alpha^n\sum_{i=1}^{n}\frac{1}{i}\right\}$$
$$\leq \frac{1}{1-\alpha}\sum_{i=0}^{n-1}\frac{\alpha^i}{n-i}$$
$$\leq \frac{1}{1-\alpha}\sum_{i=0}^{n-1}\frac{1+i}{n}\alpha^i \leq \frac{1}{n(1-\alpha)^3}\,, \tag{14}$$

where the second inequality follows from the fact that for $0 \leq i \leq n-1$ we have

$$\frac{n}{n-i} = 1 + \frac{i}{n-i} \leq 1 + i\,.$$

Finally, we obtain from (12) $\sim$ (14) that there exist positive and finite numbers M_3 and M_4 such that

$$E^{\pi^a}_{\psi_0}\left[\|\psi_t - \hat{\psi}_t\|_1\right] \leq M_3 E^{\pi^a}_{\psi_0}\left[\|\hat{\theta}_{m_n+1} - \theta\|_1\right] + \frac{M_4}{n}\,. \tag{15}$$

As $t \to \infty$, $n \geq t/N_b \to \infty$, we conclude that for each $\psi_0 \in \boldsymbol{\Psi}$

$$E^{\pi^a}_{\psi_0}\left[\|\psi_t - \hat{\psi}_t\|_1\right] \longrightarrow 0 \qquad \text{as} \quad t \to \infty$$

by Assumption 3.4-(2), the compactness of Θ, and inequality (15). ∎

In the following, we show the self-optimizing property for the adaptive policy π^a.

Theorem 3.8: Suppose Assumption 2.1, 3.2, 3.3 and 3.4 hold. For a given unknown true parameter vector $\theta \in \Theta$ the adaptive policy π^a described in (7) is self-optimizing in the long-run average sense.

Proof: By Theorem 2.2, under the assumptions we have for each $\theta \in \Theta$ a bounded solution (ρ_θ, h_θ) for equation (6) in parameterized form. Furthermore $h_\theta(\psi)$ is continuous and bounded both in $\psi \in \boldsymbol{\Psi}$ and $\theta \in \Theta$. Then the discrepancy function

$$D_\theta(\psi, u) := \tilde{c}(\psi, u) + \int_{\boldsymbol{Y}} h_\theta(\psi')\mathcal{K}(d\psi'|\psi, u, \theta) - \rho_\theta - h_\theta(\psi),$$

is uniformly continuous and bounded in $\Theta \times \boldsymbol{\Psi}$ for each $u \in \boldsymbol{U}$. Assumption 3.4 together with Theorem 3.7 imply for each $u \in \boldsymbol{U}$

$$D_{\hat{\theta}_t}(\hat{\psi}_t, u) \longrightarrow D_\theta(\psi_t, u) \quad \text{in} \quad \mathbb{P}^{\pi^a}_{\psi_0} \quad \text{as} \quad t \to \infty.$$

Due to the finiteness of $\boldsymbol{U}$ we can write as $t \to \infty$

$$D_{\hat{\theta}_t}(\hat{\psi}_t, \pi^*(\hat{\psi}_t, \hat{\theta}_t)) \longrightarrow D_\theta(\psi_t, \pi^*(\hat{\psi}_t, \hat{\theta}_t))) \quad \text{in} \quad \mathbb{P}^{\pi^a}_{\psi_0} \tag{16}$$

Since $\pi^*(\cdot, \theta)$ is optimal for $\theta \in \Theta$, we have

$$D_{\hat{\theta}_t}(\hat{\psi}_t, \pi^*(\hat{\psi}_t, \hat{\theta}_t)) = 0.$$

Define for arbitrary $\varepsilon > 0$ and $t \in \mathbb{N}$,

$$\Omega_t(\varepsilon) = \{\omega : |D_\theta(\psi_t, \pi^*(\hat{\psi}_t, \hat{\theta}_t)) - D_{\hat{\theta}_t}(\hat{\psi}_t, \pi^*(\hat{\psi}_t, \hat{\theta}_t))|(\omega) > \varepsilon\}$$
$$= \{\omega : D_\theta(\psi_t, \pi^*(\hat{\psi}_t, \hat{\theta}_t))(\omega) > \varepsilon\}\,.$$

