Zero-Sum Stochastic Games with Partial Information[1]

M. K. Ghosh,[2] D. McDonald,[3] and S. Sinha[4]

Communicated by L. D. Berkovitz

Abstract. We study a zero-sum stochastic game on a Borel state space where the state of the game is not known to the players. Both players take their decisions based on an observation process. We transform this into an equivalent problem with complete information. Then, we establish the existence of a value and optimal strategies for both players.

Key Words. Stochastic games, partial information, value, optimal strategies.

1. Introduction and Preliminaries

Stochastic games were introduced by Shapely (Ref. 1). After this pioneering work, a large number of papers were written on this topic; see Ref. 2 for a survey on zero-sum stochastic games. Most of the available literature in this area falls into the category of stochastic games with complete state information; i.e., at each stage, the state of the game is completely known to the players. Stochastic games are a generalization of Markov decision processes (MDP) in that the latter may be treated as one-player stochastic games. Though there is considerable literature on partially observable Markov decision processes (POMDP, see Refs. 3–5 and references therein), the corresponding results on stochastic games seem to be rather sparse. The

[1]This work was supported by the Indian Institute of Science DRDO Program on Advanced Mathematical Engineering, by NSF under Grants ECS-0218207 and ECS-0225448, and by the Office of Naval Research through the Electric Ship Project. The second and third authors thank the Department of Mathematics, Indian Institute of Science, Bangalore, India for hospitality.
[2]Associate Professor, Department of Mathematics, Indian Institute of Science, Bangalore, India.
[3]Professor, Department of Mathematics, University of Ottawa, Ottawa, Ontario, Canada.
[4]Senior Lecturer, Department of Mathematics, Jadavpur University, Kolkata, India.

0022-3239/04/0400-0099/0 © 2004 Plenum Publishing Corporation

treatment of POMDP is based on estimating the unobservable state using the available information. The conditional distribution of the state, given the available information, is then used as a basis for controlling systems with partial observation. In other words, one introduces a completely observable MDP (COMDP) model where the conditional distribution of the state in the POMDP model, given the available information, constitutes the state in the COMDP model. One can then show that the conditional distributions of the states based on the available information constitute a statistic sufficient for control, as does the available information itself; see Chapter 10 in Ref. 3. This does not seem to be the case for partially observable stochastic games; i.e., the conditional distributions of the states, given the available information, may not constitute a statistic sufficient for decision making. In Ref. 4, the POMDP model is treated in a somewhat different way. Here, the authors introduce also an equivalent COMDP model, but the controls are based on observations and conditional distributions of the unobservable states; i.e., these objects are the states of their COMDP model. In Ref. 3, it is shown that, if there are no control constraints, then there is nothing to be gained by using past observations in the controls. Thus, for POMDP without control constraints, the formulation in Ref. 3 is somewhat simpler and just as effective. Though the treatment of POMDP in Ref. 3 is simpler compared to the one in Ref. 4, it is the latter that can be extended to stochastic games with partial observation. This is precisely what we do in this paper. In this paper, we study a zero-sum stochastic game with partial information on general (uncountable) state and action spaces. We refer to Refs. 6–11 and references therein for related work.

For future reference, let $\mathscr{B}(S)$ and $\mathscr{P}(S)$ denote the Borel σ-field and the space of probability measures on a topological space S. In what follows, all functions are assumed to be Borel measurable.

A partially observable stochastic game (POSG) model is a stochastic dynamical system specified by a collection of objects $(X, Y, A, B, P, Q, Q_0, \psi, r)$ as follows:

(a) X and Y are Borel spaces, X is the state space and Y is the observation space;
(b) A and B are compact metric spaces, which are the action spaces of players 1 and 2;
(c) P is a stochastic kernel on X given $X \times A \times B$, describing the state transition law;
(d) Q is a stochastic kernel on Y given $A \times B \times X$, which is the observation kernel;
(e) Q_0 is a stochastic kernel on Y given X, which is the initial observation kernel;

(f) $\psi \in \mathscr{P}(X)$ is the a priori distribution of the initial state;
(g) $r: X \times Y \times A \times B \to \mathbb{R}$ is the one-stage payoff function.

The state process is described by the process $\{X_t, t \in \mathbb{N}_0\}$ taking values in X, where $\mathbb{N}_0$ is the set of all nonnegative integers. But the players cannot observe the state process. Instead, another process $\{Y_t, t \in \mathbb{N}_0\}$ taking values in Y is available to the players for decision making, and, reflecting this, Y is called the observation space, and $\{Y_t\}$ the observation process.

The game is played over an infinite-time horizon as follows:

(i) At the preinitial stage of the game, the players know $\psi \in \mathscr{P}(X)$, the distribution of the initial (unobservable) state X_0.
(ii) At the 0^{th} decision epoch, X_0 is generated by ψ.
(iii) Conditional on the event that $X_0 = x_0$, the initial observation Y_0 gets generated with the distribution $Q_0(\cdot|x_0)$.
(iv) Conditional on the event that the initial observation is $Y_0 = y_0$, the players choose actions $a_0 \in \mathrm{A}$, $b_0 \in \mathrm{B}$ independently of each other.
(v) Consequently, conditional on the event that $X_0 = x_0$, player 1 gets an immediate reward $r(x_0, y_0, a_0, b_0)$ from player 2; the next state X_1 gets generated by the stochastic kernel $P(\cdot|x_0, a_0, b_0)$; conditional on the event that $X_1 = x_1$, the next observation Y_1 is generated by the kernel $Q(\cdot|a_0, b_0, x_1)$.
(vi) Once the transition to the next state X_1 and observation Y_1 occur, the entire process from (ii) above, with $Y_0 = y_0$ replaced by $Y_1 = y_1$ and ψ replaced by $P(\cdot|x_0, a_0, b_0)$, is repeated over and over again. Thus, the game proceeds over an infinite-time horizon.
(vii) Facts (i)–(vi) above are common knowledge to the players.
(viii) Each player can recall at any time the observations and actions of the past.

We construct now the probability space on which all the random variables are defined. The canonical sample space is defined as

$$\Omega := (X \times Y \times A \times B)^{\infty}.$$

A generic element $\omega \in \Omega$ is of the form

$$\omega = (x_0, y_0, a_0, b_0, x_1, y_1, a_1, b_1, \ldots), \qquad x_i \in X, \quad y_i \in Y, \quad a_i \in A, \quad b_i \in B;$$

all random variables are defined on the measurable space $(\Omega, \mathscr{B}(\Omega))$. The history spaces are defined as

$$H_0 := X \times Y, \qquad H_{t+1} := H_t \times A \times B \times X \times Y, \qquad t \in \mathbb{N}_0.$$

The state, observation, actions, and history processes, denoted by $\{X_t\}$, $\{Y_t\}$, $\{A_t\}$, $\{B_t\}$, $\{H_t\}$ respectively, are defined by the projections

$$X_t(\omega) = x_t, \qquad Y_t(\omega) = y_t,$$
$$A_t(\omega) = a_t, \qquad B_t(\omega) = b_t.$$
$$H_t(\omega) = (x_0, y_0, a_0, b_0, \ldots, a_{t-1}, b_{t-1}, x_t, y_t),$$

for each realization

$$\omega = (x_0, y_0, a_0, b_0, \ldots, a_{t-1}, b_{t-1}, x_t, y_t, \ldots) \in \Omega;$$

A_t and B_t denote the action processes of players 1 and 2. The entire history $(x_0, y_0, a_0, b_0, \ldots, a_{t-1}, b_{t-1}, x_t, y_t)$ at time t is not available to the players for decision making. The players have to make their decisions based only on the observed history or information vector given by $i_t := (y_0, a_0, b_0, \ldots, a_{t-1}, b_{t-1}, y_t)$ and the initial distribution ψ. Thus, we define the information spaces as follows:

$$I_0 := Y, \qquad I_{t+1} := I_t \times A \times B \times Y, \qquad t \in \mathbb{N}_0.$$

The information process $\{I_t\}$ is defined by the projection

$$I_t(\omega) = (y_0, a_0, b_0, \ldots, a_{t-1}, b_{t-1}, y_t),$$

for $\omega = (x_0, y_0, a_0, b_0, \ldots, a_{t-1}, b_{t-1}, x_t, y_t, \ldots) \in \Omega$.

