On Strategic Measures and Optimality Properties in Discrete-Time Stochastic Control with Universally Measurable Policies

Huizhen Yu†

Dedicated to the memory of Professor Sanjoy K. Mitter, an inspiring mentor.

Abstract

This paper concerns discrete-time infinite-horizon stochastic control systems with Borel state and action spaces and universally measurable policies. We study optimization problems on strategic measures induced by the policies in these systems. The results are then applied to risk-neutral and risk-sensitive Markov decision processes, as well as their partially observable counterparts, to establish the measurability of the optimal value functions and the existence of universally measurable, randomized or nonrandomized, $\epsilon$-optimal policies, for a variety of average cost criteria and risk criteria. We also extend our analysis to a class of minimax control problems and establish similar optimality results under the axiom of analytic determinacy.

Keywords:
Markov decision processes; risk-sensitive control; minimax control;
Borel spaces; universally measurable policies; strategic measures

* A shorter version of this paper is to appear in the journal Mathematics of Operations Research. This research was supported by grants from DeepMind, Alberta Machine Intelligence Institute (AMII), and Alberta Innovates—Technology Futures (AITF).

† RLAI Lab, Department of Computing Science, University of Alberta, Canada (janey.hzyu@gmail.com)
# Strategic Measures in Stochastic Control on Borel Spaces

## Contents

| Section                                      | Pages |
|----------------------------------------------|-------|
| 1 Introduction                              | 3     |
| 2 Borel-Space MDPs                           | 4     |
| 2.1 Preliminaries                           | 4     |
| 2.2 MDPs with Universally Measurable Policies | 6     |
| 3 Properties of Strategic Measures           | 6     |
| 4 Applications and Extensions                | 10    |
| 4.1 MDPs with Risk-Neutral and Risk-Sensitive Average Criteria | 10    |
| 4.1.1 The Risk-Neutral Case                 | 10    |
| 4.1.2 The Risk-Sensitive Case               | 12    |
| 4.1.3 Partially Observable Problems         | 14    |
| 4.2 Minimax Control                          | 15    |
| 4.2.1 Model and Assumptions                 | 15    |
| 4.2.2 Properties of Strategic Measures and Optimality Results | 17    |
| 4.2.3 Some Implications on Optimality Equations/Inequalities | 20    |
| 5 Proofs                                     | 23    |
| 5.1 Some Definitions and Useful Lemmas      | 23    |
| 5.2 Proofs for Section 3                     | 25    |
| 5.2.1 Proof of Theorem 3.1                  | 25    |
| 5.2.2 Proof of Proposition 3.1              | 27    |
| 5.2.3 Proof of Theorem 3.2                  | 29    |
| 5.3 Proofs for Section 4                     | 31    |
| 5.3.1 Proofs of Proposition 4.1 and Theorem 4.3 | 31    |
| 5.3.2 Proofs of Propositions 4.2-4.3 and 4.5 | 31    |
| 5.3.3 Projective Class $\Sigma_2^2$ and Proof of Proposition 4.4 | 33    |
| References                                   | 35    |
1 Introduction

We consider discrete-time infinite-horizon stochastic control in the universal measurability framework as formulated by Shreve and Bertsekas [47, 48] (see also [5, Part II]), where the state and action spaces are Borel, the control constraints have analytic graphs, and the policies are universally measurable. Our particular focus will be on Markov decision processes (MDPs), their partially observable counterparts in which the controller has imperfect information about the states, and minimax control problems in which the controller plays against an opponent. The main results of this paper concern optimality properties of these systems, such as the measurability of the optimal value functions and the existence of measurable, randomized or nonrandomized, optimal or nearly optimal policies, under general conditions on the system dynamics and with respect to (w.r.t.) a broad range of performance criteria.

For MDPs with Borel measurable policies, strategic measures were analyzed and optimality results similar to ours were established a long time ago in Strauch [49], Blackwell [8], Dynkin and Yushkevich [16, Chaps. 3, 5], and Feinberg [18, 19, 20, 21]. (Although this is outside our scope, we mention that topological properties of strategic measures have also been studied by Schäl [45] and Balder [3], and by Yüksel and Saldi recently for Borel team-decision models [57].) For MDPs with universally measurable policies, however, an analysis of strategic measures seems to be missing in the literature. One of the purposes of this paper is to fill this gap. While the two MDP models are similar, their differences are not superficial, so our results are not mere repetitions of the established results for Borel measurable policies. In our proofs, besides arguments from the prior works, we use also the properties of analytic sets and semianalytic or universally measurable functions. Another important difference is in the notion of optimality. The prior results in [16, Chaps. 3, 5] and [18, 19, 49] concern the existence of Borel measurable, randomized or nonrandomized (\(\epsilon\)-)optimal policies for almost all states, w.r.t. any given finite measure on the state space. By contrast, with universally measurable policies, we can show that there exist policies that are (\(\epsilon\)-)optimal everywhere.

