Chapter 18
Optimality Conditions for Partially Observable Markov Decision Processes

Eugene A. Feinberg, Pavlo O. Kasyanov and Mikhail Z. Zgurovsky

Abstract This note describes sufficient conditions for the existence of optimal policies for Partially Observable Markov Decision Processes (POMDPs). The objective criterion is either minimization of total discounted costs or minimization of total nonnegative costs. It is well-known that a POMDP can be reduced to a Completely Observable Markov Decision Process (COMDP) with the state space being the sets of believe probabilities for the POMDP. Thus, a policy is optimal in POMDP if and only if it corresponds to an optimal policy in the COMDP. Here we provide sufficient conditions for the existence of optimal policies for COMDP and therefore for POMDP.

18.1 Introduction

Partially Observable Markov Decision Processes (POMDPs) play an important role in electrical engineering, computer science, and operations research. They have a broad range of applications including sensor networks, artificial intelligence, control and maintenance of complex systems, and medical decision making. In principle, by ignoring complexity issues, it is known how to solve POMDPs. A POMDP can be reduced to a Completely Observable Markov Decision Process (COMDP) with the state space being the sets of believe probabilities for the POMDP [2, 6, 9, 10]. After an optimal policy for the COMDP is found, it can be used to compute an optimal

E. A. Feinberg (✉)
Department of Applied Mathematics and Statistics,
Stony Brook University, Stony Brook, NY 11794-3600, USA
e-mail: eugene.feinberg@sunysb.edu

P. O. Kasyanov (✉) · M. Z. Zgurovsky
Institute for Applied System Analysis, National Technical University of Ukraine
"Kyiv Polytechnic Institute", Peremogy ave., 37, build, 35, Kyiv 03056, Ukraine
e-mail: kasyanov@i.ua

M. Z. Zgurovsky and V. A. Sadovnichiy (eds.), *Continuous and Distributed Systems*,
Solid Mechanics and Its Applications 211, DOI: 10.1007/978-3-319-03146-0_18,

policy for the POMDP. However, except the case of problems with finite state and action sets and a large variety of particular problems considered in the literature, little is known regarding the existence and properties of optimal policies for COMDPs in terms of the original POMDP.

This problem is studied in Hernández-Lerma [6, Chap. 4], where sufficient conditions for the existence of optimal policies for discounted POMDPs with Borel state spaces, compact action sets, weakly continuous transition and observation probabilities, and bounded continuous cost functions are provided. It is shown there that the weak continuity of the transition kernel in the filtration equation is sufficient for the existence of optimal policies for COMDPs and therefore for COMDPs. A sufficient condition for the case of a countable observation case is also provided in Hernández-Lerma [6, p. 92]. This condition is that the probability of observations depend continuously on the state-action pairs. Since this is the condition for a countable observation space, in the case of general Borel observation spaces, there are three continuity conditions on the observation probabilities that are equivalent to this condition, when the observation space becomes countable. These conditions are weak continuity, setwise continuity, and continuity in the total variation of observation probabilities (also called kernels or stochastic kernel).

In this paper, we a we study either minimization of expected total nonnegative costs or discounted costs with the one-step cost functions bounded below for POMDPs with Borel state spaces. The goal is to obtain sufficient conditions for the existence and characterization of optimal policies for COMDPs with possibly non-compact action sets, unbounded cost functions (they are assumed bounded below), and uncountable observation sets. The one-step cost functions are K-infcompact. The notion of K-infcompactness was introduced recently in Feinberg, Kasyanov, and Zadoianchuk [3]. As shown in Feinberg, Kasyanov, and Zadoianchuk [4], this mild condition and weak continuity of transition probabilities are sufficient for the existence of optimal policies and their characterization for fully observable Markov Decision Processes (MDPs) with the expected total costs.

Of course, for the existence of optimal policies for a POMDP, additional conditions are required for the transition observation probability. Here we show that the sufficient condition is its continuity in the total variation of the observation transition probability. We also provide a general criterion for the existence optimal policies for weakly continuous transition observation probabilities, which is different from the weak continuity of the filtration kernel considered in Hernández-Lerma [6, p. 90, Assumption 4.1(d)].

18.2 Model Description

For a metric space $\mathbb{S}$, let $\mathscr{B}(\mathbb{S})$ be its Borel σ-field, that is, the σ-field generated by all open sets of the metric space $\mathbb{S}$. For a Borel subset $E \subset \mathbb{S}$, we denote by $\mathscr{B}(E)$ the σ-field whose elements are intersections of E with elements of $\mathscr{B}(\mathbb{S})$. Observe that E is a metric space with the same metric as on $\mathbb{S}$, and $\mathscr{B}(E)$ is its Borel σ-field.

The space E is a *Borel space*, if E is a Borel subset of a Polish (complete separable metric) space $\mathbb{S}$. On E consider the induced metrizable topology. For a metric space $\mathbb{S}$, we denote by $\mathbb{P}(\mathbb{S})$ the *set of probability measures* on $(\mathbb{S}, \mathscr{B}(\mathbb{S}))$. A sequence of probability measures $\{\mu_n\}$ from $\mathbb{P}(\mathbb{S})$ *converges weakly (setwise)* to $\mu \in \mathbb{P}(\mathbb{S})$ if for any bounded continuous (bounded Borel-measurable) function f on $\mathbb{S}$

$$\int_{\mathbb{S}} f(s)\mu_n(ds) \to \int_{\mathbb{S}} f(s)\mu(ds) \qquad \text{as} \quad n \to \infty.$$

A sequence of probability measures $\{\mu_n\}$ from $\mathbb{P}(\mathbb{S})$ *converges in the total variation* to $\mu \in \mathbb{P}(\mathbb{S})$ if

$$\sup_{f \in \mathbb{F}_1(\mathbb{S})} \left\{ \int_{\mathbb{S}} f(s)\mu_n(ds) - \int_{\mathbb{S}} f(s)\mu(ds) \right\},$$

where $F_1(\mathbb{S})$ is the set of Borel-measurable functions on $\mathbb{S}$ such that $|f(s)| \le 1$ for all $s \in \mathbb{S}$.