An admissible strategy for player 1 is a sequence

$$\pi^1 = \{\pi_t^1, t \in \mathbb{N}_0\}$$

of (Borel measurable) stochastic kernels π_t^1 on A given $\mathscr{P}(X) \times I_t$. The set of all admissible strategies of player 1 is denoted by Π^1. Similarly, an admissible strategy for player 2 is a sequence

$$\pi^2 = \{\pi_t^2, t \in \mathbb{N}_0\}$$

of (Borel measurable) stochastic kernels π_t^2 on B given $\mathscr{P}(X) \times I_t$. The set of all admissible strategies of player 2 is denoted by Π^2. With $\psi \in \mathscr{P}(X)$ and a pair of admissible strategies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ specified, there exists a unique probability measure $P_\psi^{\pi^1, \pi^2}$ on $(\Omega, \mathscr{B}(\Omega))$ defined by

$$\begin{aligned} &P_\psi^{\pi^1, \pi^2}(dx_0, dy_0, da_0, db_0, \ldots, da_{t-1}, db_{t-1}, dx_t, dy_t) \\ &= \psi(dx_0) Q_0(dy_0 | x_0)\ \pi_0^1(da_0 | \psi, y_0)\ \pi_0^2(db_0 | \psi, y_0)\ P(dx_1 | x_0, a_0, b_0) \\ &Q(dy_1 | a_0, b_0, x_1) \ldots \pi_{t-1}^1(da_{t-1} | \psi, y_0, a_0, b_0, \ldots, y_{t-1}) \\ &\pi_{t-1}^2(db_{t-1} | \psi, y_0, a_0, b_0, \ldots, y_{t-1})\ P(dx_t | x_{t-1}, a_{t-1}, b_{t-1}) \\ &Q(dy_t | a_{t-1}, b_{t-1}, x_t). \end{aligned} \tag{1}$$

We turn now to payoff evaluation over the infinite-time horizon. Let $\beta > 0$ be the discount factor. For an initial distribution $\psi \in \mathscr{P}(X)$ and a pair of strategies $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, the total ($\beta$-discounted) payoff of player 1 by player 2 is given by

$$V_{\pi_1,\pi_2}(\psi) = E_{\psi}^{\pi^1,\pi^2} \sum_{t=0}^{\infty} \beta^t r(X_t, Y_t, A_t^1, A_t^2), \tag{2}$$

where $E_{\psi}^{\pi^1,\pi^2}$ denotes the expectation with respect to $P_{\psi}^{\pi^1,\pi^2}$. A strategy $\pi^{*1} \in \Pi^1$ is said to be optimal for player 1 if

$$V_{\pi^{*1},\pi^2}(\psi) \geq \bar{V}(\psi) := \inf_{\pi^2 \in \Pi^2} \sup_{\pi^1 \in \Pi^1} V_{\pi^1,\pi^2}(\psi), \tag{3}$$

for any $\pi^2 \in \Pi^2$. The function $\bar{V}$ is referred to as the upper value of the game. Similarly, a strategy $\pi^{*2} \in \Pi^2$ is said to be optimal for player 2 if

$$V_{\pi^1,\pi^{*2}}(\psi) \leq \underline{V}(\psi) := \sup_{\pi^1 \in \Pi^1} \inf_{\pi^2 \in \Pi^2} V_{\pi^1,\pi^2}(\psi), \tag{4}$$

for any $\pi^1 \in \Pi^1$. The function $\underline{V}$ is referred to as the lower value of the game. Thus a pair of optimal strategies (π^{*1}, π^{*2}) satisfies

$$V_{\pi^1,\pi^{*2}}(\psi) \leq V_{\pi^{*1},\pi^{*2}}(\psi) \leq V_{\pi^{*1},\pi^2}(\psi),$$

for any $\pi^1 \in \Pi^1, \pi^2 \in \Pi^2$. Hence, (π^{*1}, π^{*2}) constitutes a saddle-point equilibrium.

The POSG is said to have a value if

$$\bar{V}(\psi) = \underline{V}(\psi) := V(\psi). \tag{5}$$

Note that, if both the players have optimal strategies, then the POSG indeed has a value.

In this paper, we prove the existence of a value and optimal strategies for both players. The rest of the paper is organized as follows. In Section 2, we construct the completely observable model and prove some important results. In Section 3, we prove the existence of a value and optimal strategies. We carry out our program under the following assumptions which hold throughout the paper:

(A1) The payoff function $r(\cdot,\cdot,\cdot,\cdot)$ is bounded and continuous.
(A2) The stochastic kernels $P(\cdot|x,a,b)$, $Q(\cdot|a,b,x)$ are weakly continuous in (x,a,b).
(A3) The stochastic kernel $Q_0(\cdot|x)$ is weakly continuous in x.

2. Transformation into a Completely Observable Model

Given a POSG model $(X, Y, A, B, P, Q, Q_0, \psi, r)$, we construct an equivalent completely observable stochastic game (COSG) model

$(Y\times\mathscr{P}(X), A, B, \tilde{P}, \psi, \tilde{r})$ in the following manner. In the COSG model to be constructed, we introduce a preinitial state to carry forward the a priori distribution ψ of the initial unobservable state X_0. Let the canonical sample space in the COSG model be

$$\tilde{\Omega}:=(Y\times\mathscr{P}(X)\times A\times B)^{\infty}.$$

A generic element $\tilde{\omega}\in\tilde{\Omega}$ is of the form

$$\tilde{\omega}=(y_{-1}, \psi_{-1}, a_{-1}, b_{-1}, y_0, \psi_0, a_0, b_0, y_1, \psi_1, a_1, b_1, \ldots),$$

with

$$y_i\in Y, \qquad \psi_i\in\mathscr{P}(X), \qquad a_i\in A, b_i\in B.$$

The history spaces are defined by

$$\tilde{H}_{-1}=\mathscr{P}(X), \quad \tilde{H}_0=Y\times\mathscr{P}(X), \quad \tilde{H}_t=\tilde{H}_{t-1}\times A\times B\times Y\times\mathscr{P}(X), \qquad t=1,2,\ldots.$$

Let $\{Y_t, \Psi_t\}$, $\{A_t\}$, $\{B_t\}$, $\{\tilde{H}_t\}$ denote the state, actions and history processes respectively. These are defined by the usual projections

$$Y_{-1}(\tilde{\omega})=y', \qquad \Psi_{-1}(\tilde{\omega})=\psi, \qquad A_{-1}(\tilde{\omega})=a', \qquad B_{-1}(\tilde{\omega})=b', \qquad \tilde{H}_{-1}(\tilde{\omega})=\psi,$$

where a', b', y' are arbitrarily fixed. For $t=0,1,\ldots,$

$$Y_t(\tilde{\omega})=y_t, \qquad \Psi_t(\tilde{\omega})=\psi_t, \qquad A_t(\tilde{\omega})=a_t, \qquad B_t(\tilde{\omega})=b_t,$$
$$\tilde{H}_t(\tilde{\omega})=(\psi, y_0, \psi_0, a_0, b_0, \psi_1, a_1, b_1, \ldots, a_{t-1}, b_{t-1}, y_t, \psi_t).$$

We construct now the transition kernel $\tilde{P}$ in the COSG model. We define first a stochastic kernel q on $X\times Y$ given $\mathscr{P}(X)\times A\times B$ as follows. For $C\in\mathscr{B}(X)$, $D\in\mathscr{B}(Y)$, we set

$$q(C\times D|\,\psi, a, b)=\int_X\int_C Q(D|a, b, x)P(dx|x', a, b)\psi(dx'), \qquad \psi\in\mathscr{P}(X). \tag{6}$$

