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. In particular, we consider POMDPs with weakly continuous transition probabilities and bounded below $K$-infcompact cost functions. For a fully observable MDPs these two conditions guarantee the following three properties: (i) validity of finite-horizon and infinite-horizon optimality equations, (ii) convergence of value iterations to infinite-horizon value functions, (iii) existence of stationary optimal policies. We show that the single additional assumption, that the observation transition probability is continuous in the total variation, implies properties (i)-(iii) for the COMDP. Therefore, this condition also implies the existence of optimal policies for POMDPs. We also provide a more general and less constructive sufficient condition for the validity of (i)-(iii) for the COMDP and therefore for the existence of optimal policies for a POMDP and the possibility of finding them by transforming optimal policies for the corresponding COMDP.

I. 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 [10], [9], [2], [6]. After an optimal policy for the COMDP is found, it can be used to compute an optimal 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, Chapter 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 POMDPs. 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. In the case of general observation spaces, there are three continuity conditions on the observation probabilities that are equivalent to this condition, when the observation space is countable. These conditions are weak continuity, setwise continuity, and continuity in the total variation of observation probabilities (also called kernels or stochastic kernels).

In this paper, for a POMDP with a Borel state space, we study minimization of either expected total nonnegative costs or expected total discounted costs, when the one-step cost function is bounded below. The goal of this study is to obtain sufficient conditions for the existence of optimal policies and to provide methods for their computation. It is well-known that an optimal policy for the POMDP can be found from an optimal policy for the COMDP. Here we study the existence of optimal policies for COMDPs and their characterization and computation via optimality equations for problems with possibly non-compact action sets, unbounded cost functions (they are assumed bounded below), and uncountable observation sets. The one-step cost functions are assumed to be $K$-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 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 sufficient condition for the existence optimal policies for weakly continuous observation probabilities, which is different from
the weak continuity of the filtration kernel considered in Hernández-Lerma [6, p. 90, Assumption 4.1(d)].

II. MODEL DESCRIPTION

For a metric space \( S \), let \( \mathcal{B}(S) \) be its Borel \( \sigma \)-field, that is, the \( \sigma \)-field generated by all open sets of the metric space \( S \). For a Borel subset \( E \subset S \), we denote by \( \mathcal{B}(E) \) the \( \sigma \)-field whose elements are intersections of \( E \) with elements of \( \mathcal{B}(S) \). Observe that \( E \) is a metric space with the same metric as on \( S \), and \( \mathcal{B}(E) \) is its Borel-\( \sigma \)-field. The space \( E \) is a Borel space, if \( E \) is a Borel subset of a Polish (complete separable metric) space \( S \). On \( E \) consider the induced measurable topology. For a metric space \( S \), we denote by \( \mathbb{P}(S) \) the set of probability measures on \((S,\mathcal{B}(S))\). A sequence of probability measures \( \{\mu_n\} \) from \( \mathbb{P}(S) \) converges weakly (setwise) to \( \mu \in \mathbb{P}(S) \) if for any bounded continuous (bounded Borel-measurable) function \( f \) on \( S \)

\[
\int_S f(s)\mu_n(ds) \to \int_S f(s)\mu(ds) \quad \text{as} \quad n \to \infty.
\]

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

\[
\lim_{n \to \infty} \sup_{f \in F_1(S)} \left\{ \int_S f(s)\mu_n(ds) - \int_S f(s)\mu(ds) \right\} = 0,
\]

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

Note that \( \mathbb{P}(S) \) is a Polish space with respect to the weak convergence topology for probability measures; Parthasarathy [8, Chapter II]. For Borel spaces \( S_1 \) and \( S_2 \), a (Borel-measurable) transition kernel \( R(ds_1|s_2) \) on \( S_1 \) given \( S_2 \) is a mapping \( R(\cdot|\cdot) : \mathcal{B}(S_1) \times S_2 \to [0,1] \), such that \( R(\cdot|s_2) \) is a probability measure on \( S_1 \) for any \( s_2 \in S_2 \), and \( R(B|\cdot) \) is a Borel-measurable function on \( S_2 \) for any Borel set \( B \in \mathcal{B}(S_1) \). A transition kernel \( R(ds_1|s_2) \) on \( S_1 \) given \( S_2 \) defines a Borel measurable mapping \( s_2 \to R(\cdot|x) \) from the metric space \( \mathbb{P}(S_1) \) endowed with the topology of weak convergence. A transition kernel \( R(ds_1|s_2) \) on \( S_1 \) given \( S_2 \) is called weakly continuous (setwise continuous, continuous in the total variation), if \( R(\cdot|x) \) converges weakly (setwise, in the total variation) to \( R(\cdot|x) \) whenever \( x_n \) converges to \( x \) in \( S_2 \).