Note that $\mathbb{P}(\mathbb{S})$ is a Polish space with respect to the weak convergence topology for probability measures; Parthasarathy [8, Chap. 2]. For Borel spaces $\mathbb{S}_1$ and $\mathbb{S}_2$, a (Borel-measurable) *transition kernel* $R(ds_1|s_2)$ on $\mathbb{S}_1$ given $\mathbb{S}_2$ is a mapping $R(\cdot\,|\,\cdot) : \mathscr{B}(\mathbb{S}_1) \times \mathbb{S}_2 \to [0, 1]$, such that $R(\cdot\,|s_2)$ is a probability measure on $\mathbb{S}_1$ for any $s_2 \in \mathbb{S}_2$, and $R(B|\cdot)$ is a Borel-measurable function on $\mathbb{S}_2$ for any Borel set $B \in \mathscr{B}(\mathbb{S}_1)$. A transition kernel $R(ds_1|s_2)$ on $\mathbb{S}_1$ given $\mathbb{S}_2$ defines a Borel measurable mapping $s_2 \to R(\cdot|s_1)$ of $\mathbb{S}_2$ to the metric space $\mathbb{P}(\mathbb{S}_1)$ endowed with the topology of weak convergence. A transition kernel $R(ds_1|s_2)$ on $\mathbb{S}_1$ given $\mathbb{S}_2$ is called *weakly continuous (setwise continuous, continuous in the total variation)*, if $R(\cdot\,|x_n)$ converges weakly (setwise, in the total variation) to $R(\cdot\,|x)$ whenever x_n converges to x in $\mathbb{S}_2$.

Let $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{A}$ be Borel spaces, $P(dx'|x, a)$ is a transition kernel on $\mathbb{X}$ given $\mathbb{X} \times \mathbb{A}$, $Q(dy|a, x)$ is a transition kernel on $\mathbb{Y}$ given $\mathbb{A} \times \mathbb{X}$, $Q_0(dy|x)$ is a transition kernel on $\mathbb{Y}$ given $\mathbb{X}$, p_0 is a probability distribution on $\mathbb{X}$, $c : \mathbb{X} \times \mathbb{A} \to \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ is a bounded from below Borel function on $\mathbb{X} \times \mathbb{A}$.

Partially observable Markov decision process (*POMDP*) *is specified by* $(\mathbb{X}, \mathbb{Y}, \mathbb{A}, P, Q, c)$, *where* $\mathbb{X}$ *is the state space*, $\mathbb{Y}$ *is the observation set*, $\mathbb{A}$ *is the action set*, $P(dx'|x, a)$ *is the state transition law*, $Q(dy|a, x)$ *is the observation kernel*, $c : \mathbb{X} \times \mathbb{A} \to \overline{\mathbb{R}}$ *is the one-step cost*.

The partially observable Markov decision process evolves as follows:

- at time $t = 0$, the initial unobservable state x_0 has a given prior distribution p_0;
- the initial observation y_0 is generated according to the initial observation kernel $Q_0(\cdot\,|x_0)$;
- at each time epoch $n = 0, 1, 2, \ldots$, if the state of the system is $x_n \in \mathbb{X}$ and the decision-maker chooses an action $a_n \in \mathbb{A}$, then the cost $c(x_n, a_n)$ is incurred;
- the system moves to state x_{n+1} according to the transition law $P(\cdot\,|x_n, a_n)$;
- the observation $y_{n+1} \in \mathbb{Y}$ is generated by the observation kernels $Q(\cdot\,|a_n, x_{n+1})$, $n = 0, 1, \ldots$, and $Q_0(\cdot\,|x_0)$.

Define the *observable histories*: $h_0 := (p, y_0) \in H_0$ and $h_n := (p, y_0, a_0, \ldots, y_{n-1}, a_{n-1}, y_n) \in H_n$ for all $n = 1, 2, \ldots$, where $\mathbb{H}_0 := \mathbb{P}(\mathbb{X}) \times \mathbb{Y}$ and $\mathbb{H}_n := \mathbb{H}_{n-1} \times \mathbb{A} \times \mathbb{Y}$ if $n = 1, 2, \ldots$. Then a *policy* for the POMDP is defined as a sequence $\pi = \{\pi_n\}$ such that, for each $n = 0, 1, \ldots$, π_n is a transition kernel on $\mathbb{A}$ given $\mathbb{H}_n$. Moreover, π is called *nonrandomized*, if each probability measure $\pi_n(\cdot|h_n)$ is concentrated at one point. A nonrandomized policy is called *Markov*, if all of the decisions depend on the current state and time only. A Markov policy is called *stationary*, if all the decisions depend on the current state only. The *set of all policies* is denoted by Π. The Ionescu Tulcea theorem (Bertsekas and Shreve [1, pp. 140–141] or Hernández-Lerma and Lassere [7, p. 178]) implies that a policy $\pi \in \Pi$ and an initial distribution $p_0 \in \mathbb{P}(\mathbb{X})$, together with the transition kernels P, Q and Q_0 determine a unique probability measure $P^{\pi}_{p_0}$ on the set of all trajectories $\mathbb{H}_\infty = \mathbb{P}(\mathbb{X}) \times (\mathbb{Y} \times \mathbb{A})^\infty$ endowed with the product of σ-field defined by Borel σ-field of $\mathbb{P}(\mathbb{X})$, $\mathbb{Y}$, and $\mathbb{A}$ respectively. The expectation with respect to this probability measure is denoted by $E^{\pi}_{p_0}$.

Let us specify a performance criterion. For a finite horizon $N = 0, 1, \ldots$, and for a policy $\pi \in \Pi$, let us define the *expected total discounted costs*

$$v^{\pi}_{N,\alpha}(p) := \mathbb{E}^{\pi}_{p} \sum_{n=0}^{N-1} \alpha^n c(x_n, a_n), \ p \in \mathbb{P}(\mathbb{X}), \tag{18.1}$$

where $\alpha \geq 0$ is the discount factor, $v^{\pi}_{0,\alpha}(p) = 0$. When $N = \infty$, we always assume that at least one of the following two assumptions holds:

Assumption (D) c is bounded below on $\mathbb{X} \times \mathbb{A}$ and $\alpha \in [0, 1]$.
Assumption (P) c is nonnegative on $\mathbb{X} \times \mathbb{A}$ and $\alpha \in [0, 1]$.

In the both cases (18.1) defines an *infinite horizon expected total discounted cost*, and we denote it by $v^{\pi}_{\alpha}(p)$. By using notations (**D**) and (**P**), we follow Bertsekas and Shreve [1, p. 214]. However, our Assumption (**D**) is weaker than the corresponding assumption in [1], because c was assumed to be bounded under Assumption (**D**) in [1].