Our work also builds on the early work of Shreve and Bertsekas [47, 48] (see also [5, Chap. 9]) on MDPs with universally measurable policies. They focused on the discounted and total cost criteria. Instead of strategic measures, they considered the marginal distributions of the state and action pairs at each stage and related the optimal control problems to optimization problems on those sequences of marginal distributions that can be induced by the policies through time. This method has also been used by the author [53, 54] for certain total and average cost problems. But it cannot be applied when the performance criteria are not functions of those marginal state-action distributions. Such criteria include, for example, many popular risk criteria and certain long-run average cost criteria that involve costs along sample paths. To address problems with such performance criteria in the universal measurability framework is the main purpose of this paper.

For minimax control, we consider a class of problems in which the controller does not observe the opponent’s actions, and we introduce an absolute continuity condition on the finite-dimensional distributions of the processes observable to the controller while the controller applies the same policy (see Assumption 4.1). This allows us to reformulate the minimax control problems as minimax optimization problems on strategic measures. We then prove an optimality result regarding the optimal value function and the (\(\epsilon\)-)optimal policies for the controller (see Theorem 4.4) by using the properties of the set of strategic measures and a uniformization theorem of Kondō from descriptive set theory. In this result, besides the absolute continuity condition mentioned earlier and a measurability condition on the performance criterion, we assume the axiom of analytic determinacy. (The standard ZFC axioms, i.e., Zermelo–Fraenkel set theory with the axiom of choice, are assumed throughout the paper without explicit reference.) The purpose of this additional axiomatic assumption is to resolve the measurability questions that cannot be settled in the ZFC system. Previously, such an approach and a closely related axiomatic assumption (the universal measurability of \(\Sigma^1_2\) sets) had been discussed and used in Maitra, Purves, and Sudderth [36] for Borel leavable gambling problems, and in Prikry and Sudderth [43] for one-stage Borel parametrized games. Further discussions about
analytic determinacy (and $\Sigma^0_3$ sets) will be given in Remarks 2.1, 4.4 and Section 5.3.3.

We are not aware of any prior result similar to ours in the minimax control/stochastic game literature. The majority of published results (for discounted, total, or average cost criteria) concern the validity of dynamic programming equations, the existence of stationary optimal policies, and/or minimax equalities, under various continuity/compactness conditions on the control models (see e.g., [26, 31, 32, 44] and the references therein). Nowak [40] and Maitra and Sudderth [37] placed such conditions on only one of the players. Maitra and Sudderth [38] did not require such conditions; to avoid measurability issues, they took the approach that is based on the theory of finitely additive (instead of countably additive) probability measures.

We mention that this paper does not aim to establish dynamic programming equations for the optimal value functions, because the strategic-measure approach alone is generally insufficient for that purpose (especially for average-cost and risk-sensitive problems). However, the optimality results it produces can serve as starting point for deriving such equations. We will demonstrate this for minimax control with risk-neutral criteria (see Section 4.2.3). Some other related examples can be found in our earlier analyses of discounted, total-cost, and average-cost MDPs with universally measurable policies ([54, Thm. 2.3], [53, App. A], and [55, Sec. 2], respectively), and in Yüksel [56] for finite-horizon decentralized team-decision problems with Borel measurable policies.

This paper is organized as follows. In Section 2 we give background materials about Borel-space MDPs. In Section 3 we present general results regarding the measurability of various sets of strategic measures and the optimality properties of associated optimization problems. We then apply these results to risk-neutral and risk-sensitive MDPs in Section 4, where we also extend these results to partially observable problems and minimax control. We collect some of our proofs in Section 5.

## 2 Borel-Space MDPs

In this section we describe the universal measurability framework for Borel-space MDPs. We start with the definitions and important properties of certain sets/functions that underly this framework.

### 2.1 Preliminaries

A Borel subset of a Polish space is called a **Borel space** (or **standard Borel space**) [5, Def. 7.7]. For a Borel space $X$, $\mathcal{B}(X)$ denotes the Borel $\sigma$-algebra on $X$, and $\mathcal{P}(X)$ denotes the set of all probability measures on $\mathcal{B}(X)$. We endow the space $\mathcal{P}(X)$ with the topology of weak convergence; then $\mathcal{P}(X)$ is also a Borel space [5, Sec. 7.4.2]. The *universal $\sigma$-algebra* on $X$ is given by $\mathcal{U}(X) := \cap_{p \in \mathcal{P}(X)} \mathcal{B}_p(X)$. Here, for each $p \in \mathcal{P}(X)$, $\mathcal{B}_p(x)$ is the $\sigma$-algebra generated by $\mathcal{B}(X)$ and all the subsets of $X$ with $p$-outer measure 0, and on $\mathcal{B}_p(X)$, $p$ admits a unique extension, called the *completion of $p$* (cf. [15, Sec. 3.3]). The sets in $\mathcal{U}(X)$ and the $\mathcal{U}(X)$-measurable mappings on $X$, being measurable w.r.t. the completion of any $p \in \mathcal{P}(X)$, are called *universally measurable*.