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

Partially observable Markov decision process (POMDP) is specified by \( (X,Y,A,P,Q,c) \), where \( X \) is the state space, \( Y \) is the observation set, \( A \) is the action set, \( P(dx'|x,a) \) is the state transition law, \( Q(dy|a,x) \) is the observation kernel, \( c : X \times A \to \mathbb{R} = \mathbb{R} \cup \{+\infty\} \) 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 \( p \);
- 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 \( x_n \in X \) and the decision-maker chooses an action \( a_n \in A \), 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 \( y_{n+1} \in Y \) 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: \( h_0 := (p,y_0) \in H_0 \), \( h_n := (p,y_0,a_0,y_1-a_0,y_1-a_1,y_n) \in H_n \) for all \( n = 1, 2, \ldots \), where \( H_0 := P(X) \times Y \) and \( H_n := H_{n-1} \times A \times Y \) if \( n = 1, 2, \ldots \). Then a policy for the POMDP is defined as a sequence \( \pi = \{\pi_n\} \) such that, for each \( n = 0, 1, \ldots \), \( \pi_n \) is a transition kernel on \( A \) given \( H_n \). Moreover, \( \pi \) 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 \( \Pi \). The Ionescu Tulcea theorem (Bertsekas and Shreve [1, pp. 140-141] or Hernández-Lerma and Lasserre [7, p.178]) implies that a policy \( \pi \in \Pi \) and an initial distribution \( p \in P(X) \), together with the transition kernels \( P, Q \) and \( Q_0 \) determine a unique probability measure \( P^\pi \) on the set of all trajectories \( \mathbb{H} = P(X) \times (Y \times \mathbb{A}) \) endowed with the product of-\( \sigma \)-field defined by Borel \( \sigma \)-field of \( P(X), Y, \) and \( A \) respectively. The expectation with respect to this probability measure is denoted by \( E^\pi_p \).

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

\[
v_{N,\alpha}^\pi(p) := E_{N,\alpha}^p \sum_{n=0}^{N-1} \alpha^n c(x_n,a_n), \quad p \in P(X), \quad (1)
\]

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

**Assumption (D).** \( c \) is bounded below on \( X \times A \) and \( \alpha \in [0,1) \).

**Assumption (P).** \( c \) is nonnegative on \( X \times A \) and \( \alpha \in [0,1] \).

In both cases (1) defines an infinite horizon expected total discounted cost, and we denote it by \( v_{N,\alpha}^\pi(p) \). 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 \( X \times A \), a discounted model can be converted into a positive model by shifting the cost function. In particular, let \( c(x,a) \geq -K \) for all \((x,a) \in X \times A\). Consider a new cost function \( \tilde{c}(x,a) := c(x,a) + K \) for all \((x,a) \in X \times A\). Then the corresponding total discounted reward is equal to

\[
\tilde{v}_{\alpha}^\pi(p) := v_{\infty,\alpha}^\pi(p) + \frac{K}{1-\alpha}, \quad \pi \in \Pi, \quad p \in P(X).
\]
Thus, optimizing $v_{\pi}^\alpha$ 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 larger classes of applications. Thus we formulate the results for either of these models.