$$E^{\pi^a}_{\psi_0}\{D_\theta(\psi_t, \pi^*(\hat{\psi}_t, \hat{\theta}_t))\} = \int_{\Omega_t(\varepsilon)} D_\theta(\psi_t, \pi^*(\hat{\psi}_t, \hat{\theta}_t))d\mathbb{P}^{\pi^a}_{\psi_0} + \int_{\Omega\setminus\Omega_t(\varepsilon)} D_\theta(\psi_t, \pi^*(\hat{\psi}_t, \hat{\theta}_t))d\mathbb{P}^{\pi^a}_{\psi_0}$$
$$\leq K\mathbb{P}^{\pi^a}_{\psi_0}(\Omega_t(\varepsilon)) + \varepsilon$$

for some finite $K > 0$. By (16) and letting $\varepsilon \to 0$, we have

$$E^{\pi^a}_{\psi_0}\{D_\theta(\psi_t, \pi^*(\hat{\psi}_t, \hat{\theta}_t))\} \longrightarrow 0 \qquad \text{as} \quad t \to \infty\,.$$

Therefore,

$$\frac{1}{T}\sum_{t=0}^{N-1} E^{\pi^a}_{\psi_0}\{D_\theta(\psi_t, \pi^*(\hat{\psi}_t, \hat{\theta}_t))\} \longrightarrow 0 \qquad \text{as} \quad t \to \infty$$

and the result follows from Theorem 2.2. ∎

IV. CONCLUSION

In this paper we address the adaptive control problems of a class of discrete-time partially observed Markov decision processes whose transition kernels are parameterized by a unknown vector. Suppose a sequence of parameter estimates converging to the true value with probability 1 is given, we propose an adaptive control policy and follow the methodology in [14] to show that under some conditions this policy is self-optimizing in the long-run average sense. The major conditions, including the property on the structure of transition matrices and the convergence speed of the parameter estimates to the true value, are shown to be the sufficient conditions to justify the self-optimality of the adaptive policy. We note that our conditions are either weaker or easily verifiable and thus can be of practical interest.

APPENDIX

Proof: When $\psi_1, \psi_2 \in \boldsymbol{\Psi}$ and B_k is row-allowable for $k \in \mathbb{N}$,

$$\left\|\frac{\psi_1 B^n}{\psi_1 B^n \mathbf{1}} - \frac{\psi_2 B^n}{\psi_2 B^n \mathbf{1}}\right\|_1 = \sum_{i=1}^{N_x}\left|\sum_{s=1}^{N_x}\frac{\psi_{1s}B^n_{si}}{\psi_1 B^n\mathbf{1}} - \sum_{s=1}^{N_x}\frac{\psi_{2s}B^n_{si}}{\psi_2 B^n\mathbf{1}}\right|$$
$$:= \sum_{i=1}^{N_x}\left|\sum_{s=1}^{N_x}(\hat{\psi}_{1s} - \hat{\psi}_{2s})\hat{B}^n_{si}\right| = \left\|(\hat{\psi}_1 - \hat{\psi}_2)\hat{B}^n\right\|_1 \leq \tau_1(\hat{B}^n)\left\|\hat{\psi}_1 - \hat{\psi}_2\right\|_1$$

where

$$\hat{\psi}_{1s} := \frac{\psi_{1s}B^n_{s\cdot}\mathbf{1}}{\psi_1 B^n\mathbf{1}}\,,\ \hat{\psi}_{2s} := \frac{\psi_{2s}B^n_{s\cdot}\mathbf{1}}{\psi_2 B^n\mathbf{1}}\,,\ \hat{B}^n_{si} := \frac{B^n_{si}}{B^n_{s\cdot}\mathbf{1}}\,.$$