Disintegrating q, we obtain

$$q(dx, dy|\,\psi, a, b)=\bar{q}(dy|\,\psi, a, b)\bar{Q}(dx|\,\psi, a, b, y), \tag{7}$$

where $\bar{q}(dy|\psi, a, b)$ is the marginal of $q(dx, dy|\psi, a, b)$ on Y and $\bar{Q}$ $(dx|\psi, a, b, y)$ is a version of the regular conditional law, defined $\bar{q}(dy|\psi, a, b)$ almost surely; we pick any version from this equivalence class and keep it fixed thereafter. Equation (7) is the filtering equation. Along with the probability measure $\bar{q}$ on Y, $\bar{Q}$ induces a probability measure $\tilde{P}$ on $Y\times\mathscr{P}(X)$, which is a measurable function of (ψ, a, b). In other words, it defines a (Borel measurable) stochastic kernel on $Y\times\mathscr{P}(X)$ given $\mathscr{P}(X)\times A\times B$ defined by

$$\tilde{P}(D\times\Lambda|\psi,a,b):=\int_{\mathbf{D}} I\{\bar{Q}(\cdot\,|\psi,a,b,y)\in\Lambda\}\bar{q}(dy|\psi,a,b),$$
$$D\in\mathscr{B}(Y),\ \Lambda\in\mathscr{B}(\mathscr{P}(X)). \tag{8}$$

The initial distribution Ξ of (Y_0,Ψ_0) in the COSG model corresponding to the initial distribution ψ of X_0 and the initial observation kernel $Q_0(\cdot|x)$ in the POSG model is constructed as follows. Let $q_0(\cdot|\psi)$ be a stochastic kernel on $\mathscr{P}(X\times Y)$ given $\mathscr{P}(X)$ defined by

$$q_0(C\times D|\psi)=\int_C Q_0(D|x)\psi(dx), \qquad C\in\mathscr{B}(X), \qquad D\in\mathscr{B}(Y).$$

Disintegrate $q_0(\cdot|\psi)$ as

$$q_0(dx,dy|\psi)=\bar{q}_0(dy|\psi)\bar{Q}_0(dx|\psi,y), \tag{9}$$

where $\bar{q}_0(dy|\psi)$ is the marginal of $q_0(dx,dy|\psi)$ on Y and $\bar{Q}_0(dx|\psi,y)$ is a version of the regular conditional law. Define Ξ by

$$\Xi(D\times\Lambda):=\int_D I\{\bar{Q}_0(\cdot|\psi,y)\in\Lambda\}\bar{q}_0(dy|\psi), \qquad D\in\mathscr{B}(Y),\ \Lambda\in\mathscr{B}(\mathscr{P}(X)). \tag{10}$$

The above equation establishes a map between $\mathscr{P}(X)$ and $\mathscr{P}(Y\times\mathscr{P}(\mathrm{X}))$. Let this map be denoted by Φ; i.e.,

$$\Phi(\psi)=\Xi. \tag{11}$$

It is easy to see that Φ is Borel measurable. The payoff function $\tilde{r}$ in the COSG model is defined by

$$\tilde{r}(y,\psi,a,b)=\int_X r(x,y,a,b)\psi(dx), \qquad \psi\in\mathscr{P}(X), y\in Y, a\in A, b\in B. \tag{12}$$

The game in the COSG model is played over an infinite-time horizon as follows:

(i) At the preinitial epoch (i.e., at $t=-1$), the a priori distribution ψ is known to both the players, but they do not take any action based on this.
(ii) At the 0^{th} decision epoch, the initial state (Y_0,Ψ_0) is generated by the distribution $\Phi(\psi)\in\mathscr{P}(Y\times\mathscr{P}(X))$.
(iii) Conditional on the event that $Y_0=y_0, \Psi_0=\psi_0$, the players choose actions $a_0\in A, b_0\in B$ independently of each other.
(iv) Consequently, player 1 gets an immediate reward $\tilde{r}(y_0,\psi_0,a_0,b_0)$ from player 2; the next state (Y_1,Ψ_1) gets generated by the stochastic kernel $\tilde{P}(\cdot|\psi_0,a_0,b_0)$.

(v) Once the transition to the next state (Y_1, Ψ_1) occurs, the entire process from (ii) above, with $(Y_0 = y_0, \Psi_0 = \psi_0)$ replaced by $(Y_1 = y_1, \Psi_1 = \psi_1)$ and $\Phi(\psi)$ replaced by $P(\cdot|\psi_0, a_0, b_0)$, is repeated over and over again. Thus, the game proceeds over an infinite-time horizon.

(vi) All the facts (i)–(v) above are common knowledge to the players.

(vii) Each player can recall at any time the state and actions of the past. Thus, the COSG model is a stochastic game with perfect recall.

In the COSG model, the entire history is available to the players for decision making. Thus, an admissible strategy for player 1 is a sequence

$$\tilde{\pi}^1 = \{\tilde{\pi}_t^1, t \in \mathbb{N}_0\}$$

of (Borel measurable) stochastic kernels $\tilde{\pi}_t^1$ on A given $\tilde{H}_t$. The set of all admissible strategies of player 1 is denoted by $\tilde{\Pi}^1$. Similarly, an admissible strategy for player 2 is a sequence

$$\tilde{\pi}^2 = \{\tilde{\pi}_t^2, t \in \mathbb{N}_0\}$$

of (Borel measurable) stochastic kernels $\tilde{\pi}_t^2$ on B given $\tilde{H}_t$. The set of all admissible strategies of player 2 is denoted by $\tilde{\Pi}^2$.

An admissible strategy $\tilde{\pi}_1$ of player 1 in the COSG model is said to be Markov if there exists a sequence of measurable maps $\{\lambda_t\}$, λ_t: $Y \times \mathscr{P}(X) \rightarrow \mathscr{P}(A)$ such that

$$\tilde{\pi}_1(\cdot|\tilde{h}_t) = \lambda_t(y_t, \psi_t)(\cdot).$$

The set of all Markov strategies for player 1 is denoted by $\tilde{\Pi}_M^1$. A Markov strategy $\{\lambda_t\}$ of player 1 is said to be stationary if there exists a measurable map λ: $Y \times \mathscr{P}(X) \rightarrow \mathscr{P}(A)$ such that $\lambda_t = \lambda$ for all t. With an abuse of terminology, the map λ itself is referred to as the stationary strategy of player 1. The set of all stationary strategies of player 1 in the COSG model is denoted by $\tilde{\Pi}_S^1$. Markov and stationary strategies for player 2 are defined analogously. Let $\tilde{\Pi}_M^2$ and $\tilde{\Pi}_S^2$ denote the set of all Markov strategies and stationary strategies of player 2.

Remark 2.1. Note that the Markov strategies and stationary strategies do not make much sense in the POSG model, since the initial distribution, past observations, and past actions of the players are used in estimating the current state. Thus, we require the strategies to depend on the initial distribution and the information vector.