For any function $g^\pi(p)$, including $g^\pi(p) = v_{\pi}^\alpha(p)$ and $g^\pi(p) = \hat{v}_{\pi}^\alpha(p)$ define the optimal cost

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

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

We recall that a function $c$ defined on $\mathcal{X} \times \mathcal{A}$ is inf-compact (or lower semi-compact) if the set $\{(x, a) \in \mathcal{X} \times \mathcal{A} : c(x, a) \leq \lambda\}$ is compact for any finite number $\lambda$. A function $c$ defined on $\mathcal{X} \times \mathcal{A}$ is called $K$-inf-compact on $\mathcal{X} \times \mathcal{A}$, if for any compact subset $K$ of $\mathcal{X}$, the function $c$ is inf-compact on $K \times \mathcal{A}$; Feinberg, Kasyanov, and Zadoianchuk [3, Definition 11]. $K$-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 $\mathcal{X}$ and $\mathcal{A}$ if and only if it satisfies the following two conditions:

(a) $c$ is lower semi-continuous;

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

As a POMDP $(\mathcal{X}, \mathcal{Y}, \mathcal{A}, P, Q, c)$, consider the classical MDP $(\mathcal{X}, \mathcal{A}, P, c)$, whenever all the states are observable. An MDP can be viewed as a particular POMDPs with $\mathcal{Y} = \mathcal{X}$ and $Q(B|a, x) = Q(B|x) = 1\{x \in B\}$ for all $x \in \mathcal{X}$, $a \in \mathcal{A}$, and $B \subseteq \mathcal{X}$. In fact, this POMDP 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 are equal to $\mathcal{A}$ for any state.

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

(i) $c$ is K-inf-compact on $\mathcal{X} \times \mathcal{A}$;

(ii) the transition probability $P(\cdot | x, a)$ is weakly continuous in $(x, a) \in \mathcal{X} \times \mathcal{A}$.

**Theorem 2.1**: (cf. Feinberg, Kasyanov, and Zadoianchuk [4, Theorem 2]). Let MDP $(\mathcal{X}, \mathcal{A}, P, c)$ satisfies Assumption ($W^*$). Consider either positive or discounted model. Then:

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

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

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

(2)

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

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

satisfy the following properties: (a) the graph $Gr(A_{n, \alpha}) := \{(x, a) : x \in \mathcal{X}, a \in A_{n, \alpha}(x)\}$ of $A_{n, \alpha}$, $n = 0, 1, \ldots$, is a Borel subset of $\mathcal{X} \times \mathcal{A}$, 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 $\phi_{N-1-n}(x) \in A_{n, \alpha}(x)$, $x \in \mathcal{X}$, $n = 0, \ldots, N - 1$, hold, then this policy is $N$-horizon optimal;

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

$$v_{\alpha}(x) = \min_{a \in A(x)} \left\{ c(x, a) + \alpha \int_{\mathcal{X}} v_{\alpha}(y) P(dy|x, a) \right\}, \quad x \in \mathcal{X},$$

(3)

and the nonempty sets

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

satisfy the following properties: (a) the graph of $A_\alpha$, $Gr(A_\alpha) := \{(x, a) : x \in \mathcal{X}, a \in A_\alpha(x)\}$ is a Borel subset of $\mathcal{X} \times \mathcal{A}$, 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 $\phi_\alpha$, and a stationary policy is optimal if and only if $\phi_\alpha(x) \in A_\alpha(x)$ for all $x \in \mathcal{X}$;

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

**III. REDUCTION OF POMDPs TO COMDPs**

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 4.1 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 4.2, 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, 12). To simplify notations, we sometimes drop the time parameter. Given a posterior distribution \( z \) of the state \( z \) at time epoch \( n = 0, 1 \), 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 \( B \in \mathbb{B}(\mathbb{X}) \) and the observation at time \( n \) belongs to the set \( C \in \mathbb{B}(\mathbb{Y}) \),

\[
R(B \times C|z, a) := \int_X \int_X Q(C|a, x') P(dx'|x, a) z(dx),
\]

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 Hernandez-Lerma [6]; or Yushkevich [12] for details. Therefore, the probability \( R'(C|z, a) \) that the observation \( y \) at time \( n \) belongs to the set \( C \in \mathbb{B}(\mathbb{Y}) \) is

\[
R'(C|z, a) = \int_X R(B \times C|z, a) \int_X Q(C|a, x') P(dx'|x, a) z(dx),
\]

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

\[
R(B \times C|z, a) = \int_C H(B|z, a, y) R'(dy|z, a),
\]