Let $X$ and $Y$ be Borel spaces. A function $q(\cdot \mid x) : \mathcal{B}(Y) \times X \to [0,1]$ is called a *universally measurable stochastic kernel* (resp. *Borel measurable stochastic kernel*) on $Y$ given $X$, if for each $x \in X$, $q(\cdot \mid x)$ is a probability measure on $\mathcal{B}(Y)$ and for each $B \in \mathcal{B}(Y)$, $q(B \mid \cdot)$ is universally measurable (resp. Borel measurable). Equivalently, such a kernel can also be defined as a measurable mapping $x \mapsto q(\cdot \mid x)$ from the space $(X, \mathcal{U}(X))$ (resp. $(X, \mathcal{B}(X))$) into the space $\left(\mathcal{P}(Y), \mathcal{B}(\mathcal{P}(Y))\right)$ of probability measures; cf. [5, Def. 7.12, Prop. 7.26, Lem. 7.28]. To refer to the stochastic kernel, we will often use the notation $q$ or $q(dy \mid x)$.

For $p \in \mathcal{P}(X)$ and $D \in \mathcal{U}(X)$, $p(D)$ is the probability of $D$ w.r.t. the completion of $p$. (Notation-wise, we do not distinguish between $p$ and its completion.) Let $\mathbb{R} := (-\infty, +\infty)$ and $\tilde{\mathbb{R}} := [-\infty, +\infty]$. If $f : X \to \tilde{\mathbb{R}}$ is universally measurable, then $\int f dp := \int f^+ dp - \int f^- dp$, where $f^+$ (respectively $f^-$) is the positive (negative) part of $f$ and the integration is w.r.t. the completion of $p$. (For summations involving extended real numbers, the convention $\infty - \infty = -\infty + \infty = \infty$ is adopted unless otherwise stated.)
The same interpretations apply to \( q(D \mid y) \) and \( \int f(x)q(dx \mid y) \), if \( D \) and \( f \) are as defined above and \( q(dx \mid y) \) is a Borel or universally measurable stochastic kernel on \( X \) given \( Y \).

A subset of \( X \) is called analytic, if it is either empty or the image of a Borel subset of some Polish space under a Borel measurable mapping (cf. [5, Prop. 7.41], [15, Sec. 13.2]). The complement of an analytic set is called coanalytic. A lower semianalytic function is a function \( f : D \to \mathbb{R} \) such that the domain \( D \) is an analytic set and for every \( r \in \mathbb{R} \), the level set \( \{ x \in D \mid f(x) \leq r \} \) is analytic [5, Def. 7.21, Lem. 7.30(1)]. (This definition is equivalent to that the epigraph of \( f \), \( \{ (x, r) \in D \times \mathbb{R} \mid f(x) \leq r \} \), is analytic; cf. [5, p. 186].) Replacing \( \leq r \) with \( \geq r \) in the preceding statement gives the definition of an upper semianalytic function. Borel subsets of \( X \) are analytic (and coanalytic); Borel measurable functions from \( X \) into \( \mathbb{R} \) are lower (and upper) semianalytic. The \( \sigma \)-algebra \( \mathcal{A}(X) \) generated by analytic subsets of \( X \) contains \( \mathcal{B}(X) \) and is contained in \( \mathcal{U}(X) \). Thus analytically measurable (i.e., \( \mathcal{A}(X) \)-measurable) sets and functions, in particular, analytic subsets of \( X \) and lower (or upper) semianalytic functions on \( X \), are universally measurable.

In addition, analytic sets and lower semianalytic functions have a number of important properties that play fundamental roles in the mathematical framework for Borel-space MDPs. Let us recount a few of these properties here, which will be used frequently in this paper (see the article [9] for a more comprehensive review, and the monograph [5, Chap. 7] for an in-depth study).

(a) For any analytic set \( D \subset X \) and real number \( r \geq 0 \), \( \{ p \in \mathcal{P}(X) \mid p(D) > r \} \) and \( \{ p \in \mathcal{P}(X) \mid p(D) \geq r \} \) are analytic subsets of \( \mathcal{P}(X) \) ([5, Prop. 7.43, Cor. 7.43.1]; see also [9, Lem. (25)]).

(b) If \( f : X \times Y \to \mathbb{R} \) is lower semianalytic and \( q(dy \mid x) \) is a Borel measurable stochastic kernel on \( Y \) given \( X \), then \( \phi(x) := \int f(x, y)q(dy \mid x) \) is lower semianalytic on \( X \) [5, Prop. 7.48].

For \( D \subset X \times Y \), \( \text{proj}_X(D) := \{ x \in X \mid (x, y) \in D \text{ for some } y \in Y \} \) is the projection of \( D \) on \( X \), and we call a function \( \psi : \text{proj}_X(D) \to Y \) a selection of \( D \) if the graph of \( \psi \) is contained in \( D \), i.e., \((x, \psi(x)) \in D \) for all \( x \in \text{proj}_X(D) \).

(c) If \( D \subset X \times Y \) is analytic, then \( \text{proj}_X(D) \) is analytic [5, Prop. 7.39] and \( D \) admits an analytically measurable selection (the Jankov-von Neumann selection theorem [5, Prop. 7.49]).