Since

$$\left\|\hat{\psi}_1 - \hat{\psi}_2\right\|_1 \leq \left\{\left\|\hat{\psi}_1 - \hat{\psi}_0\right\|_1 + \left\|\hat{\psi}_2 - \hat{\psi}_0\right\|_1\right\}$$

where $\hat{\psi}_0 = [\hat{\psi}_{01}, \hat{\psi}_{02}, \cdots, \hat{\psi}_{0N_x}]$ with s^{th} component $\frac{\psi_{2s}B^n_{s.}\mathbf{1}}{\psi_1 B^n \mathbf{1}}$, so

$$\begin{aligned}\left\|\hat{\psi}_2 - \hat{\psi}_0\right\|_1 &= \sum_{s=1}^{N_x}\left|\frac{\psi_{2s}B^n_{s.}\mathbf{1}}{\psi_2 B^n\mathbf{1}} - \frac{\psi_{2s}B^n_{s.}\mathbf{1}}{\psi_1 B^n\mathbf{1}}\right| \\ &= \sum_{s=1}^{N_x}\left|\frac{(\psi_1-\psi_2)B^n\mathbf{1}\psi_{2s}B^n_{s.}\mathbf{1}}{\psi_1 B^n\mathbf{1}\psi_2 B^n\mathbf{1}}\right| \\ &= \left|\frac{(\psi_1-\psi_2)B^n\mathbf{1}}{\psi_1 B^n\mathbf{1}}\right| \le \sum_{s=1}^{N_x}\left|\frac{\psi_{1s}B^n_{s.}\mathbf{1}}{\psi_1 B^n\mathbf{1}} - \frac{\psi_{2s}B^n_{s.}\mathbf{1}}{\psi_1 B^n\mathbf{1}}\right| \\ &= \left\|\hat{\psi}_1 - \hat{\psi}_0\right\|_1 .\end{aligned}$$

That is,

$$\begin{aligned}\left\|\hat{\psi}_1 - \hat{\psi}_2\right\|_1 &\le 2\sum_{s=1}^{N_x}\left|\frac{(\psi_{1s}-\psi_{2s})B^n_{s.}\mathbf{1}}{\psi_1 B^n\mathbf{1}}\right| \\ &\le 2\sum_{s=1}^{N_x}|\psi_{1s}-\psi_{2s}|\cdot\max_s\left\{\frac{B^n_{s.}\mathbf{1}}{\psi_1 B^n\mathbf{1}}\right\} .\end{aligned}$$

Since $\tau_1(B) \le 1$ for a stochastic matrix B, if given $(B_k)_{i_1.}\mathbf{1} \ge \varepsilon\cdot(B_k)_{i_2.}\mathbf{1}$, then from Lemma 3.4

$$\left\|\hat{\psi}_1 - \hat{\psi}_2\right\|_1 \le \frac{2}{\varepsilon^n}\left\|\psi_1-\psi_2\right\|_1$$

and part (1) is proved. If $(B_k)_{i_1 j} \ge \varepsilon\cdot(B_k)_{i_2 j}$, then from Lemma 3.5

$$\begin{aligned}\tau_1(\hat{B}^n) &= \frac{1}{2}\max_{i_1,i_2}\sum_{j=1}^{N_x}\left|\hat{B}^n_{i_1 j} - \hat{B}^n_{i_2 j}\right| \\ &= \frac{1}{2}\max_{i_1,i_2}\sum_{j=1}^{N_x}\left|\frac{B^n_{i_1 j}}{B^n_{i_1.}\mathbf{1}} - \frac{B^n_{i_2 j}}{B^n_{i_2.}\mathbf{1}}\right| \\ &\le \frac{N_x}{2}(1-\varepsilon^2)^{n-1}\end{aligned}$$

and by Corollary 3.3

$$\left\|\hat{\psi}_1 - \hat{\psi}_2\right\|_1 \le \frac{2}{\varepsilon}\left\|\psi_1-\psi_2\right\|_1 ,$$

therefore part (2) is proved. ■

REFERENCES

[1] A. Arapostathis, S. I. Marcus. Analysis of an identification algorithm arising in the adaptive estimation of Markov chains. *Mathematics of Control, Signal and System*, 3:1–29, 1990.