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 \( H(z, a, y)\{ \cdot \} \) is defined \( R' \)-a.s. uniquely in \( y \); Dynkin and Yushkevich [2, p. 309]. It is known that for a posterior distribution \( z_n, a_n \in \mathbb{P}(\mathbb{X}) \), action \( a_n \in A(x) \), and an observation \( y_{n+1} \in \mathbb{Y} \), the posterior distribution \( z_{n+1} \in \mathbb{P}(\mathbb{X}) \)

\[
z_{n+1} = H(z_n, a_n, y_{n+1}).
\]

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 \( z_{n+1} \) is a stochastic equation that maps \((z_n, a_n) \in \mathbb{P}(\mathbb{X}) \times \mathbb{A} \) to \( \mathbb{P}(\mathbb{X}) \). 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

\[
q(D|z, a) := \int_X \mathbf{1}_D(H(z, a, y)) R'(dy|z, a),
\]

where for \( D \in \mathbb{B}(\mathbb{P}(\mathbb{X})) \)

\[
\mathbf{1}_D[u] = \begin{cases} 1, & u \in D, \\ 0, & u \notin D; \end{cases}
\]

Hernández-Lerma [7, p. 87]. The measurable particular choice of stochastic kernel \( H \) from (6) does not effect on the definition of \( q \) from (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 \( R' \)-a.s. uniquely in \( y \); Dynkin and Yushkevich [2, p. 309].

The COMDP is defined as an MDP with parameters \((\mathbb{P}(\mathbb{X}), \mathbb{A}, q, \bar{c})\), 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 \( \bar{c}: \mathbb{P}(\mathbb{X}) \times \mathbb{A} \to \mathbb{R} \), defined as

\[
\bar{c}(z, a) := \int_X c(x, a) z(dx), \quad z \in \mathbb{P}(\mathbb{X}), \quad a \in \mathbb{A};
\]

(iv) transition probabilities \( q \) on \( \mathbb{P}(\mathbb{X}) \) given \( \mathbb{P}(\mathbb{X}) \times \mathbb{A} \) as defined in (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 POMDP. First, we recall how the initial state distribution \( z_0 \in \mathbb{P}(\mathbb{X}) \) can be computed for the COMDP. Similar to the transition kernels \( R, R' \), and \( H \), consider a transition kernel

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

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

\[
R_0(B \times C|p) = \int_C H_0(B|p, y) R_0'(dy|p),
\]

where

\[
R_0'(C|p) = \int_X Q_0(C|x) p(dx), \quad C \in \mathbb{B}(\mathbb{Y}), \quad p \in \mathbb{P}(\mathbb{X}),
\]

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 \( \mathbb{P}(\mathbb{X}) \times \mathbb{Y} \). Similar to (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 \( R_0' \)-a.s. uniquely in \( y \) for each \( p \in \mathbb{P}(\mathbb{X}) \). Similar to (8), the transition kernel

\[
q_0(D|p) := \int_Y \mathbf{1}_D(H_0(p, y)) R_0'(dy|p),
\]

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

\[
z_0 := q_0(D|p), \quad D \in \mathbb{P}(\mathbb{X}).
\]

Define a sequence of information vectors

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

where \( z_0 \in \mathbb{P}(\mathbb{X}) \) is defined in (12), \( z_n \in \mathbb{P}(\mathbb{X}) \) is recursively defined by equation (7), \( I_n := \mathbb{P}(\mathbb{X}) \times (\mathbb{A} \times \mathbb{P}(\mathbb{X}))^n \) for all \( n = 0, 1, \ldots \), with \( I_0 := \mathbb{P}(\mathbb{X}) \). An information policy (I-policy) is a policy in the COMDP, that is, I-policy is a sequence \( \delta = \{\delta_n : n = 0, 1, \ldots \} \) such that, for each \( n = 0, 1, \ldots, \delta_n(\cdot|i_n) \) is a transition kernel on \( \mathbb{A} \) given
In; Hernández-Lerma [6, p. 88]. Denote by $\Delta$ the set of all $I$-policies.
Consider $\Delta$ as a subset of $\Pi$; Dynkin and Yushkevich [2, pp. 251, 238] and references therein. Moreover, for all $n = 0, 1, \ldots$

$$\pi_n^{\delta}(\cdot | h_n) := \delta_n(\cdot | i_n(h_n)) \text{ for all } h_n \in \mathbb{H}_n. \quad (13)$$