For partial minimization of a function \( f \) on \( X \times Y \) (that is, minimizing \( f \) over \( Y \) for each \( x \in X \)), by applying (c) to the level sets or epigraph of a lower semianalytic \( f \) and using the properties of universally measurable functions, one obtains the following:

(d) If \( D \subset X \times Y \) is analytic and \( f : D \to \mathbb{R} \) is lower semianalytic, then the function \( f^* : \text{proj}_X(D) \to \mathbb{R} \) defined by

\[
   f^*(x) = \inf_{y \in D_x} f(x, y), \quad \text{where } D_x = \{ y \in Y \mid (x, y) \in D \},
\]

is also lower semianalytic [5, Prop. 7.47]. Furthermore, for each \( \epsilon > 0 \), there exists a universally measurable function \( \psi : \text{proj}_X(D) \to Y \) such that \( \psi(x) \in D_x \) for all \( x \in \text{proj}_X(D) \) and

\[
   f(x, \psi(x)) = f^*(x), \quad \forall x \in E^*,
\]

\[
   f(x, \psi(x)) \leq \begin{cases} f^*(x) + \epsilon & \text{if } f^*(x) > -\infty, \\ -1/\epsilon & \text{if } f^*(x) = -\infty, \end{cases} \quad \forall x \in \text{proj}_X(D) \setminus E^*,
\]

where \( E^* := \{ x \in \text{proj}_X(D) \mid \arg \min_{y \in D_x} f(x, y) \neq \emptyset \} \subseteq \mathcal{U}(X) \) (cf. [5, Prop. 7.50]).

Remark 2.1. We will introduce later a counterpart of (d) for upper semianalytic functions as well as for a more general class of functions under the axioms of ZFC and analytic determinacy (see Prop. 4.4 and Remark 5.1). Based on Gödel’s work [25], it is known that this is not possible under ZFC alone (cf. the discussions in [5, p. 302] and Remark 4.4(b)).

Miscellaneous notation: Throughout the paper, for \( x \in X \), \( \delta_x \) denotes the Dirac measure concentrating at \( x \), and \( \mathcal{D}(X) \) denotes the set of all Dirac measures on \( X \). For \( B \subset X \), \( B^c := X \setminus B \) and \( 1_B \) is the indicator function for \( B \). For an event \( E \) in a probability space with a probability measure \( p \), we write \( \bar{p}(E) \) for the indicator of \( E \) and \( p(E) \) for the probability of \( E \).
2.2 MDPs with Universally Measurable Policies

We consider a Borel-space MDP in the universal measurability framework (cf. [5, Sec. 8.1]). Specifically, the following will be assumed throughout the paper unless otherwise stated:

- The state space \( \mathcal{X} \) and the action space \( \mathcal{A} \) are \textit{Borel spaces}.
- State transitions are governed by \( q(dy \mid x, a) \), a \textit{Borel measurable} stochastic kernel on \( \mathcal{X} \) given \( \mathcal{X} \times \mathcal{A} \).
- The control constraint is specified by a set-valued mapping \( A : x \mapsto A(x) \) on \( \mathcal{X} \), where \( A(x) \subset \mathcal{A} \) is a nonempty set of admissible actions at the state \( x \). The graph of \( A(\cdot) \), \( \Gamma := \{(x, a) \mid x \in \mathcal{X}, a \in A(x)\} \), is \textit{analytic}.
- The one-stage cost function \( c : \Gamma \rightarrow \bar{\mathbb{R}} \) is \textit{lower semianalytic}.

Our primary interest is in infinite-horizon control problems (several performance criteria will be introduced in Section 4). Let \( \omega := (x_0, a_0, x_1, a_1, \ldots) \), where \( x_n \) and \( a_n \) denote the state and action, respectively, at the \( n \)th stage. For \( n \geq 0 \), let \( h_n := (x_0, a_0, x_1, a_1, \ldots, x_n) \) and \( h'_n := (x_0, a_0, x_1, a_1, \ldots, x_n, a_n) \). We denote the space of \( \omega \) by \( \Omega := (\mathcal{X} \times \mathcal{A})^\infty \), the space of \( h_n \) by \( H_n := (\mathcal{X} \times \mathcal{A})^n \times \mathcal{X} \), and the space of \( h'_n \) by \( H'_n := (\mathcal{X} \times \mathcal{A})^{n+1} \). All these spaces are endowed with the product topology to make them Borel spaces [5, Prop. 7.13].

A \textit{universally measurable policy} (or \textit{a policy} for short) is a sequence of \textit{universally measurable} stochastic kernels, \( \pi := (\mu_0, \mu_1, \ldots) \), where for each \( n \geq 0 \), \( \mu_n(da_n \mid h_n) \) is a universally measurable stochastic kernel on \( \mathcal{A} \) given \( H_n \) that satisfies the control constraint

\[
\mu_n(A(x_n) \mid h_n) = 1, \quad \forall h_n = (x_0, a_0, \ldots, a_{n-1}, x_n) \in H_n. \tag{2.4}
\]