Since the function c is bounded below on $\mathbb{X} \times \mathbb{A}$, a discounted model can be converted into a positive model by shifting the cost function. In particular, let $c(x, a) \geq -K$ for any $(x, a) \in \mathbb{X} \times \mathbb{A}$. Consider a new cost function $\hat{c}(x, a) := c(x, a) + K$ for any $(x, a) \in \mathbb{X} \times \mathbb{A}$. Then the corresponding total discounted reward is equal to

$$\hat{v}^{\pi}_{\alpha}(p) := v^{\pi}_{\alpha}(p) + \frac{K}{1-\alpha}, \qquad \pi \in \Pi, \ p \in \mathbb{P}(\mathbb{X}).$$

Thus, optimizing v^{π}_{α} and $\hat{v}^{\pi}_{\alpha}$ are equivalent problems, but $\hat{v}^{\pi}_{\alpha}$ is the objective function for the positive model. Though positive models are more general, discounted models are met in lager classes of applications. Thus we formulate the results for either of these models.

For any function $g^\pi(p)$, including $g^\pi(p) = v^\pi_{N,\alpha}(p)$ and $g^\pi(p) = v^\pi_\alpha(p)$ define the *optimal cost*

$$g(p) := \inf_{\pi \in \Pi} g^\pi(p), \qquad p \in \mathbb{P}(\mathbb{X}),$$

where Π is *the set of all policies*. A policy π is called *optimal* for the respective criterion, if $g^\pi(p) = g(p)$ for all $p \in \mathbb{P}(\mathbb{X})$. For $g^\pi = v^\pi_{n,\alpha}$, the optimal policy is called *n-horizon discount-optimal*; for $g^\pi = v^\pi_\alpha$, it is called *discount-optimal*.

We recall that a function c defined on $\mathbb{X} \times \mathbb{A}$ is inf-compact (or lower semi-compact) if the set $\{(x, a) \in \mathbb{X} \times \mathbb{A} : c(x, a) \le \lambda\}$ is compact for any finite number λ. A function c defined on $\mathbb{X} \times \mathbb{A}$ is called K-inf-compact on $\mathbb{X} \times \mathbb{A}$, if for any compact subset K of $\mathbb{X}$, the function c is inf-compact on $K \times \mathbb{A}$; Feinberg, Kasyanov, and Zadoianchuk [3, Definition 11]. K-inf-inf-compactness is a mild assumption that is weaker than inf-compactness. Essentially, K-inf-compactness of the cost function c is almost equivalent to lower-semicontinuity of c in the state variable x and lower semi-continuity in the action variable a. This property holds for many applications including inventory control and various problems with least square criteria. According to Feinberg, Kasyanov, and Zadoianchuk [3, Lemma 2.5], a bounded below function c is K-inf-compact on the product of metric spaces $\mathbb{X}$ and $\mathbb{A}$ if and only if it satisfies the following two conditions:

(a) c is lower semi-continuous;
(b) if a sequence $\{x_n\}_{n=1,2,\ldots}$ with values in $\mathbb{X}$ converges and its limit x belongs to $\mathbb{X}$ then any sequence $\{a_n\}_{n=1,2,\ldots}$ with $a_n \in \mathbb{A}$, $n = 1, 2, \ldots$, satisfying the condition that the sequence $\{\overline{c}(x_n, a_n)\}_{n=1,2,\ldots}$ is bounded above, has a limit point $a \in \mathbb{A}$.

As an POMDP $(\mathbb{X}, \mathbb{Y}, \mathbb{A}, P, Q, c)$, consider the classical MDP $(\mathbb{X}, \mathbb{A}, P, c)$, when all the states are observable. An MDP can be viewed as a particular POMDPs with $\mathbb{Y} = \mathbb{X}$ and $Q(B|a, x) = Q(B|x) = \mathbf{I}\{x \in B\}$ for all $x \in \mathbb{X}$, $a \in A$, and $\in \mathscr{B}(\mathbb{X})$. In fact, this POMP possesses a special property that action sets at all the states are equal. For MDPs, Feinberg, Kasyanov, and Zadoianchuk [4] the following general general conditions for the existence of optimal policies, validity of optimality equations, and convergence of value iterations. Here we formulate these conditions for an MDP whose action sets at different states are equal.

Assumption (W*) (cf. Feinberg, Kasyanov, and Zadoianchuk [4] and Lemma 2.5 in [3]).

(i) c is K-inf-compact on $\mathbb{X} \times \mathbb{A}$;
(ii) the transition probability $P(\cdot\,|x, a)$ is weakly continuous in $(x, a) \in \mathbb{X} \times \mathbb{A}$.

Theorem 18.1 (cf. Feinberg, Kasyanov, and Zadoianchuk [4, Theorem 2]). *Let MDP* $(\mathbb{X}, \mathbb{A}, P, c)$ *satisfies Assumption* $(\mathbf{W}^*)$. *Consider either positive or discounted model. Then:*

(*i*) *the functions* $v_{n,\alpha}$, $n = 0, 1, 2, \ldots$, *and* v_α *are lower semi-continuous on* $\mathbb{X}$, *and* $v_{n,\alpha}(x) \to v_\alpha(x)$ *as* $n \to \infty$ *for all* $x \in \mathbb{X}$;

(*ii*) *for any* $x \in \mathbb{X}$, *and* $n = 0, 1, \ldots,$

$$v_{n+1,\alpha}(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_{\mathbb{X}} v_{n,\alpha}(y) P(dy|x,a) \right\}, \tag{18.2}$$

where $v_{0,\alpha}(x) = 0$ *for all* $x \in \mathbb{X}$, *and the nonempty sets*

$$A_{n,\alpha}(x) := \{a \in \mathbb{A} : v_{n+1,\alpha}(x) = c(x,a) + \alpha \int_{\mathbb{X}} v_{n,\alpha}(y) P(dy|x,a)\}$$

satisfy the following properties: (a) the graph $\mathrm{Gr}(A_{n,\alpha}) = \{(x,a) : x \in \mathbb{X}, a \in A_{n,\alpha}(x)\}$, $n = 0, 1, \ldots,$ *is a Borel subset of* $\mathbb{X} \times \mathbb{A}$, *and (b) if* $v_{n+1,\alpha}(x) = \infty$, *then* $A_{n,\alpha}(x) = A(x)$ *and, if* $v_{n+1,\alpha}(x) < \infty$, *then* $A_{n,\alpha}(x)$ *is compact;*