where $i_n(h_n) \in I_n$ is the information vector determined by the observable history $h_n$ via (7). Thus $\delta$ and $\pi^{\delta}$ are equivalent in the sense that, for every $n = 0, 1, \ldots$, $\pi_n^{\delta}$ assigns the same conditional probability on $\mathbb{H}$ as that assigned by $\delta_n$ for any observable history $h_n$; Dynkin and Yushkevich [2, pp. 251, 238]; Hernández-Lerma [6, p. 89]. Equality (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 $\rho$ necessary and sufficient the optimality of respective $\delta^\pi \in \Delta$ with respective initial distribution $\gamma^\pi$ from (12). If $\delta$ is stationary, then respective $\pi$ is stationary too. Therefore, consider an I-policy $\delta \in \Delta$ as a policy $\pi \in \Pi$; see, for example, Dynkin and Yushkevich [2, p. 251], Sawaragi and Yoshikawa [10], Rhenius [9], Yushkevich [12]. The set of policies for the COMDP $(P(\mathbb{X}), A, q, \bar{c}, \bar{c})$ is the set $\Delta$ of I-policies; Sawaragi and Yoshikawa [10], Rhenius [9], Yushkevich [12].

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

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

where $z \in \mathbb{P}(\mathbb{X})$, and the property that $v_{\alpha}$ is a minimal solution of this equation. In addition if the function $\bar{c}$ is bounded on $\mathbb{P}(\mathbb{X}) \times A$, and $\alpha \in [0, 1)$, $v_{\alpha}$ is the unique bounded solution of the optimality equation and can be found by value iterations. However, if $c$ is just bounded below on $\mathbb{X} \times A$, 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.

IV. Optimality Results

For the COMDP, Assumption (W') can be rewritten in the following form:

(i) $\bar{c}$ is $K$-inf-compact on $\mathbb{P}(\mathbb{X}) \times A$;

(ii) the transition probability $q(\cdot | z, a)$ is weakly continuous in $(z, a) \in \mathbb{P}(\mathbb{X}) \times A$.

Theorem 2.1 has the following form for the COMDP $(\mathbb{P}(\mathbb{X}), A, q, \bar{c})$:

Theorem 4.1: (cf. Feinberg, Kasyanov, and Zadoianchuk [4, Theorem 2]). Let COMDP $(\mathbb{P}(\mathbb{X}), A, q, \bar{c})$ satisfy Assumption (W') and, in addition, either Assumption (D) or Assumption (P) holds. Then:

(i) the functions $v_{n,\alpha}$, $n = 0, 1, 2, \ldots$, and $v_{\alpha}$ 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 $n = 0, 1, \ldots$

$$v_{n+1,\alpha}(z) = \min_{a \in A} \left\{ \bar{c}(z, a) + \alpha \int_{\mathbb{P}(\mathbb{X})} v_{\alpha}(z')q(dz'|z, a) \right\} = \alpha Q(dy|a, x)P(dx'|x, a)q(dx),$$

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

$$A_{\alpha,\alpha}(z) := \{a \in A : v_{n+1,\alpha}(z) \leq v_{n,\alpha}(z)\},$$

where $z \in \mathbb{P}(\mathbb{X})$, satisfy the following properties: (a) the graph $Gr(A_{\alpha,\alpha}) = \{(z, a) : z \in \mathbb{P}(\mathbb{X}), a \in A_{\alpha,\alpha}(z)\}$, $n = 0, 1, \ldots$, is a Borel subset of $\mathbb{P}(\mathbb{X}) \times A$, and (b) if $v_{n+1,\alpha}(z) = \infty$, then $A_{\alpha,\alpha}(z) = A$ and, if $v_{n+1,\alpha}(z) < \infty$, then $A_{\alpha,\alpha}(z)$ 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 $\phi_{N-1,\alpha}(z) \in A_{\alpha,\alpha}(z)$, $z \in \mathbb{P}(\mathbb{X})$, $n = 0, \ldots, N - 1$, hold, then this I-policy is $N$-horizon optimal;