[2] A. Arapostathis, V. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus. Discrete-time controlled Markov processes with average cost criterion: a survey. *SIAM Journal on Control & Optimization*, 31:282–344, 1993.

[3] R. Atar and O. Zeitouni. Exponential stability for nonlinear filtering. *Probabilites et Statistiques*, 33:697–725, 1997.

[4] R. Bellman. *Dynamic Programming*. Princeton University Press, 1957.

[5] G. Birkhoff. Extensions of jentzsch's theorem. *Transactions of the American Mathematical Society*, 85:219–227, 1957.

[6] D. P. Bertsekas. *Dynamic programming and stochastic control*. Academic Press, 1976.

[7] D. P. Bertsekas and S. E. Shreve. *Stochastic optimal control: the discrete time case*. Academic Press, 1978.

[8] V. S. Borkar. Ergodic control of partially observed Markov chains. *System & Control Letters*, 99:185–189, 1998.

[9] D.-M. Chuang and A. Arapostathis. Some new results on the ergodic control of partially observed Markov chains. *Proceedings of the 38th IEEE conference on decision and control*, 1908–1909, 1999.

[10] D.-M. Chuang. *Risk-sensitive control of discrete-time partially observed Markov decision processes*. Ph.D Dissertation, The University of Texas at Austin, 1999.

[11] T. E. Duncan, B. Pasik-Duncan, and L. Stettner. Adaptive control of a partially observed discrete time Markov process. *Applied Mathematics and Optimization*, 37:269–293, 1998.

[12] E. Fernández-Gaucherand, A. Arapostathis, and S. I. Marcus. On the adaptive control of partially observable Markov decision processes. *Proceedings of the 27th IEEE Conference Decision and Control*, 1204–1210, 1988.

[13] E. Fernández-Gaucherand. *Controlled Markov processes on the infinite planning horizon: optimal & adaptive control*. Ph.D Disertation, The University of Texas at Austin, 1991.

[14] E. Fernández-Gaucherand, A. Arapostathis, and S. I. Marcus. A methodology for the adaptive control of Markov chains under partial state information. *Proceedings of the 31th IEEE Conference Decision and Control*, 2750–2752, 1992.

[15] E. Fernández-Gaucherand, A. Arapostathis, and S. I. Marcus. Convex stochastic control problems. *Proceedings of the 31th IEEE Conference Decision and Control*, 2179–2180, 1992.

[16] E. Fernández-Gaucherand, A. Arapostathis, and S. I. Marcus. Analysis of an adaptive control scheme for a partially observed controlled Markov chain. *IEEE transactions on Automatic Control*, 38:987–993, 1993.

[17] O. Hernández-Lerma. *Adaptive Markov control processes*. Springer Verlag, 1989.

[18] O. Hernández-Lerma, J. B. Lasserre. *Discrete-Time Markov control processes*. Springer Verlag, 1996.

[19] S.-P. Hsu, D.-M. Chuang and A. Arapostathis. On the existence of stationary optimal policies for partially observed MDPs under the long-run average cost criterion. *Systems and Control Letters*, 55:165–173, 2006.

[20] P. R. Kumar and P. Varaiya. *Stochastic Systems: estimation, identification and adaptive control*. Prentice-Hall, 1986.

[21] H. J. Kushner and C. G. Yin. *Stochastic approximation algorithm and applications*. Springer-Verlag, New York, 1997.

[22] L. K. Platzman. Optimal infinite-horizon undiscounted control of finite probabilistic systems. *SIAM Journal on Control & Optimization*, 18:362–380, 1980.

[23] W. J. Runggaldier and L. Stettner. *Approximations of discrete time partially observed control problems*. Applied Mathematics Monographs, 6, 1994, Giardini Editori E Stampatori in Pisa, Italy.

[24] E. Seneta. *Non-negative matrices and Markov chains*. Springer Veralg, 1981.