(*iii*) *for any* $N = 1, 2, \ldots,$ *there exists a Markov optimal* N*-horizon policy* $(\phi_0, \ldots, \phi_{N-1})$ *and if, for an* N*-horizon Markov policy* $(\phi_0, \ldots, \phi_{N-1})$ *the inclusions* $\phi_{N-1-n}(x) \in A_{n,\alpha}(x)$, $x \in \mathbb{X}$, $n = 0, \ldots, N-1$, *hold then this policy is* N*-horizon optimal;*

(*iv*) *for* $\alpha \in [0, 1]$

$$v_\alpha(x) = \min_{a \in A(x)} \{c(x,a) + \alpha \int_{\mathbb{X}} v_\alpha(y) P(dy|x,a)\}, \qquad x \in \mathbb{X}, \tag{18.3}$$

and the nonempty sets

$$A_\alpha(x) := \{a \in \mathbb{A} : v_\alpha(x) = c(x,a) + \alpha \int_{\mathbb{X}} v_\alpha(y) P(dy|x,a)\}, \quad x \in \mathbb{X},$$

satisfy the following properties: (a) the graph $\mathrm{Gr}(A_\alpha) = \{(x,a) : x \in \mathbb{X}, a \in A_\alpha(x)\}$ *is a Borel subset of* $\mathbb{X} \times \mathbb{A}$, *and (b) if* $v_\alpha(x) = \infty$, *then* $A_\alpha(x) = A(x)$ *and, if* $v_\alpha(x) < \infty$, *then* $A_\alpha(x)$ *is compact;*

(*v*) *for an infinite-horizon there exists a stationary discount-optimal policy* ϕ_α, *and a stationary policy is optimal if and only if* $\phi_\alpha(x) \in A_\alpha(x)$ *for all* $x \in \mathbb{X}$;

(*vi*) (Feinberg and Lewis [5, Proposition 3.1(iv)]) *if* c *is inf-compact on* $\mathbb{X} \times \mathbb{A}$, *then the functions* $v_{n,\alpha}$, $n = 1, 2, \ldots,$ *and* v_α *are inf-compact on* $\mathbb{X}$.

18.3 Reduction of POMDPs to COMDPs and Optimality Results

In this section, we formulate the known reduction of a POMDP to the completely observable Markov decision process (COMDP). Based on general results for MDPs (Feinberg, Kasyanov, Zadoianchuk [4, Theorem 4.1], Theorem 18.2 states sufficient

conditions for the validity of the following results for the COMDP: the existence of stationary optimal policies, the validity of optimality equations, the characterization of optimal policies via optimality equations, and the convergence of value iterations. Then, we formulate the main result of this paper, Theorem 18.3, that states sufficient conditions of these properties in terms of the parameters of the original POMDP.

First, we formulate the well-known reduction of a POMDP to the COMDP ([1, 2, 6, 9, 11]). To simplify notations, we drop sometimes the time parameter. Given a posterior distribution z of the state x at time epoch $n = 0, 1, \ldots$ and given an action a selected at epoch n, denote by $R(B \times C|z, a)$ the joint probability that the state at time $(n + 1)$ belongs to the set $B \in \mathscr{B}(\mathbb{X})$ and the observation at time n belongs to the set $C \in \mathscr{B}(\mathbb{Y})$,

$$R(B \times C|z, a) := \int_{\mathbb{X}} \int_{B} Q(C|a, x')P(dx'|x, a)z(dx), \tag{18.4}$$

where R is a transition kernel on $\mathbb{X} \times \mathbb{Y}$ given $\mathbb{P}(\mathbb{X}) \times \mathbb{A}$; see Bertsekas and Shreve [1]; or Dynkin and Yushkevich [2]; or Hernández-Lerma [6]; or Yushkevich [11] for details. Therefore, the probability $R'(C|z, a)$ that the observation y at time n belongs to the set $C \in \mathscr{B}$ is

$$R'(C|z, a) = \int_{\mathbb{X}} \int_{\mathbb{X}} Q(C|a, x')P(dx'|x, a)z(dx), \tag{18.5}$$

where R' is a transition kernel on $\mathbb{Y}$ given $\mathbb{P}(\mathbb{X}) \times \mathbb{A}$. By Bertsekas and Shreve [1, Proposition 7.27], there exist a transition kernel H on $\mathbb{X}$ given $\mathbb{P}(\mathbb{X}) \times \mathbb{A} \times \mathbb{Y}$ such that

$$R(B \times C|z, a) = \int_{C} H(B|z, a, y)R'(dy|z, a), \tag{18.6}$$

The transition kernel $H(\cdot\,|z, a, y)$ defines a measurable mapping $H : \mathbb{P}(\mathbb{X}) \times \mathbb{A} \times \mathbb{Y} \to \mathbb{P}(\mathbb{X})$, where $H(z, a, y)[\,\cdot\,] = H(\cdot\,|z, a, y)$. For each pair $(z, a) \in \mathbb{P}(\mathbb{X}) \times \mathbb{A}$, the mapping $H(z, a, \cdot) : \mathbb{Y} \to \mathbb{P}(\mathbb{Y})$ is defined $R'(\cdot\,|z, a)$-a.s. uniquely in y; Dynkin and Yushkevich [2, p. 309]. It is known that for a posterior distribution $z_n \in \mathbb{P}(\mathbb{X})$, action $a_n \in A(x)$, and an observation $y_{n+1} \in \mathbb{Y}$, the posterior distribution $z_{n+1} \in \mathbb{P}(\mathbb{X})$ is

$$z_{n+1} = H(z_n, a_n, y_{n+1}). \tag{18.7}$$

However, the observation y_{n+1} is not available in the COMDP model, and therefore y_{n+1} is a random variable with the distribution $R'(\cdot\,|z_n, a_n)$, and (18.7) is a stochastic equation that maps $(z_n, a_n) \in \mathbb{P}(\mathbb{X}) \times \mathbb{A}$ to $\mathbb{P}(\mathbb{P}(\mathbb{X}))$. The stochastic kernel that defines the distribution of z_{n+1} on $\mathbb{P}(\mathbb{X})$ given $\mathbb{P}(\mathbb{X}) \times \mathbb{X}$ is defined uniquely as

$$q(D|z,a) := \int_{\mathbb{Y}} \mathbf{1}_D[H(z,a,y)]R'(dy|z,a), \tag{18.8}$$

where

$$\mathbf{1}_D[u] = \begin{cases} 1, u \in D \in \mathscr{B}(\mathbb{P}(\mathbb{X})), \\ 0, u \notin D \in \mathscr{B}(\mathbb{P}(\mathbb{X})); \end{cases}$$

Hernández-Lerma [7, p. 87]. The measurable particular choice of stochastic kernel H from (18.6) does not effect on the definition of q from (18.8), since for each pair $(z,a) \in \mathbb{P}(\mathbb{X}) \times \mathbb{A}$, the mapping $H(z,a,\cdot) : \mathbb{Y} \to \mathbb{P}(\mathbb{Y})$ is defined $R'(\cdot\,|z,a)$-a.s. uniquely in y; Dynkin and Yushkevich [2, p. 309].