Note that the set \( A(y), y \in \mathcal{X} \), is analytic since it is a section of the analytic set \( \Gamma \) (cf. [5, Prop. 7.40]), so the probability of \( A(x_n) \) in the above expression is well-defined. A \textit{Borel measurable policy} is a policy that consists of Borel measurable stochastic kernels \( \{\mu_n\}_{n \geq 0} \). If for every \( n \geq 0 \) and every \( h_n \in H_n \), \( \mu_n(da_n \mid h_n) \) is a Dirac measure, \( \pi \) is called a \textit{nonrandomized} policy. A \textit{Markov} (resp. \textit{semi-Markov}) policy is a policy such that for every \( n \geq 0 \), as a function of \( h_n \), the probability measure \( \mu_n(da_n \mid h_n) \) depends only on \( x_n \) (resp. \( (x_0, x_n) \)). If a Markov policy \( \pi \) has identical stochastic kernels, i.e., \( \pi = (\mu, \mu, \ldots) \), we call it a \textit{stationary} policy and write it simply as \( \mu \). Likewise, if a semi-Markov policy \( \pi \) satisfies that for some stochastic kernel \( \tilde{\mu} \) on \( \mathcal{X} \) given \( \mathcal{X}^2 \), \( \mu_0(da_0 \mid x_0) = \tilde{\mu}(da_0 \mid x_0, x_0) \) and \( \mu_n(da_n \mid x_0, x_n) = \tilde{\mu}(da_n \mid x_0, x_n) \) for all \( x_0, x_n \in \mathcal{X} \) and \( n \geq 1 \), we call \( \pi \) a \textit{semi-stationary} policy and write it simply as \( \tilde{\mu} \).

Let \( \Pi \) denote the set of all policies, and let \( \Pi_m, \Pi_{m}, \Pi_{ss} \), and \( \Pi_s \) denote the subsets of semi-Markov, Markov, semi-stationary, and stationary policies, respectively. If \( \Pi_s \) is any one of these sets, we write \( \Pi_s' \) for the subset of nonrandomized policies in \( \Pi_s \). All these sets are nonempty, because by the Jankov-von Neumann selection theorem [5, Prop. 7.49], the analytic set \( \Gamma \) admits an analytically measurable selection, which corresponds to a universally measurable nonrandomized stationary policy. By contrast, a Borel measurable policy may not exist even if \( \Gamma \) is Borel [7].

For \( \pi := \{\mu_n\}_{n \geq 0} \in \Pi \), an initial state distribution \( p_0 \in \mathcal{P}(\mathcal{X}) \) together with the collection of stochastic kernels \( \mu_0(da_0 \mid x_0), q(dx_1 \mid x_0, a_0), \mu_1(da_1 \mid h_1), q(dx_2 \mid x_1, a_1), \ldots \) determines uniquely a probability measure \( P^\pi_{p_0} \) on \( \mathcal{U}(\Omega) \) [5, Prop. 7.45]. The expectation w.r.t. \( P^\pi_{p_0} \) is denoted by \( E^\pi_{p_0} \). If \( p_0 = \delta_x \), we will also use the notation \( P^\pi_x \) and \( E^\pi_x \). Throughout the paper, we will simply write a stochastic process on \( \Omega \) as \( \{(x_n, a_n)\}_{n \geq 0} \), using \( (x_n, a_n) \) to denote the random variables \( (x_n(\omega), a_n(\omega)) \) instead, which are the \((x_n, a_n)\)-components of \( \omega \).

## 3 Properties of Strategic Measures

The probability measures on \( \Omega \) induced by the policies are also known as strategic measures in the literature [16, Sec. 3.5]. In the universal measurability framework, these are probability measures on \( \mathcal{U}(\Omega) \); however, we will work with their restrictions to \( \mathcal{B}(\Omega) \), in order to make use of well-studied
properties of Borel probability measures. Since a probability measure on \( \mathcal{U} (\Omega) \) is uniquely determined by its restriction to \( \mathcal{B} (\Omega) \) (cf. [15, Sec. 3.3] on completion of measures), the two are effectively the same. Specifically, for a policy \( \pi \in \Pi \) and an initial distribution \( p_0 \in \mathcal{P} (X) \), let \( \rho_{p_0}[\pi] \) denote the restriction of \( \mathbb{F}^\pi_{p_0} \) to \( \mathcal{B} (\Omega) \); if \( p_0 = \delta_x \), we also write it as \( \rho_x[\pi] \). Let
\[ \mathcal{S} := \{ p \in \mathcal{P} (\Omega) \mid p = \rho_{p_0}[\pi], \pi \in \Pi, p_0 \in \mathcal{P} (X) \} \]
be the set of all strategic measures, and let \( \mathcal{S}' \) be the subset of strategic measures corresponding to nonrandomized policies in \( \Pi' \). Similarly, with \( \Pi_m \) (resp. \( \Pi_n \)) in place of \( \Pi \), we define \( \mathcal{S}_m \) (resp. \( \mathcal{S}_n \)) to be the set of all strategic measures induced by Markov (resp. stationary) policies, and let \( \mathcal{S}'_m \) (resp. \( \mathcal{S}'_n \)) denote its subset corresponding to those nonrandomized policies.

**Theorem 3.1.** For \( \mathcal{S}_* \in \{ \mathcal{S}, \mathcal{S}_m, \mathcal{S}_n \} \), \( \mathcal{S}_* \) and \( \mathcal{S}'_* \) are analytic; they are Borel if \( \Gamma \) is Borel.