Let $\psi \in \mathscr{P}(X)$ and let $(\tilde{\pi}^1, \tilde{\pi}^2) \in \tilde{\Pi}^1 \times \tilde{\Pi}^2$ be specified. Then, there exists a unique probability measure $P_\psi^{\tilde{\pi}^1, \tilde{\pi}^2}$ on $(\tilde{\Omega}, \mathscr{B}(\tilde{\Omega}))$, given by

$$\begin{aligned} &P_\psi^{\tilde{\pi}^1, \tilde{\pi}^2}(d(y_{-1}, \psi_{-1}), da_{-1}, db_{-1}, d(y_0, \psi_0), da_0, db_0, \ldots, da_{t-1}, db_{t-1}, d(y_t, \psi_t)) \\ &= \delta_{(y', \psi, a', b')} \Phi(\psi)(d(y_0, \psi_0)) \tilde{\pi}_0^1(da_0 | y_0, \psi_0) \tilde{\pi}_0^2(db_0 | y_0, \psi_0) \\ &\quad \tilde{P}(d(y_1, \psi_1) | y_0, \psi_0, a_0, b_0) \ldots \tilde{\pi}_{t-1}^1(da_{t-1} | y_0, \psi_0, a_0, b_0, \ldots, y_{t-1}, \psi_{t-1}) \\ &\quad \tilde{\pi}_{t-1}^2(db_{t-1} | y_0, \psi_0, a_0, b_0, \ldots, y_{t-1}, \psi_{t-1})\, \tilde{P}(d(y_t, \psi_t) | y_{t-1}, \psi_{t-1}, a_{t-1}, b_{t-1}), \end{aligned} \tag{13}$$

where $\delta_{(y', \psi, a', b')}$ is the Dirac measure at (y', ψ, a', b').

Clearly, the model $(Y \times \mathscr{P}(X), A, B, \tilde{P}, \psi, \tilde{r})$ thus constructed is a completely observable stochastic game model.

For $\psi \in \mathscr{P}(X)$ and a pair of strategies $(\tilde{\pi}^1, \tilde{\pi}^2) \in \tilde{\Pi}^1 \times \tilde{\Pi}^2$, the total ($\beta$-discounted) payoff of player 1 by player 2 in the (COSG) model is given by

$$\tilde{V}_{\tilde{\pi}_1, \tilde{\pi}_2}(\psi) = \tilde{E}_\psi^{\tilde{\pi}^1, \tilde{\pi}^2} \sum_{t=0}^{\infty} \beta^t \tilde{r}(Y_t, \Psi_t, A_t^1, A_t^2), \tag{14}$$

where $\tilde{E}_\psi^{\tilde{\pi}^1, \tilde{\pi}^2}$ denotes the expectation with respect to $\tilde{P}_\psi^{\tilde{\pi}^1, \tilde{\pi}^2}$. A strategy $\tilde{\pi}^{*1} \in \tilde{\Pi}^1$ is said to be optimal for player 1 in the COSG model if

$$\tilde{V}_{\tilde{\pi}^{*1}, \tilde{\pi}^2}(\psi) \geq \bar{\tilde{V}}(\psi) := \inf_{\tilde{\pi}^2 \in \tilde{\Pi}^2} \sup_{\tilde{\pi}^1 \in \tilde{\Pi}^1} \tilde{V}_{\tilde{\pi}^1, \tilde{\pi}^2}(\psi), \tag{15}$$

for any $\tilde{\pi}^2 \in \tilde{\Pi}^2$. The function $\bar{\tilde{V}}$ is referred to as the upper value of the game in the COSG model.

Similarly, a strategy $\tilde{\pi}^{*2} \in \tilde{\Pi}^2$ is said to be optimal for player 2 (for the initial distribution Ξ) in the COSG model if

$$\tilde{V}_{\tilde{\pi}^1, \tilde{\pi}^{*2}}(\psi) \leq \underline{\tilde{V}}(\psi) := \sup_{\tilde{\pi}^1 \in \tilde{\Pi}^1} \inf_{\tilde{\pi}^2 \in \tilde{\Pi}^2} \tilde{V}_{\tilde{\pi}^1, \tilde{\pi}^2}(\psi), \tag{16}$$

for any $\tilde{\pi}^1 \in \tilde{\Pi}^1$. The function $\underline{\tilde{V}}$ is referred to as the lower value of the game in the COSG model.

The COSG is said to have a value if

$$\bar{\tilde{V}}(\psi) = \underline{\tilde{V}}(\psi) := \tilde{V}(\psi). \tag{17}$$

Given a $\psi \in \mathscr{P}(X)$ and an information vector

$$i_t = (y_0, a_0, b_0, \ldots, y_{t-1}, a_{t-1}, b_{t-1}, y_t) \in I_t$$

in the POSG model, we construct $\psi_0, \psi_1, \ldots$ in a recursive manner from (7), (9) in the following way. Set

$$\begin{aligned} \psi_0 &= \psi_0(\psi, i_0) = \bar{Q}_0(\cdot | \psi, y_0), \\ \psi_{s+1} &= \psi_{s+1}(\psi, i_{s+1}) = \bar{Q}(\cdot | \psi_s, a_s, b_s, y_{s+1}), \qquad \text{for } s = 0, 1, \ldots, t-1. \end{aligned} \tag{18}$$

We can modify the arguments from Chapter 8 in Ref. 4 to obtain the following result in the POSG model. Since this result plays a crucial role in establishing its equivalence with the COSG model, we present its proof. Let

$$\mathscr{F}_t^I = \sigma(I_s, s \leq t).$$

Lemma 2.1. Let $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and let $\psi \in \mathscr{P}(X)$. Then, for any $t \in \mathbb{N}_0$, $C \in \mathscr{B}(X)$,

$$P_\psi^{\pi^1, \pi^2}(X_t \in C | \mathscr{F}_t^I) = \psi_t(\psi, I_t)(C), \qquad P_\psi^{\pi^1, \pi^2} - \text{a.s.} \tag{19}$$

Proof. We first prove (19) for $t = 0$. Let $C \in \mathscr{B}(X)$, $D \in \mathscr{B}(Y)$. Using (1), (6), (7), (9), (18), and the Fubini theorem, we have

$$\begin{aligned}
&\int_{\{I_0(\omega) \in D\}} \psi_0(\psi, I_0(\omega))(C) P_\psi^{\pi^1, \pi^2}(d\omega) \\
&= \int_{\{x_0 \in X,\, y_0 \in D\}} \psi_0(\psi, y_0)(C) P_\psi^{\pi^1, \pi^2}(dx_0, dy_0) \\
&= \int_D \int_X \bar{Q}_0(C | \psi, y_0) \psi(dx_0) Q_0(dy_0 | x_0) \\
&= \int_D \int_X \bar{Q}_0(C | \psi, y_0) q_0(dx_0, dy_0 | \psi) \\
&= \int_D \bar{Q}_0(C | \psi, y_0) \bar{q}_0(dy_0 | \psi) = q_0(C \times D | \psi) \\
&= \int_C Q_0(D | x_0) \psi(dx_0) = P_\psi^{\pi^1, \pi^2}(X_0 \in C,\, Y_0 \in D).
\end{aligned}$$

This proves (19) for $t = 0$. Assume that (19) holds for some $t > 0$; we prove it for $t + 1$. Let

$$C \in \mathscr{B}(X), \qquad D \in \mathscr{B}(Y), \qquad U \in \mathscr{B}(A), \qquad V \in \mathscr{B}(B), \qquad I \in \mathscr{B}(I_t).$$

Then, using the induction hypothesis, the Fubini theorem, (1), (6), (7), (18), we have

$$\begin{aligned}
&\int_{\{I_t(\omega) \in I,\, A_t(\omega) \in U,\, B_t(\omega) \in V,\, Y_{t+1}(\omega) \in D\}} \psi_{t+1}(\psi, I_{t+1}(\omega))(C) P_\psi^{\pi^1, \pi^2}(d\omega) \\
&= \int_{\{i_t \in I,\, a_t \in U,\, b_t \in V,\, y_{t+1} \in D,\, x_0, \ldots, x_{t+1} \in X\}} \psi_{t+1}(\psi, i_t, a_t, b_t, y_{t+1})(C) \\
&\quad P_\psi^{\pi^1, \pi^2}(dx_0, dy_0, da_0, db_0, \ldots, da_t, db_t, dx_{t+1}, dy_{t+1}) \\
&= \int_{\{i_t \in I,\, x_0, \ldots, x_t \in X\}} \int_V \int_U \int_X \int_D \psi_{t+1}(\psi, i_t, a_t, b_t, y_{t+1})(C) Q(dy_{t+1} | a_t, b_t, x_{t+1})
\end{aligned}$$