The COMDP is defined as an MDP with parameters $(\mathbb{P}(\mathbb{X}), \mathbb{A}, q, \overline{c})$, where

(i) $\mathbb{P}(\mathbb{X})$ is the state space;
(ii) $\mathbb{A}$ is the action set available at all state $z \in \mathbb{P}(\mathbb{X})$;
(iii) the one-step cost function $\overline{c} : \mathbb{P}(\mathbb{X}) \times \mathbb{A} \to \overline{\mathbb{R}}$, defined as

$$\overline{c}(z,a) := \int_{\mathbb{X}} c(x,a)z(dx), \quad z \in \mathbb{P}(\mathbb{X}),\ a \in \mathbb{A}; \tag{18.9}$$

(iv) transition probabilities q on $\mathbb{P}(\mathbb{X})$ given $\mathbb{P}(\mathbb{X}) \times \mathbb{A}$ defined in (18.8).

see Bertsekas and Shreve [1, Corollary 7.27.1, p. 139] or Dynkin and Yushkevich [2, p. 215], or Hernández-Lerma [6] for details.

If a stationary optimal policy for the COMDP exists and found, it allows the decision maker to compute an optimal policy for the COMDP. First, we recall how the initial state distribution $z_0 \in \mathbb{P}(\mathbb{P}(X))$ can be computed for the COMDP. Similarly to transition kernels R, R', and H, consider a transition kernel

$$R_0(B \times C|p) := \int_B Q_0(C|x)p(dx), \quad B \in \mathscr{B}(\mathbb{X})$$

on $\mathbb{X} \times \mathbb{Y}$ given $\mathbb{P}(\mathbb{X})$. It can be decomposed as

$$R_0(B \times C|p) = \int_C H_0(B|p,y)R_0'(dy|p), \tag{18.10}$$

where

$$R_0'(C|p) = \int_{\mathbb{X}} Q_0(C|x)p(dx), \quad C \in \mathscr{B}(\mathbb{Y}),\ p \in \mathbb{P}(\mathbb{X}),$$

is a transition kernel on $\mathbb{Y}$ given $\mathbb{P}(\mathbb{X})$ and $H_0(\cdot|\cdot,\cdot)$ is a transition kernel on $\mathbb{P}(\mathbb{X})$ given $\mathbb{P}(\mathbb{X}) \times \mathbb{Y}$ that for any initial prior distribution $p_0 \in \mathbb{P}(\mathbb{X})$ and the inital observation y_0 sets the initial posteriori distribution $z_0 = H_0(p_0, y_0)$. Similarly to

(18.7), the observation y_0 is not available in the COMDP, and this equation is a stochastic equation. In addition, $H_0(p, y)$ is defined $R_0'(dy|p)$-a.s. uniquely in y for each $p \in \mathbb{P}(X)$.

Similarly to (18.8), the transition kernel

$$q_0(D|p) := \int_{\mathbb{Y}} \mathbf{1}_D[H_0(p, y)]R_0'(dy|p), \tag{18.11}$$

on $\mathbb{P}(\mathbb{X})$ given $\mathbb{P}(\mathbb{X})$ defines the *initial posterior distribution*. In particular,

$$z_0 := q_0(D|p_0), \qquad D \in \mathbb{P}(\mathbb{X}). \tag{18.12}$$

Define a sequence of *information vectors*

$$i_n := (z_0, a_0, \dots, z_{n-1}, a_{n-1}, z_n) \in I_n, \quad n = 0, 1, \dots,$$

where $z_0 \in \mathbb{P}(\mathbb{X})$ is defined in (18.12), $z_n \in \mathbb{P}(\mathbb{X})$ is recursively defined by Eq. (18.7), $I_n := \mathbb{P}(\mathbb{X}) \times (\mathbb{A} \times \mathbb{P}(\mathbb{X}))^n$ for all $n = 0, 1, \dots$, with $I_0 := \mathbb{P}(\mathbb{X})$. An *information policy* (I-policy) is a policy in a new COMDP, i.e. I-policy is a sequence $\delta = \{\delta_n : n = 0, 1, \dots\}$ such that, for each $n = 0, 1, \dots$, $\delta_n(\cdot\,|i_n)$ is a transition kernel on $\mathbb{A}$ given I_n; Hernández-Lerma [6, p. 88]. Denote by $\triangle$ the set of all I-policies. Identify the set of all Markov I-policies with a subset of Δ.

Consider Δ as a subset of Π; Hernández-Lerma [6, p. 89]. The correspondence of policies in a new COMDP (I-policies) $\delta = \{\delta_n : n = 0, 1, \dots\}$ in $\triangle$ with respective policies $\pi^\delta = \{\pi_n^\delta : n = 0, 1, \dots\}$ in Π is given; Dynkin and Yushkevich [2, pp. 251, 238] and references therein. Moreover, for all $n = 0, 1, \dots$,

$$\pi_n^\delta(\cdot\,|h_n) := \delta_n(\cdot\,|i_n(h_n)) \text{ for all } h_n \in H_n. \tag{18.13}$$

where $i_n(h_n) \in I_n$ is the information vector determined by the observable history h_n via (18.7). Thus δ and π^δ are equivalent in the sense that, for every $n = 0, 1, \dots$, π_n^δ assigns the same conditional probability on $\mathbb{A}$ as that assigned by δ_n for any observable history h_n; Dynkin and Yushkevich [2, pp. 251, 238]; Hernández-Lerma [6, p. 89]. Equality (18.13) yields that I-policy in COMDP is optimal, then the respective policy in initial POMDP is optimal too. For optimality of policy $\pi \in \Pi$ with initial distribution p necessary and sufficient the optimality of respective $\delta^\pi \in \triangle$ with respective initial distribution z^p from (18.12). If δ is stationary, then respective π is stationary too. Therefore, consider an I-policy $\delta \in \triangle$ as a policy $\pi \in \Pi$; see, for example, Dynkin and Yushkevich [2, p. 251], Sawaragi and Yoshikawa [10], Rhenius [9], Yushkevich [11]. The set of policies for the COMDP $(\mathbb{P}(\mathbb{X}), \mathbb{A}, q, q_0, \overline{c})$ is the set $\triangle$ of I-policies; Sawaragi and Yoshikawa [10], Rhenius [9], Yushkevich [11].