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

$$v_{\alpha}(z) = \min_{a \in A} \left\{ \bar{c}(z, a) + \alpha \int_{\mathbb{P}(\mathbb{X})} v_{\alpha}(z')q(dz'|z, a) \right\} = \min_{\phi \in \phi_{\alpha}} \int_{\mathbb{P}(\mathbb{X})} \int_{\mathbb{X}} \int_{\mathbb{Y}} v_{\alpha}(H(z, a, y)) \times Q(dy|a, x)P(dx'|x, a)q(dx),$$

and the nonempty sets

$$A_\alpha(z) := \{a \in A : v_{\alpha}(z) = \bar{c}(z, a) + \alpha \int_{\mathbb{P}(\mathbb{X})} v_{\alpha}(z')q(dz'|z, a)\},$$

satisfy the following properties: (a) the graph $Gr(A_\alpha) = \{(z, a) : z \in \mathbb{P}(\mathbb{X}), a \in A_\alpha(z)\}$ is a Borel subset of $\mathbb{P}(\mathbb{X}) \times A$, and (b) if $v_{\alpha}(z) = \infty$, then $A_\alpha(z) = A$ 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 $\phi_\alpha$, and a stationary I-policy is optimal if and only if $\phi_\alpha(z) \in A_\alpha(z)$ for all $z \in P(X)$.

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

Note that statement (vi) of Theorem 4.1 follows from compactness of $P(X)$; Parthasarathy [8]), properties of $K$-inf-compact functions; Feinberg, Kasyanov, and Zadoianchuk [3], and Feinberg and Lewis [5, Proposition 3.1(iv)].

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

**Assumption (H).** There exists a transition kernel $H$ on $X$ given $P(X) \times A \times Y$ satisfying (6) such that if a sequence $\{z_n\} \subseteq P(X)$ converges weakly to $z \in P(X)$, and $\{a_n\} \subseteq A$ converges to $a \in A$, $n \to \infty$, then there exists a subsequence $\{(z_{n_k}, a_{n_k})\}_{k \geq 1} \subseteq \{(z_n, a_n)\}_{n \geq 1}$ such that

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

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

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

**Theorem 4.2:** Under the following four conditions:

(a) either Assumption (D) or Assumption (P) holds;

(b) Assumption (W*) holds for the MDP $(X, A, P, c)$;

(c) either the stochastic kernel $R'(dy|z, a)$ is setwise continuous and Assumption (H) holds, or the stochastic kernel $Q(dy|a, x)$ on $Y$ given $A \times X$ is weakly continuous and there exists a weakly continuous $H: P(X) \times A \times Y \to P(X)$ satisfying (6);

the COMDP $(P(X), A, q, c)$ satisfies Assumption (W*) and therefore statements (i)–(vi) of Theorem 4.1 hold.

Theorem 4.2 follows from Theorems 4.3 and 4.4. In particular, Theorem 4.3 states $K$-inf-compactness of the cost function $c$, and Theorem 4.4 states weak continuity of the transition kernel $q$ for the COMDP.

**Theorem 4.3:** If the function $c: X \times A \to \mathbb{R}$ is bounded from below $K$-inf-compact function on $X \times A$, then the cost function $\tilde{c}: P(X) \times A \to \mathbb{R}$ defined for the COMDP in (9) is bounded from below $K$-inf-compact function on $P(X) \times A$.

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

**Theorem 4.4:** Let the transition kernel $P(dx'|x, a)$ on $X$ given $X \times A$ be weakly continuous and let the transition kernel $Q(dy|a, x)$ on $Y$ given $A \times X$ be continuous in the total variation. Then: (i) the transition kernel $R'(dy|z, a)$ on $Y$ given $P(X) \times A$ is setwise continuous, Assumption (H) holds, and (ii) the transition kernel $q$ on $P(X)$ given $P(X) \times A$ is weakly continuous.

Theorems 4.2 and 4.4 imply the following result.

**Theorem 4.5:** Let assumptions of (a) and (b) from Theorem 4.2 hold and let the transition kernel $Q(dy|a, x)$ on $Y$ given $A \times X$ be continuous in the total variation. Then statements (i)–(vi) of Theorem 4.1 hold.

**V. 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 such sufficient condition is the continuity of the observation transition kernel in the total variation; see Theorem 4.5. Therefore, this paper provides theoretical foundations to analyze POMDPs with general state and action spaces and with expected total cost criteria.

**Acknowledgements.** The authors thank Dr. Huizhen Janey Yu and Dr. N.V. Zadoianchuk for their useful remarks.

**REFERENCES**