We prove this theorem in Section 5.2.1. There we also consider the strategic measures induced by policies with certain structural constraints and give a more general result (cf. Prop. 5.1) that entails the conclusions of the above theorem for \( \mathcal{S} \) and \( \mathcal{S}_m \).

**Remark 3.1.** We make a few comments on the preceding theorem and related earlier results.

(a) For the MDP model with a Borel \( \Gamma \) and Borel measurable policies, the Borel measurability of \( \mathcal{S} \) was shown in Strauch [49, Lem. 7.2] and in Dynkin and Yushkevich [16, Secs. 3.5, 3.6, 5.5]; the Borel measurability of \( \mathcal{S}' \) was implied by the results of Blackwell [8]; and the Borel measurability of the other sets involved in Theorem 3.1 were established by Feinberg [21]. All these results were summarized in Feinberg [21, Thm. 3.2]. A small difference between [21, Thm. 3.2] and the part of Theorem 3.1 for the case of a Borel \( \Gamma \) is worth noting: in our case \( \Gamma \) need not admit a Borel measurable selection and hence a Borel measurable policy need not exist. (By the Blackwell–Ryll-Nardzewski selection theorem [10, Thm. 2], if \( \Gamma \) is Borel, then \( \Gamma \) admits a Borel measurable selection if and only if there exists a Borel measurable stationary policy.)

(b) Suppose \( \Gamma \) admits a Borel measurable selection, so that there is at least one Borel measurable policy \( \mu^0 \in \Pi'_m \). Then the sets \( \mathcal{S}_*, \mathcal{S}'_* \) in Theorem 3.1 all coincide with their respective subsets of strategic measures induced by Borel measurable policies. This follows from the fact that w.r.t. any given finite measure on its domain space, a universally measurable function coincides with some Borel measurable function almost everywhere [5, Lems. 7.27, 7.28(c)]. More specifically, using this fact and the policy \( \mu^0 \), for any initial distribution \( p_0 \) and policy \( \pi := \{ \mu_n \}_{n \geq 0} \in \Pi \), one can modify the stochastic kernels \( \mu_n \), one by one to construct a Borel measurable policy \( \tilde{\pi} \) such that \( \rho_{p_0}[\tilde{\pi}] = \rho_{p_0}[\pi] \) and \( \tilde{\pi} \) and \( \pi \) belong to the same class of policies (any class \( \Pi_* \) or \( \Pi'_* \) defined in Section 2.2). However, for optimal control, universally measurable policies are still essential in this case; cf. Remark 3.3(b).

(c) Let \( \Delta \subset \mathcal{P} (X) \times (\mathcal{P}(X \times A))^\infty \) consist of all those sequences \( \{ p_0, \gamma_0, \gamma_1, \ldots \} \) such that \( p_0 \in \mathcal{P} (X) \) and for some \( \pi \in \Pi_m, \gamma_n \in \mathcal{P} (X \times A) \) is the marginal distribution of \( (x_n, a_n) \) w.r.t. \( \mathbb{F}^\pi_{p_0} \) for all \( n \geq 0 \). It has been shown by Shreve and Bertsekas [48, Lem. 1] (see also [5, Lem. 9.1]) that the set \( \Delta \) is analytic; their proof also shows that \( \Delta \) is Borel if \( \Gamma \) is Borel. As to the relation between \( \Delta \) and strategic measures, we can show that \( \mathcal{S}_m \) is Borel-isomorphic to \( \Delta \), whereas \( \mathcal{S}'_m, \mathcal{S}_n, \) and \( \mathcal{S}'_n \) are Borel-isomorphic to certain subsets of \( \Delta \).

Many optimal control problems can be equivalently formulated as optimization problems on strategic measures. We now give some results that will be useful for constructing (nearly) optimal policies from (nearly) optimal solutions of those optimization problems.

Let \( \mathcal{S}_* \in \{ \mathcal{S}, \mathcal{S}_m, \mathcal{S}_n \} \). Define sets \( \tilde{\mathcal{S}}_*, \tilde{\mathcal{S}}_* \) by
\[ \tilde{\mathcal{S}}_* := \{ (x, p) \in X \times \mathcal{P} (\Omega) \mid p \in \mathcal{S}_*, p_0(p) = \delta_x \}, \quad \tilde{\mathcal{S}}'_* := \{ (x, p) \in X \times \mathcal{P} (\Omega) \mid p \in \mathcal{S}'_*, p_0(p) = \delta_x \}, \]
where \( p_0(p) \) denotes the marginal distribution of \( x_0 \) w.r.t. \( p \). For \( x \in X \), denote by \( \tilde{\mathcal{S}}_*(x) \) the \( x \)-section of \( \tilde{\mathcal{S}}_* \); i.e., \( \tilde{\mathcal{S}}_*(x) := \{ p \in \mathcal{P} (\Omega) \mid (x, p) \in \tilde{\mathcal{S}}_* \} = \{ p \in \mathcal{S}_* \mid p_0(p) = \delta_x \} \). Similarly, the \( x \)-section of \( \tilde{\mathcal{S}}'_* \) is denoted by \( \tilde{\mathcal{S}}'_*(x) \).
Lemma 3.1. The sets $\hat{\mathcal{S}}_*$ and $\hat{\mathcal{S}}'_*$ are analytic; they are Borel if $\Gamma$ is Borel.