$$P(dx_{t+1}|x_t, a_t, b_t)\pi_t^1(da_t|\psi, i_t)\pi_t^2(db_t|\psi, i_t)$$

$$P_\psi^{\pi^1,\pi^2}(dx_0, dy_0, da_0, db_0, \ldots, da_{t-1}, db_{t-1}, dx_t, dy_t)$$

$$= \int_{\{i_t \in I, x_0, \ldots, x_{t-1} \in X\}} \int_X \int_V \int_U \int_X \int_D \psi_{t+1}(\psi, i_t, a_t, b_t, y_{t+1})(C) Q(dy_{t+1}|a_t, b_t, x_{t+1})$$

$$P(dx_{t+1}|x_t, a_t, b_t)\pi_t^1(da_t|\psi, i_t)\pi_t^2(db_t|\psi, i_t)\psi_t(\psi, i_t)(dx_t)$$

$$P_\psi^{\pi^1,\pi^2}(dx_0, dy_0, da_0, db_0, \ldots, da_{t-1}, db_{t-1}, dy_t)$$

$$= \int_{\{i_t \in I, x_0, \ldots, x_{t-1} \in X\}} \int_V \int_U \int_X \int_X \int_D \bar{Q}(C|\psi_t(\psi, i_t), a_t, b_t, y_{t+1}) Q(dy_{t+1}|a_t, b_t, x_{t+1})$$

$$P(dx_{t+1}|x_t, a_t, b_t)\psi_t(\psi, i_t)(dx_t)\pi_t^1(da_t|\psi, i_t)\pi_t^2(db_t|\psi, i_t)$$

$$P_\psi^{\pi^1,\pi^2}(dx_0, dy_0, da_0, db_0, \ldots, da_{t-1}, db_{t-1}, dy_t)$$

$$= \int_{\{i_t \in I, x_0, \ldots, x_{t-1} \in X\}} \int_V \int_U \int_D \bar{Q}(C|\psi_t(\psi, i_t), a_t, b_t, y_{t+1}) \bar{q}(dy_{t+1}|\psi_t(\psi, i_t), a_t, b_t)$$

$$\pi_t^1(da_t|\psi, i_t)\pi_t^2(db_t|\psi, i_t)P_\psi^{\pi^1,\pi^2}(dx_0, dy_0, da_0, db_0, \ldots, da_{t-1}, db_{t-1}, dy_t)$$

$$= \int_{\{i_t \in I, x_0, \ldots, x_{t-1} \in X\}} \int_V \int_U q(C \times D|\psi_t(\psi, i_t), a_t, b_t)$$

$$\pi_t^1(da_t|\psi, i_t)\pi_t^2(db_t|\psi, i_t)P_\psi^{\pi^1,\pi^2}(dx_0, dy_0, da_0, db_0, \ldots, da_{t-1}, db_{t-1}, dy_t)$$

$$= \int_{\{i_t \in I, x_0, \ldots, x_{t-1} \in X\}} \int_V \int_U \int_X \int_C Q(D|a_t, b_t, x_{t+1}) P(dx_{t+1}|x_t, a_t, b_t)$$

$$\psi_t(\psi, i_t)(dx_t)\pi_t^1(da_t|\psi, i_t)\pi_t^2(db_t|\psi, i_t)P_\psi^{\pi^1,\pi^2}$$
$$\times (dx_0, dy_0, da_0, db_0, \ldots, da_{t-1}, db_{t-1}, dy_t)$$

$$= \int_{\{i_t \in I, x_0, \ldots, x_{t-1} \in X\}} \int_X \int_V \int_U \int_C Q(D|a_t, b_t, x_{t+1}) P(dx_{t+1}|x_t, a_t, b_t)$$

$$\pi_t^1(da_t|\psi, i_t)\pi_t^2(db_t|\psi, i_t)\psi_t(\psi, i_t)(dx_t)P_\psi^{\pi^1,\pi^2}$$
$$\times (dx_0, dy_0, da_0, db_0, \ldots, da_{t-1}, db_{t-1}, dy_t)$$

$$= \int_{\{i_t \in I, x_0, \ldots, x_t \in X\}} \int_V \int_U \int_C Q(D|a_t, b_t, x_{t+1}) P(dx_{t+1}|x_t, a_t, b_t)$$

$$\pi_t^1(da_t|\psi, i_t)\pi_t^2(db_t|\psi, i_t)P_\psi^{\pi^1,\pi^2}(dx_0, dy_0, da_0, db_0, \ldots, da_{t-1}, db_{t-1}, dx_t, dy_t)$$

$$= P_\psi^{\pi^1,\pi^2}(I_t \in I, A_t \in U, B_t \in V, X_{t+1} \in C, Y_{t+1} \in D).$$

Equation (19) follows from the above for $t + 1$. □

The following results follow immediately from the above lemma.

Corollary 2.1. Let $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, $\psi \in \mathscr{P}(X)$, $D \in \mathscr{B}(Y)$, $\Lambda \in \mathscr{B}(\mathscr{P}(X))$. Then,

$$P_\psi^{\pi^1, \pi^2}(Y_{t+1} \in D, \psi_{t+1}(\psi, I_{t+1}) \in \Lambda | \mathscr{F}_t^I) = \tilde{P}(D \times \Lambda | \psi_t(\psi, I_t), A_t, B_t), \qquad P_\psi^{\pi^1, \pi^2} - \text{a.s.} \tag{20}$$

Corollary 2.2. Let $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, $\psi \in \mathscr{P}(X)$. Then,

$$E_\psi^{\pi^1, \pi^2}[r(X_t, Y_t, A_t, B_t) | \mathscr{F}_t^I] = \tilde{r}(Y_t, \psi_t(\psi, I_t), A_t, B_t), \qquad P_\psi^{\pi^1, \pi^2} - \text{a.s.} \tag{21}$$

Given

$$\psi \in \mathscr{P}(X) \quad \text{and} \quad i_t = (y_0, a_0, b_0, \ldots, y_{t-1}, a_{t-1}, b_{t-1}, y_t) \in I_t,$$

we use (18) to obtain

$$(\psi, \tilde{h}_t) = (\psi, y_0, \psi_0, a_0, b_0, \ldots, y_{t-1}, \psi_{t-1}, a_{t-1}, b_{t-1}, y_t, \psi_t) \in \mathscr{P}(X) \times \tilde{H}_t.$$

We denote this correspondence by the map $g_t: \mathscr{P}(X) \times I_t \to \tilde{H}_t$. We can then assign to each strategy $\tilde{\pi}^j \in \tilde{\Pi}^j$, $j = 1, 2$, in the COSG model a corresponding strategy $\pi^j = g^*(\tilde{\pi}^j)$ in the POSG model, defined by

$$\pi_t^j(\cdot | \psi, i_t) := \tilde{\pi}_t^j(\cdot | g_t(\psi, i_t)). \tag{22}$$

Lemma 2.2. The map g^* is onto.