This reduction holds for measurable transition kernels P, Q, Q_0. The measurability of these kernels and cost function c lead to the measurability of transition probabilities for the corresponding COMDP. However, it is well known that, except the case of

finite action sets, measurability of transition probabilities is not sufficient for the existence of optimal policies in COMDPs. In spite of this certain properties hold if COMDP satisfies stronger measurability conditions. These properties are provide the validity of optimality equations

$$v_\alpha(z) = \inf_{a\in\mathbb{A}} \left\{ \bar{c}(z,a) + \alpha \int_{\mathbb{P}(\mathbb{X})} v_\alpha(s) q(ds|z,a) \right\},$$

where $z \in \mathbb{P}(\mathbb{X})$, and the property that v_α is a minimal solution of this equation. In addition if the function $\bar{c}$ is bounded on $\mathbb{P}(\mathbb{X})\times\mathbb{A}$, and $\alpha \in [0,1]$, v_α is unique bounded solution of the optimality equation and can be found by value iterations. However, if c is just bounded below on $\mathbb{X}\times\mathbb{A}$, value iterations cannot be applied; Bertsekas [1]. For COMDPs there are sufficient conditions for the existence of stationary optimal policies. If the equivalent COMDP satisfies these conditions, then the optimal policy exists, the value function can be computed by value iterations, the infimum can be substituted with minimum in the optimality equations, and the optimal policy can be derived from the optimality equations. We show below that, if POMDP satisfies these conditions then the COMDP also satisfies them.

For the COMDP, Assumption ($\mathbf{W}^*$) can be rewritten in the following form:

(i) $\bar{c}$ is K-inf-compact on $\mathbb{P}(\mathbb{X})\times\mathbb{A}$;
(ii) the transition probability $q(\cdot|z,a)$ is weakly continuous in $(z,a)\in\mathbb{P}(\mathbb{X})\times\mathbb{A}$.

Theorem 18.1 has the following form for the COMDP $(\mathbb{P}(\mathbb{X}), \mathbb{A}, q, \bar{c})$:

Theorem 18.2 (cf. Feinberg, Kasyanov, and Zadoianchuk [4, Theorem 2]). *Let COMDP $(\mathbb{P}(\mathbb{X}), \mathbb{A}, q, \bar{c})$ satisfy Assumption ($\mathbf{W}^*$) and, in addition, either Assumption ($\mathbf{D}$) or Assumption ($\mathbf{P}$) holds. Then:*

(*i*) *the functions $v_{n,\alpha}$, $n = 0, 1, 2, \ldots$, and v_α are lower semi-continuous on $\mathbb{P}(\mathbb{X})$, and $v_{n,\alpha}(z) \to v_\alpha(z)$ as $n\to\infty$ for all $z\in\mathbb{P}(\mathbb{X})$;*
(*ii*) *for any $z\in\mathbb{P}(\mathbb{X})$, and $n = 0, 1, \ldots,$*

$$\begin{aligned} v_{n+1,\alpha}(z) &= \min_{a\in\mathbb{A}} \{\bar{c}(z,a) + \alpha \int_{\mathbb{P}(\mathbb{X})} v_{n,\alpha}(z') q(dz'|z,a)\} \\ &= \min_{a\in\mathbb{A}} \Big\{ \int_{\mathbb{X}} c(x,a) z(dx) + \int_{\mathbb{X}}\int_{\mathbb{X}}\int_{\mathbb{Y}} v_{n,\alpha}(H(z,a,y)) \\ &\quad \times \alpha Q(dy|a,x') P(dx'|x,a) z(dx) \Big\}, \end{aligned} \tag{18.14}$$

where $v_{0,\alpha}(z) = 0$ for all $z\in\mathbb{P}(\mathbb{X})$, and the nonempty sets

$$\begin{aligned} A_{n,\alpha}(z) := \Big\{ a\in\mathbb{A} :\ & v_{n+1,\alpha}(z) \\ & = c(z,a) + \alpha \int_{\mathbb{P}(\mathbb{X})} v_{n,\alpha}(z') q(dz'|z,a) \Big\}, \end{aligned}$$

where $z \in \mathbb{P}(\mathbb{X})$ *, satisfy the following properties: (a) the graph* $\mathrm{Gr}(A_{n,\alpha}) = \{(z,a) : z \in \mathbb{P}(\mathbb{X}), a \in A_{n,\alpha}(z)\}$, $n = 0, 1, \ldots$, *is a Borel subset of* $\mathbb{P}(\mathbb{X}) \times \mathbb{A}$, *and (b) if* $v_{n+1,\alpha}(z) = \infty$, *then* $A_{n,\alpha}(z) = \mathbb{A}$ *and, if* $v_{n+1,\alpha}(z) < \infty$, *then* $A_{n,\alpha}(z)$ *is compact;*

(*iii*) *for any* $N = 1, 2, \ldots$, *there exists a Markov optimal N-horizon I-policy* $(\phi_0, \ldots, \phi_{N-1})$ *and if, for an N-horizon Markov I-policy* $(\phi_0, \ldots, \phi_{N-1})$ *the inclusions* $\phi_{N-1-n}(z) \in A_{n,\alpha}(z)$, $z \in \mathbb{P}(\mathbb{X})$, $n = 0, \ldots, N-1$, *hold then this I-policy is N-horizon optimal;*