Proof. The mapping $\psi_1 : p \mapsto (p_P(p), p)$ from $\mathcal{P}(\Omega)$ into $\mathcal{P}(X) \times \mathcal{P}(\Omega)$ and the mapping $\psi_2 : (x, p) \mapsto (\delta_x, p)$ from $X \times \mathcal{P}(\Omega)$ into $\mathcal{P}(X) \times \mathcal{P}(\Omega)$ are homeomorphisms. We can express the sets $\hat{\mathcal{S}}_*$ and $\hat{\mathcal{S}}'_*$ as

$$\hat{\mathcal{S}}_* = \psi_1^{-1}(\psi_2(\mathcal{S}_*))$$

and

$$\hat{\mathcal{S}}'_* = \psi_2^{-1}(\psi_1(\mathcal{S}_*)) .$$

By [5, Prop. 7.40], images and preimages of analytic sets are analytic under Borel measurable mappings; by [42, Sec. I.3, Cor. 3.3], under one-to-one Borel measurable mappings, images of Borel sets are Borel. Therefore, $\hat{\mathcal{S}}_*$ and $\hat{\mathcal{S}}'_*$ are analytic (Borel) if $\mathcal{S}_*$ and $\mathcal{S}_*$ are analytic (Borel). We then obtain the desired conclusion from Theorem 3.1. □

Proposition 3.1. Let $\mathcal{S}_* \in \{\mathcal{S}, \mathcal{S}_m, \mathcal{S}_s\}$. Let $\zeta : X \to \mathcal{P}(\Omega)$ be a universally measurable mapping such that $\zeta(x) \in \hat{\mathcal{S}}_*(x)$ for all $x \in X$. Then there exists a policy $\pi \in \Pi$ such that:

(i) $\rho x[\pi] = \zeta(x)$ for all $x \in X$;

(ii) for $\mathcal{S}_* = \mathcal{S}_m$ or $\mathcal{S}_s$, $\pi$ is semi-Markov or semi-stationary, respectively.

If, in addition, $\zeta$ is such that $\zeta(x) \in \hat{\mathcal{S}}'_*(x)$ for all $x \in X$, then $\pi$ can be taken to be nonrandomized.

The proof of this result is given in Section 5.2.2. A similar result for partially observable problems will be given later in Section 4.1.3 (see Prop. 4.1).

The next theorem is similar to Feinberg [19, Thm. 1] for MDPs with Borel measurable policies. It represents the strategic measures of randomized policies as “mixtures” of the strategic measures of nonrandomized policies, and will be important in studying the optimality of randomized policies.

We prove this theorem and a more general version of it in Section 5.2.3. The proof is similar to that of [19, Thm. 1] and makes use of a result in Gilkhman and Skorokhod [24, Lem. 1.2], in addition to the properties of universally measurable sets and functions.

Let $\lambda$ be the countable product of Lebesgue measures on $[0, 1]$.

Theorem 3.2. For each $\pi \in \Pi$, there exists a family of nonrandomized policies $\tilde{f}(\theta) \in \Pi'$, parametrized by $\theta \in [0, 1]^\infty$, such that

(i) the mapping $(p_0, \theta) \mapsto \rho_{p_0}[\tilde{f}(\theta)]$ from $\mathcal{P}(X) \times [0, 1]^\infty$ into $\mathcal{P}(\Omega)$ is universally measurable;

(ii) for all $p_0 \in \mathcal{P}(X)$,

$$\mathcal{P}^\pi_{p_0}(E) = \int_{[0, 1]^\infty} \mathcal{P}^{\tilde{f}(\theta)}(E) \lambda(d\theta), \quad \forall E \in \mathcal{U}(\Omega). \tag{3.1}$$

Moreover, if $\pi$ is Markov (semi-Markov), the policies $\tilde{f}(\theta)$ are Markov (semi-Markov) as well.

Remark 3.2. The representation (3.1) for $\mathcal{P}^\pi_{p_0}$ involves the probability measure $\tilde{\lambda}$ on $[0, 1]^\infty$. Alternatively, similarly to Feinberg [21, Thm. 5.2], we can introduce a probability measure $\eta$ on $\mathcal{U}(\mathcal{P}(\Omega))$ to express $\mathcal{P}^\pi_{p_0}$ as

$$\mathcal{P}^\pi_{p_0}(E) = \int_{\mathcal{P}(\Omega)} p(E) \eta(dp), \quad \forall E \in \mathcal{U}(\Omega), \tag{3.2}$$

where $\eta$ depends on both $p_0$ and $\pi$. In particular, let $\gamma_{p_0} : [0, 1]^\infty \to \mathcal{P}(\Omega)$ with $\gamma_{p_0}(\theta) := \rho_{p_0}[\tilde{f}(\theta)]$. By Theorem 3.2, $\gamma_{p_0}$ is universally measurable and we can let $\eta$ be the image measure of $\lambda$ under $\gamma_{p_0}$: $\eta = \lambda \circ \gamma_{p_0}^{-1}$. Moreover, by Theorem 3.2, $\eta(S') = 1$, and if $\pi$ is Markov, then $\eta(S'_m) = 1$. □

By combining the preceding results with [5, Props. 7.47, 7.50], we obtain the following theorem about optimization problems on strategic measures.