Proof. For each $t \in \mathbb{N}_0$, $\mathscr{P}(X) \times I_t$ can be imbedded into $\tilde{H}_t$ in an obvious manner. Thus, Π^j can be imbedded into $\tilde{\Pi}^j$, $j = 1, 2$. Let i denote this imbedding. Clearly,

$$(g^* \circ i)(\pi^j) = \pi^j, \qquad \text{for any } \pi^j \in \Pi^j, \ j = 1, 2. \qquad \square$$

Lemma 2.3. Let $\psi \in \mathscr{P}(\mathbf{X})$, $(\tilde{\pi}^1, \tilde{\pi}^2) \in \tilde{\Pi}^1 \times \tilde{\Pi}^2$. Let $(\pi^1, \pi^2) = (g^*(\tilde{\pi}_1), g^*(\tilde{\pi}_2))$ be the corresponding strategies in the POSG model. Then,

$$V_{\pi^1, \pi^2}(\psi) = \tilde{V}_{\tilde{\pi}^1, \tilde{\pi}^2}(\psi). \tag{23}$$

Proof. In view of Corollary 2.2, it suffices to show that, for any $t \in \mathbb{N}_0$, $D_t \in \mathscr{B}(Y)$, $\Lambda_t \in \mathscr{B}(\mathscr{P}(X))$, $U_t \in \mathscr{B}(A)$, $V_t \in \mathscr{B}(B)$, we have

$$\begin{aligned} &P_\psi^{\pi^1,\pi^2}(Y_0 \in D_0, \psi_0(\psi, I_0) \in \Lambda_0, A_0 \in U_0, B_0 \in V_0, \ldots, \\ &\quad Y_t \in D_t, \psi_t(\psi, I_t) \in \Lambda_t, A_t \in U_t, B_t \in V_t) \\ &= P_\psi^{\tilde{\pi}^1,\tilde{\pi}^2}(Y_0 \in D_0, \Psi_0 \in \Lambda_0, A_0 \in U_0, B_0 \in V_0, \ldots, \\ &\quad Y_t \in D_t, \Psi_t \in \Lambda_t, A_t \in U_t, B_t \in V_t). \end{aligned} \tag{24}$$

Indeed, for $t = 0$, (24) follows from (1), (9), (11), (13). If we assume that (24) holds for some $t > 0$, then again using (1), (6), (7), (13), Corollary 2.1, and the induction hypothesis, we can show that (24) does hold for $t + 1$. □

We prove now the main result of this section.

Theorem 2.1. If the COSG model has a value, then the POSG model has a value and

$$V(\psi) = \tilde{V}(\psi), \tag{25}$$

for any $\psi \in \mathscr{P}(\mathbf{X})$. Furthermore, if $(\tilde{\pi}^1, \tilde{\pi}^2) \in \tilde{\Pi}^1 \times \tilde{\Pi}^2$ is a pair of saddle-point strategies in the COSG model, then $(\pi^1, \pi^2) = (g^*(\tilde{\pi}_1), g^*(\tilde{\pi}_2))$ is a pair of saddle-point strategies in the POSG model.

Proof. Let $\epsilon > 0$. Then, from the definition of $\bar{\tilde{V}}(\psi)$ $(= \tilde{V}(\psi)$, by assumption), there exists $\tilde{\pi}^{2,\epsilon} \in \tilde{\Pi}^2$ such that

$$\tilde{V}_{\tilde{\pi}^1, \tilde{\pi}^{2,\epsilon}}(\psi) \leq \tilde{V}(\psi) + \epsilon, \tag{26}$$

for every $\tilde{\pi}^1 \in \tilde{\Pi}^1$. Now, using (26) together with Lemma 2.2 and Lemma 2.3, we obtain

$$\bar{V}(\psi) \leq \tilde{V}(\psi) + \epsilon.$$

Since $\epsilon > 0$ is arbitrary, it follows that

$$\bar{V}(\psi) \leq \tilde{V}(\psi). \tag{27}$$

Similarly, we can show that

$$\underline{\mathrm{V}}(\psi) \geq \tilde{V}(\psi). \tag{28}$$

Now, (25) follows from (27) and (28). The rest can be proved using Lemma 2.3. □

Some comments are in order.

Remark 2.2.

(i) Using the above arguments, we can show that, if the POSG model has a value, then the COSG model has also a value. Similarly, if $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ is a pair of saddle point strategies in the POSG model, then any $(\tilde{\pi}^1, \tilde{\pi}^2) \in \tilde{\Pi}^1 \times \tilde{\Pi}^2$ belonging to the fibre $(g^{*-1}(\pi^1), g^{*-1}(\pi^2))$ is a pair of saddle point strategies for the COSG model. But this result may not be very useful from the practical point of view, since in the POSG model state estimation is done at each stage using the initial distribution and the information vector. Thus, the processes of state estimation and taking actions are linked together, whereas in the COSG model these two aspects are separated out or delinked. Hence, the COSG model is more useful from the practical viewpoint.

(ii) It may be interesting to compare the equivalence of COSG and POSG models with the corresponding result in partially-observable Markov decision processes (POMDP). In the POMDP model without control constraints, the state estimate process ψ_t summarizes all the information relevant for decision making (cf. Chapter 10 in Ref. 3) whereas in the stochastic game problem this does not seem to be the case. Indeed, in Ref. 3, the state estimate process ψ_t constitutes the (sole) state process in the corresponding completely observable MDP (COMDP). In such a setup, one cannot claim that the map g^* as in (22) is onto. The equivalence of the POMDP and COMDP models is established in Ref. 3 in a different way. Given a strategy in the POMDP model, one disintegrates the joint distribution of the corresponding state and action processes. This yields a Markov strategy in the COMDP model such that the corresponding costs in these two models are equal; see the proof of Lemma 10.2, pp. 255–256 in Ref. 3. Together with the counterpart of Corollary 2.2, this result completes the equivalence of POMDP and the corresponding COMDP models. Note that the technique of disintegrating the joint distribution of state and action processes to obtain Markov strategies cannot be extended to the case where there are two or more players who do not cooperate with one another. Thus, the methodology of Chapter 10 in Ref. 3 cannot be extended to POSG.

(iii) The treatment of the POMDP problem in Ref. 4 is different from that in Ref. 3. In the latter, both the state estimates and the observation constitute the states in the corresponding COMDP model. Even though, in the absence of control constraints, the observation process is not really necessary for control purposes in the COMDP model, this approach has the advantage that it can be generalized to POSG. Thus, following Ref. 4, we have incorporated the observation process Y_t together with the state estimate process ψ_t in the COSG model to establish its equivalence with the POSG models. We have also incorporated the a priori distribution ψ in the COSG model. Our construction of the COSG model in stochastic games with partial observation differs from that in Ref. 4. In Ref. 4, the authors assume that an a priori distribution of (X_0, Y_0) is given. This distribution is then disintegrated to obtain an a posteriori (conditional) distribution of X_0 given Y_0. This a posteriori distribution is added in the observable history. In our case, we assume that an a priori distribution of X_0 is given. We do not add this information in the observable history. Instead we carry forward this distribution to the COSG model by introducing a pre-initial state. Thus, our construction of the COSG model is somewhat different from that in Ref. 4.

The following results may be of independent interest. The COSG model is said to have a value in Markov strategies if

$$\inf_{\tilde{\pi}^2 \in \tilde{\Pi}^2_M} \sup_{\tilde{\pi}^1 \in \tilde{\Pi}^1_M} \tilde{V}_{\tilde{\pi}^1, \tilde{\pi}^2}(\psi) = \sup_{\tilde{\pi}^1 \in \tilde{\Pi}^1_M} \inf_{\tilde{\pi}^2 \in \tilde{\Pi}^2_M} \tilde{V}_{\tilde{\pi}^1, \tilde{\pi}^2}(\psi) := \tilde{V}_M(\psi). \tag{29}$$

Similarly, we say that the COSG model has a value in stationary strategies if

$$\inf_{\tilde{\pi}^2 \in \tilde{\Pi}^2_S} \sup_{\tilde{\pi}^1 \in \tilde{\Pi}^1_S} \tilde{V}_{\tilde{\pi}^1, \tilde{\pi}^2}(\psi) = \sup_{\tilde{\pi}^1 \in \tilde{\Pi}^1_S} \inf_{\tilde{\pi}^2 \in \tilde{\Pi}^2_S} \tilde{V}_{\tilde{\pi}^1, \tilde{\pi}^2}(\psi) := \tilde{V}_S(\psi). \tag{30}$$

Proposition 2.1. The COSG model has a value if and only if it has value in Markov strategies and in that case the two values are equal.