(*iv*) *for* $\alpha \in [0, 1]$

$$
\begin{aligned}
v_\alpha(z) &= \min_{a \in \mathbb{A}} \left\{ \overline{c}(z,a) + \alpha \int_{\mathbb{P}(\mathbb{X})} v_\alpha(z') q(dz'|z,a) \right\} \\
&= \min_{a \in \mathbb{A}} \Bigg\{ \int_{\mathbb{X}} c(x,a) z(dx) + \alpha \int_{\mathbb{X}} \int_{\mathbb{X}} \int_{\mathbb{Y}} v_\alpha(H(z,a,y)) \\
&\qquad \times Q(dy|a,x') P(dx'|x,a) z(dx) \Bigg\}, \quad z \in \mathbb{P}(\mathbb{X}),
\end{aligned}
$$

and the nonempty sets

$$
\begin{aligned}
A_\alpha(z) := \{ a \in \mathbb{A} :\ & v_\alpha(z) = \overline{c}(z,a) \\
& + \alpha \int_{\mathbb{P}(\mathbb{X})} v_\alpha(z') q(dz'|z,a) \}, \quad z \in \mathbb{P}(\mathbb{X}),
\end{aligned}
$$

satisfy the following properties: (a) the graph $\mathrm{Gr}(A_\alpha) = \{(z,a) : z \in \mathbb{P}(\mathbb{X}), a \in \mathbb{A}_\alpha(z)\}$ *is a Borel subset of* $\mathbb{P}(\mathbb{X}) \times \mathbb{A}$, *and (b) if* $v_\alpha(z) = \infty$, *then* $A_\alpha(z) = \mathbb{A}$ *and, if* $v_\alpha(z) < \infty$, *then* $A_\alpha(z)$ *is compact.*

(*v*) *for an infinite horizon there exists a stationary discount-optimal I-policy* ϕ_α, *and a stationary I-policy is optimal if and only if* $\phi_\alpha(z) \in A_\alpha(z)$ *for all* $z \in \mathbb{P}(\mathbb{X})$.

(*vi*) *if the function c is inf-compact, the functions* $v_{n,\alpha}$, $n = 1, 2, \ldots$, *and* v_α *are inf-compact on* $\mathbb{P}(\mathbb{X})$.

Note that statement (vi) of Theorem 18.2 follows from Feinberg and Lewis [5, Proposition 3.1(iv)].

Hernández-Lerma [6, Sect. 4.4] provided the following conditions for the existence of optimal policies for the COMDP: (a) $\mathbb{A}$ is compact, (b) the cost function c is bounded and continuous, (c) the transition probability $P(\cdot|x,a)$ and the observation kernel $Q(\cdot|a,x)$ are weakly continuous transition kernels; (d) there exists a weakly continuous $H : \mathbb{P}(\mathbb{X}) \times \mathbb{A} \times \mathbb{Y} \to \mathbb{P}(\mathbb{X})$ satisfying (18.6). Consider the following relaxed version of Assumption (d).

Assumption (H) There exists a transition kernel H on $\mathbb{X}$ given $\mathbb{P}(\mathbb{X}) \times \mathbb{A} \times \mathbb{Y}$ satisfying (18.6) such that: if a sequence $\{z_n\} \subseteq \mathbb{P}(\mathbb{X})$ converges weakly to $z \in \mathbb{P}(\mathbb{X})$, and $\{a_n\} \subseteq \mathbb{A}$ converges to $a \in \mathbb{A}$, $n \to \infty$, then there exists a subsequence $\{(z_{n_k}, a_{n_k})\}_{k \geq 1} \subseteq \{(z_n, a_n)\}_{n \geq 1}$ such that

$$H(z_{n_k}, a_{n_k}, y) \text{ converges weakly to } H(z, a, y), \ n \to \infty,$$

and this convergence takes place $R'(\cdot|z, a)$ almost surely for all $y \in \mathbb{Y}$.

The following theorem relaxes assumptions (a), (b), and (d) in Hernández-Lerma [6, Sect. 4.4].

Theorem 18.3 *Under the following four conditions:*

(*a*) *either Assumption* (**D**) *or Assumption* (**P**) *holds;*
(*b*) *Assumption* (**W***) *holds for the MDP* $(\mathbb{X}, \mathbb{A}, P, c)$*;*
(*c*) *either the stochastic kernel* $R'(dy|z, a)$ *on* $\mathbb{Y}$ *given* $\mathbb{P}(\mathbb{X}) \times \mathbb{A}$ *is setwise continuous and Assumption* (**H**) *holds, or the stochastic kernel* $Q(dy|a, x)$ *on* $\mathbb{Y}$ *given* $\mathbb{A} \times \mathbb{X}$ *is weakly continuous and there exists a weakly continuous* $H : \mathbb{P}(\mathbb{X}) \times \mathbb{A} \times \mathbb{Y} \to \mathbb{P}(\mathbb{X})$ *satisfying* (18.6)*;*

the COMDP $(\mathbb{P}(\mathbb{X}), \mathbb{A}, q, \overline{c})$ *satisfies Assumption* (**W***) *and therefore statements* (*i*)–(*vi*) *of Theorem 18.2 hold.*

If transition kernel $Q(dy|a, x)$ on $\mathbb{Y}$ given $\mathbb{A} \times \mathbb{X}$ is continuous in the total variation, then Assumption **(H)** holds, and this leads to the following theorem.

Theorem 18.4 *Let the transition kernel* $P(dx'|x, a)$ *on* $\mathbb{X}$ *given* $\mathbb{X} \times \mathbb{A}$ *be weakly continuous and let the transition kernel* $Q(dy|a, x)$ *on* $\mathbb{Y}$ *given* $\mathbb{A} \times \mathbb{X}$ *be continuous in the total variation. Then:* (*i*) *the transition kernel* $R'(dy|z, a)$ *on* $\mathbb{Y}$ *given* $\mathbb{P}(\mathbb{X}) \times \mathbb{A}$ *is setwise continuous, Assumption* (**H**) *holds, and* (*iii*) *the transition kernel* q *on* $\mathbb{P}(\mathbb{X})$ *given* $\mathbb{P}(\mathbb{X}) \times \mathbb{A}$ *is setwise continuous.*

Theorems 18.3 and 18.4 imply the following result.