Proof. We first show that, if the COSG model has a value in Markov strategies, then the COSG has a value and

$$\tilde{V}(\psi) = \tilde{V}_M(\psi). \tag{31}$$

Let

$$(\tilde{\pi}^1, \tilde{\pi}^2) \in \tilde{\Pi}^1_M \times \tilde{\Pi}^2 \quad \text{and} \quad \Xi \in \mathscr{P}(Y \times \mathscr{P}(X)).$$

For $t \in \mathbb{N}_0, D \in \mathscr{B}(Y), \Lambda \in \mathscr{B}(\mathscr{P}(X)), U \in \mathscr{B}(A), V \in \mathscr{B}(B)$, define a probability measure $\nu_t \in \mathscr{P}(Y \times \mathscr{P}(X) \times A \times B)$ by

$$\nu_t(D \times \Lambda \times U \times V) = P_{\Xi}^{\tilde{\pi}^1, \tilde{\pi}^2}(Y_t \in D, \Psi_t \in \Lambda, A_t \in U, B_t \in V). \tag{32}$$

Let $\tilde{\nu}_t$ be the marginal of ν_t on $\mathscr{P}(Y \times \mathscr{P}(X) \times B)$. Disintegrate $\tilde{\nu}_t$ as

$$\tilde{\nu}_t(d(y_t, \psi_t), db_t) = \bar{\nu}_t(d(y_t, \psi_t))\tilde{\pi}_t'^2(db_t | y_t, \psi_t). \tag{33}$$

Clearly,

$$\tilde{\pi}'^2 := \{\tilde{\pi}_t'^2\} \in \tilde{\Pi}^2_M.$$

From the construction of $\tilde{\pi}'^2$, one can use the arguments in Ref. 12 to verify that

$$\tilde{V}_{\tilde{\pi}^1, \tilde{\pi}^2}(\psi) = \tilde{V}_{\tilde{\pi}^1, \tilde{\pi}'^2}(\psi). \tag{34}$$

It follows from (34) that

$$\sup_{\tilde{\pi}^1\in\tilde{\Pi}^1_M} \inf_{\tilde{\pi}^2\in\tilde{\Pi}^2} \tilde{V}_{\tilde{\pi}^1,\tilde{\pi}^2}(\psi) = \sup_{\tilde{\pi}^1\in\tilde{\Pi}^1_M} \inf_{\tilde{\pi}^2\in\tilde{\Pi}^2_M} \tilde{V}_{\tilde{\pi}^1,\tilde{\pi}^2}(\psi). \tag{35}$$

Similarly, we can show that

$$\inf_{\tilde{\pi}^2\in\tilde{\Pi}^2_M} \sup_{\tilde{\pi}^1\in\tilde{\Pi}^1} \tilde{V}_{\tilde{\pi}^1,\tilde{\pi}^2}(\psi) = \inf_{\tilde{\pi}^2\in\tilde{\Pi}^2_M} \sup_{\tilde{\pi}^1\in\tilde{\Pi}^1_M} \tilde{V}_{\tilde{\pi}^1,\tilde{\pi}^2}(\psi). \tag{36}$$

Clearly,

$$\inf_{\tilde{\pi}^2\in\tilde{\Pi}^2_M} \sup_{\tilde{\pi}^1\in\tilde{\Pi}^1} \tilde{V}_{\tilde{\pi}^1,\tilde{\pi}^2}(\psi) \geq \inf_{\tilde{\pi}^2\in\tilde{\Pi}^2} \sup_{\tilde{\pi}^1\in\tilde{\Pi}^1} \tilde{V}_{\tilde{\pi}^1,\tilde{\pi}^2}(\psi), \tag{37}$$

$$\sup_{\tilde{\pi}^1\in\tilde{\Pi}^1_M} \inf_{\tilde{\pi}^2\in\tilde{\Pi}^2} \tilde{V}_{\tilde{\pi}^1,\tilde{\pi}^2}(\psi) \leq \sup_{\tilde{\pi}^1\in\tilde{\Pi}^1} \inf_{\tilde{\pi}^2\in\tilde{\Pi}^2} \tilde{V}_{\tilde{\pi}^1,\tilde{\pi}^2}(\psi). \tag{38}$$

By our assumption,

$$\inf_{\tilde{\pi}^2\in\tilde{\Pi}^2_M} \sup_{\tilde{\pi}^1\in\tilde{\Pi}^1_M} \tilde{V}_{\tilde{\pi}^1,\tilde{\pi}^2}(\psi) = \sup_{\tilde{\pi}^1\in\tilde{\Pi}^1_M} \inf_{\tilde{\pi}^2\in\tilde{\Pi}^2_M} \tilde{V}_{\tilde{\pi}^1,\tilde{\pi}^2}(\psi). \tag{39}$$

Since $\inf\sup \geq \sup\inf$, (31) follows from (35)–(39). The converse can be proved using analogous arguments. □

Proposition 2.2. The COSG model has a value if and only if it has value in stationary strategies and in that case the two values are equal.

Proof. Let $(\tilde{\pi}^1, \tilde{\pi}^2) \in \tilde{\Pi}^1_S \times \tilde{\Pi}^2$. Let $\psi \in \mathscr{P}(X)$ be fixed. Define a measure $\mu \in \mathscr{P}(Y \times \mathscr{P}(X) \times A \times B)$ implicitly by

$$\int_{Y\times\mathscr{P}(X)\times A\times B} f(y, \psi, a, b)\mu(d(y, \psi), da, db)$$
$$= (1-\beta)\tilde{E}^{\tilde{\pi}^1,\tilde{\pi}^2}_{\psi} \sum_{t=0}^{\infty} \beta^t f(Y_t, \Psi_t, A_t, B_t). \tag{40}$$

Note that, in terms of the measure μ thus defined,

$$\tilde{V}_{\tilde{\pi}^1,\tilde{\pi}^2}(\Xi) = (1-\beta)^{-1} \int_{Y\times\mathscr{P}(X)\times A\times B} \tilde{r}(y, \psi, a, b)\mu(d(y, \psi), da, db). \tag{41}$$

Let $\tilde{\mu}$ be the marginal of μ on $\mathscr{P}(Y \times \mathscr{P}(X) \times B)$. Disintegrate $\tilde{\mu}$ as

$$\tilde{\mu}(d(y, \psi), db) = \bar{\mu}(d(y, \psi))\tilde{\pi}'^2(db|y, \psi). \tag{42}$$

Clearly,

$$\tilde{\pi}'^2 := \{\tilde{\pi}'^2\} \in \tilde{\Pi}_S^2.$$

From the construction of $\tilde{\pi}'^2$, one can verify using the arguments in Ref. 12 that

$$\tilde{V}_{\tilde{\pi}^1, \tilde{\pi}^2}(\psi) = \tilde{V}_{\tilde{\pi}^1, \tilde{\pi}'^2}(\psi). \tag{43}$$

We can now mimic the arguments used in the proof of the Proposition 2.1 to complete the proof of this proposition. $\square$

3. Existence of Value and Optimal Strategies

In this section, we establish first the value and optimal strategies for the COSG model and then as a corollary we get the analogous results for the POSG model. To this end, we make the following assumption.

(A4) The stochastic kernel $\bar{Q}(\cdot|\psi, a, b, y)$, as in (7), is weakly continuous in (ψ, a, b, y).

Remark 3.1. If the observation space Y is countable, then one can verify that (A2) implies (A4). For uncountable Y, we give a typical example where (A4) is indeed true. Consider the POSG model whose state and observation processes evolve as follows:

$$\begin{aligned} X_{t+1} &= f(X_t, A_t, B_t, \xi_t), \qquad && t = 0, 1, \ldots, \\ Y_{t+1} &= g(A_t, B_t, X_{t+1}, \eta_{t+1}), \qquad && t = 0, 1, \ldots, \\ Y_0 &= g_0(X_0, \eta_0), \end{aligned}$$

where

$$f: \mathbb{R}^l \times \mathbb{R}^m \times R^n \times \mathbb{R}^p \to R^l, \qquad g: \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^q \to R^s, \qquad g_0: \mathbb{R}^l \times \mathbb{R}^q \to \mathbb{R}^s.$$

Here, $\{\xi_t\}$ is an iid sequence of random variables which is the state disturbance noise and $\{\eta_t\}$ is another sequence of iid random variables which is the observation noise. If we assume that the functions f, g, g_0 are continuous, then assumptions (A2) and (A3) are satisfied. Further, if we assume that the distributions of ξ_t and η_t have densities with respect to the Lebesgue measures on $\mathbb{R}^p$ and $\mathbb{R}^q$ respectively, then one can verify that (A4) is satisfied.

To establish the existence of a value and optimal strategies, we study the following Shapley equations:

$$
\begin{aligned}
W(y,\chi) = &\min_{\mu\in\mathscr{P}(B)} \max_{\lambda\in\mathscr{P}(A)} \bigg[\int_B \int_A \tilde{r}(y,\chi,a,b)\lambda(da)\mu(db) \\
&+ \beta \int_{Y\times\mathscr{P}(X)} \int_B \int_A W(y',\chi')\lambda(da)\mu(db)\tilde{P}(d(y',\chi')|y,\chi,a,b)\bigg] \\
= &\max_{\lambda\in\mathscr{P}(A)} \min_{\mu\in\mathscr{P}(B)} \bigg[\int_A \int_B \tilde{r}(y,\chi,a,b)\mu(db)\lambda(da) \\
&+ \beta \int_{Y\times\mathscr{P}(X)} \int_A \int_B W(y',\psi')\mu(db)\lambda(da)\tilde{P}(d(y',\chi')|y,\chi,a,b)\bigg].
\end{aligned}
\tag{44}
$$

Let $B(Y\times\mathscr{P}(X))$ be the space of bounded and measurable functions on $Y\times\mathscr{P}(X)$ endowed with the sup-norm topology. We can use the methods in Refs. 13–15 to obtain the following result. We omit the proof.

Theorem 3.1. Assume (A1)–(A4). Then (44) has a unique solution W in $B(Y\times\mathscr{P}(X))$. The value in the COSG model exists and is given by

$$\tilde{V}(\psi) = \int_{Y\times\mathscr{P}(X)} W(y,\chi)\Phi(\psi)(dy,d\chi). \tag{45}$$

Further, let $\lambda^*(\cdot)\ \tilde{\Pi}_S^1$ be such that

$$
\begin{aligned}
&\min_{\mu\in\mathscr{P}(B)} \bigg[\int_A \int_B \tilde{r}(y,\chi,a,b)\mu(db)\lambda^*(y,\chi)(da) \\
&\quad+ \beta \int_{Y\times\mathscr{P}(X)} \int_A \int_B W(y',\chi')\mu(db)\lambda^*(y,\chi)(da)\tilde{P}(d(y',\chi')|y,\chi,a,b)\bigg] \\
&= \max_{\lambda\in\mathscr{P}(A)} \min_{\mu\in\mathscr{P}(B)} \bigg[\int_A \int_B \tilde{r}(y,\chi,a,b)\mu(db)\lambda(da) \\
&\quad+ \beta \int_{Y\times\mathscr{P}(X)} \int_A \int_B W(y',\chi')\mu(db)\lambda(da)\tilde{P}(d(y',\chi')|y,\chi,a,b)\bigg],
\end{aligned}
\tag{46}
$$

for every $(y,\chi)\in Y\times\mathscr{P}(X)$. Then, $\lambda^*(\cdot)$ is a stationary optimal strategy for player 1 in the COSG model.

Similarly, let $\mu^*(\cdot)\in\tilde{\Pi}_S^2$ be such that

$$\begin{aligned}
&\max_{\lambda\in\mathscr{P}(A)}\Bigg[\int_B\int_A \tilde{r}(y,\chi,a,b)\lambda(da)\mu^*(y,\chi)(da)\\
&\quad+\beta\int_{Y\times\mathscr{P}(X)}\int_B\int_A W(y',\chi')\lambda(da)\mu^*(y,\chi)(db)\tilde{P}(d(y',\chi')|y,\chi,a,b)\Bigg]\\
&=\min_{\mu\in\mathscr{P}(B)}\max_{\lambda\in\mathscr{P}(A)}\Bigg[\int_B\int_A \tilde{r}(y,\chi,a,b)\lambda(da)\mu(db)\\
&\quad+\beta\int_{Y\times\mathscr{P}(X)}\int_B\int_A W(y',\chi')\lambda(da)\mu(db)\tilde{P}(d(y',\chi')|y,\chi,a,b)\Bigg],
\end{aligned}\tag{47}$$

for every $(y,\chi)\in Y\times\mathscr{P}(X)$. Then, $\mu^*(\cdot)$ is a stationary optimal strategy for player 2 in the COSG model for any initial distribution.

Corollary 3.1. Assume (A1)–(A4). Let $\lambda^*(\cdot)$, $\mu^*(\cdot)$ be as in (46), (47) respectively. Let g^* be the map as defined in (22). Let $(\pi^{*1},\pi^{*2})=(g^*(\lambda^*(\cdot)), g^*(\mu^*(\cdot)))$. Then, (π^{*1},π^{*2}) is a pair of optimal strategies in the POSG model.

Proof. This follows from Theorem 2.1 and Theorem 3.1. □

References

1. Shapley, L., *Stochastic Games*, Proceedings of National Academy of Sciences, Vol. 39, pp. 1095–1100, 1953.
2. Vrieze, K., *Zero-Sum Stochastic Games: A Survey*, CWI Quarterly, Vol. 2, pp. 147–170, 1989.
3. Bertsekas, D. P., and Shreve, S. E., *Stochastic Optimal Control*, Academic Press, New York, NY, 1978.
4. Dynkin, E. B., and Yushkevich, A. A., *Controlled Markov Processes*, Springer Verlag, Berlin, Germany, 1979.
5. Hernandez-Lerma, O., *Adaptive Control of Markov Processes*, Springer Verlag, New York, NY, 1989.
6. Borkar, V. S., *N-Person Noncooperative Stochastic Games with Partial Information*, Game Theory and Economic Applications, Lecture Notes in Economics and Mathematical Systems, Springer Verlag, New York, NY, Vol. 389, pp. 98–113, 1992.
7. Rosenberg, D., and Vielle, N., *The Maxmin of Recursive Games with Incomplete Information on One Side*, Mathematics of Operations Research, Vol. 25, pp. 23–35, 2000.

8. Sorin, S., *Big Match with Lack of Information on One Side, Part 1*, International Journal of Game Theory, Vol. 13, pp. 201–255, 1984.
9. Sorin, S., *Big Match with Lack of Information on One Side, Part 2*, International Journal of Game Theory, Vol. 14, pp. 173–204, 1985.
10. Sorin, S., and Zamir, S., *A 2-Person Game with Lack of Information on $1\frac{1}{2}$ Sides*, Mathematics of Operations Research, Vol. 10, pp. 17–23, 1985.
11. Zamir, S., *Repeated Games of Incomplete Information: Zero-Sum*, Handbook of Game Theory with Economic Applications, Elsevier Science Publishers, Amsterdam, Holland, Vol. 1, pp. 109–154, 1992.
12. Borkar, V. S., and Ghosh, M. K., *Stochastic Differential Games: An Occupation Measure Based Approach*, Journal of Optimization Theory and Applications, Vol. 73, pp. 359–385, 1992 (Errata Corrige, Vol. 88, pp. 251–252, 1996).
13. Kumar, P. R., and Shiau, T. H., *Existence of Value and Randomized Strategies in Zero-Sum Discrete-Time Stochastic Dynamic Games*, SIAM Journal on Control and Optimization, Vol. 19, pp. 617–634, 1981.
14. Maitra, A., and Parthasarathy, T., *On Stochastic Games*, Journal of Optimization Theory and Applications, Vol. 5, pp. 289–300, 1970.
15. Nowak, A. S., *On Zero-Sum Stochastic Games with General State Space, II*, Probability and Mathematical Statistics, Vol. 4, 143–152, 1984.