Theorem 18.5 *Let assumptions of* (*a*) *and* (*b*) *from Theorem 18.3 hold and let the transition kernel* $Q(dy|a, x)$ *on* $\mathbb{Y}$ *given* $\mathbb{A} \times \mathbb{X}$ *be continuous in the total variation. Then statements (i)–(vi) of Theorem 18.2 hold.*

18.4 Example

Let $\mathbb{X}$, $\mathbb{A}$ and $\mathbb{Y}$ are nonempty Borel subsets of $\mathbb{R}$, $\{\xi_n\}_{n\geq 1}$ is a sequence of independent and identically distributed random vectors with values in some Borel subset $\mathbb{S}$ of a Polish space. Assume that the generic disturbance ξ has a distribution μ on $\mathbb{S}$. Let also $\{\eta_n\}_{n\geq 1}$ is a sequence of independent and identically distributed random variables, that uniformly distributed on [0, 1]. The goal is to minimize the expected discounted total costs over the infinite time horizon.

Consider a stochastic partially observable control system of the form

$$x_{n+1} = F(x_n, a_n, \xi_n), \ n = 0, 1, \ldots, \tag{18.15}$$

$$y_{n+1} = G(a_n, x_{n+1}, \eta_n), \ \ n = 0, 1, \ldots, \tag{18.16}$$

where F and G are given measurable function from $\mathbb{X} \times \mathbb{A} \times \mathbb{S}$ to $\mathbb{X}$ and from $\mathbb{A} \times \mathbb{X} \times [0, 1]$ to $\mathbb{Y}$ respectively. The states x_n are not observable, while the states y_n are observable.

The transition law of the system can be written as

$$P(B|x, a) = \int_{\mathbb{S}} \mathbf{1}\{F(x, a, s) \in B\}\mu(ds).$$

The observation kernel is given by

$$Q(C|a, x) = \int_{[0,1]} \mathbf{1}\{G(a, x, s) \in C\}\lambda(ds),$$

where $\lambda \in \mathbb{P}([0, 1])$ is a Lebesgue measure on $[0, 1]$.

It is clear that, if $(x, a) \to F(x, a, s)$ is continuous mapping on $\mathbb{X} \times \mathbb{A}$ for every $s \in \mathbb{S}$, then stochastic kernel $P(dx'|x, a)$ on $\mathbb{X}$ given $\mathbb{X} \times \mathbb{A}$ is weakly continuous.

Assume that G is a continuous mapping on $\mathbb{A} \times \mathbb{X} \times [0, 1]$, its derivative by the last variable exists (we denote it by g) is a continuous mapping on $\mathbb{A} \times \mathbb{X} \times [0, 1]$ and it has a fixed sign, i.e. for some constant $\beta > 0$ we have $|g(a, x, s)| \geq \beta$ for any $a \in \mathbb{A}, x \in \mathbb{X}, s \in G(a, x, [0, 1])$, where $G(a, x, [0, 1]) = \{G(a, x, s') : s' \in [0, 1]\}$. Then it is possible to show that that the observation transition kernel Q on $\mathbb{Y}$ given $\mathbb{A} \times \mathbb{X}$ is continuous in the total variation.

Finally, we assume that one-period cost $c : \mathbb{X} \times \mathbb{A} \to \overline{\mathbb{R}}$ is K-inf-compact function (see for details Feinberg, Kasyanov, and Zadoianchuk [3]), it is bounded from below. Then the MDP satisfies Assumption ($\mathbf{W}^*$), that is, K-inf-compactness of the cost function c and weak continuity of the transition kernel P that describes transition probabilities for the MDP. In addition, the observation transition kernel Q is continuous in the total variation. Therefore, the corresponding COMDP satisfies Assumption ($\mathbf{W}^*$). Thus, in view of Theorems 18.3–18.5 for the COMDP there exist a stationary optimal, they satisfy optimality equations, and the value function can be computed via value iterations. By using the standard known procedures [6, Chap. 4], an optimal policy for the COMDP can be used to construct an optimal policy for the initial problem, which is typically nonstationary.

18.5 Conclusions

This presentation studies POMDPs with Borel state, action, and observation spaces satisfying mild continuity assumptions that guarantee the following properties for the underlying fully observable MDP: (i) the existence of stationary optimal policies, (ii) validity of optimality equations, and (iii) convergence of value iterations for the expected total discounted costs as well as for the expected total costs, when the one-

step cost function is nonnegative. This presentation provides additional sufficient conditions under which the COMDP possesses the same continuity assumptions as the underlying MDP and, therefore, properties (i)–(iii) are also satisfied for the COMDP. One of such sufficient conditions is the continuity of the observation transition kernel in the total probability; see Theorem 18.5. Therefore, this paper provides theoretical foundations to analyze POMDPs with general state and action spaces and with expected total cost criteria.

Acknowledgments The authors thank Dr. Huizhen Janey Yu and Dr. N.V. Zadoianchuk for their useful remarks. Research of the first coauthor was partially supported by NSF grants CMMI-0928490 and CMMI-1335296.

References

1. Bertsekas, D.P., Shreve, S.E.: Stochastic Optimal Control: The Discrete-Time Case. Academic Press, New York (1978) (reprinted by Athena Scientific, Belmont, 1996)
2. Dynkin, E.B., Yushkevich, A.A.: Controlled Markov Processes. Springer, New York (1979)
3. Feinberg, E.A., Kasyanov, P.O., Zadoianchuk, N.V.: Berge's theorem for noncompact image sets. J. Math. Anal. Appl. **397**(1), 255–259 (2013)
4. Feinberg, E.A., Kasyanov, P.O., Zadoianchuk, N.V.: Average-cost Markov decision processes with weakly continuous transition probabilities. Math. Oper. Res. **37**(4), 591–607 (2012)
5. Feinberg, E.A., Lewis, M.E.: Optimality inequalities for average cost Markov decision processes and the stochastic cash balance problem. Math. Oper. Res. **32**(4), 769–783 (2007)
6. Hernández-Lerma, O.: Adaptive Markov Control Processes. Springer, New York (1989)
7. Hernández-Lerma, O., Lassere, J.B.: Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer, New York (1996)
8. Parthasarathy, K.R.: Probability Measures on Metric Spaces. Academic Press, New York (1967)
9. Rhenius, D.: Incomplete information in Markovian decision models. Ann. Statist. **2**, 1327–1334 (1974)
10. Sawaragi, Y., Yoshikawa, T.: Descrete-time markovian decision processes with incomplete state observations. Ann. Math. Statist. **41**, 78–86 (1970)
11. Yushkevich, A.A.: Reduction of a controlled Markov model with incomplete data to a problem with complete information in the case of Borel state and control spaces. Theor. Probab. Appl. **21**, 153–158 (1976)