Theorem 3.3. Let $(\mathcal{S}_*, \Pi*) = (\mathcal{S}, \Pi), (\mathcal{S}_m, \Pi_{sm}),$ or $(\mathcal{S}_s, \Pi_{ss})$. Let $\phi : \mathcal{S}_* \to \mathcal{R}$ be lower semianalytic. Then the following hold:

(i) The function $\phi^*(x) := \inf_{\pi \in \mathcal{S}_*(x)} \phi(x, p)$, $x \in X$, is lower semianalytic.
The policy \( \pi \) of Prop. 3.1, we then obtain a nonrandomized policy

\[
\phi(x, \rho_x[\pi]) = \phi^*(x), \quad \forall x \in E^* := \left\{ x \in \mathcal{X} \mid \arg \min_{p \in \mathcal{S}} \phi(x, p) \neq \emptyset \right\};
\]

\[
\phi(x, \rho_x[\pi]) \leq \begin{cases} 
\phi^*(x) + \epsilon & \text{if } \phi^*(x) > -\infty, \\
-1/\epsilon & \text{if } \phi^*(x) = -\infty,
\end{cases} \quad \forall x \in \mathcal{X} \setminus E^*.
\]

(iii) The policy \( \pi^* \) in (ii) can be taken to be nonrandomized if the function \( \phi \) satisfies the condition

\[
\phi^*(x) = \inf_{p \in \mathcal{S}_x(x)} \phi(x, p) \quad \text{for all } x \in \mathcal{X}.
\]

In particular, for \( \mathcal{S}_* = \mathcal{S} \) or \( \mathcal{S}_m \), this condition holds if \( \phi \) is such that for each \( \pi \in \Pi \) or \( \Pi_m \), respectively,

\[
\phi(x, \rho_x[\pi]) \geq \int_{[0,1]^{\infty}} \phi(x, \rho_x[\bar{f}(\tilde{\theta})]) \lambda(d\tilde{\theta}), \quad \forall x \in \mathcal{X},
\]

where the function \( \bar{f}(\tilde{\theta}) \) associated with \( \pi \) is as given in Theorem 3.2.

Proof. Since the set \( \tilde{\mathcal{S}} \) is analytic by Lemma 3.1 and the function \( \phi \) is lower semianalytic by assumption, part (i) holds by [5, Prop. 7.47]. Moreover, by [5, Prop. 7.50(b)], there exists a universally measurable function \( \zeta^* : \mathcal{X} \to \mathcal{P}(\Omega) \) such that its graph lies in \( \tilde{\mathcal{S}} \) and the relations (3.3)-(3.4) hold with \( \zeta^*(x) \) in place of \( \rho_x[\pi^*] \). Letting \( \pi^* \in \Pi \), be the policy given by Prop. 3.1 for the function \( \zeta^* \), we obtain part (ii).

If \( \phi^*(x) = \inf_{p \in \mathcal{S}_x(x)} \phi(x, p) \) for all \( x \in \mathcal{X} \), then, since the set \( \tilde{\mathcal{S}}_x \) is analytic by Lemma 3.1, we can replace \( \tilde{\mathcal{S}}_x \) with \( \tilde{\mathcal{S}}_x' \) in the preceding proof so that the graph of \( \zeta^* \) lies in \( \tilde{\mathcal{S}}_x' \). From the last statement of Prop. 3.1, we then obtain a nonrandomized policy \( \pi^* \in \Pi \) that satisfies (3.3)-(3.4).

Finally, for the last statement in part (ii), if \( \mathcal{S}_* = \mathcal{S} \) or \( \mathcal{S}_m \), then by Theorem 3.2, \( \bar{f}(\tilde{\theta}) \in \Pi' \) or \( \Pi'_m \), respectively, for all \( \tilde{\theta} \in [0,1]^{\infty} \). Hence by assumption (3.5), \( \inf_{p \in \tilde{\mathcal{S}}_x(x)} \phi(x, p) \geq \inf_{p \in \tilde{\mathcal{S}}_x'(x)} \phi(x, p) \) for all \( x \in \mathcal{X} \), which implies \( \phi^*(x) = \inf_{p \in \tilde{\mathcal{S}}_x'(x)} \phi(x, p) \) for all \( x \in \mathcal{X} \).

Remark 3.3. We make two comments regarding the preceding theorem and related results for MDPs with Borel measurable policies.

(a) For MDPs with Borel \( \Gamma \) and Borel measurable policies, optimality results similar to Theorem 3.3(i)-(ii) were given in Strauch [49, Thms. 7.1, 8.1] and Dynkin and Yushkevich [16, Sec. 5.5] for the discounted and total cost criteria, and in Feinberg [18, Thm. 3.1] and [21, Lem. 4.1, Thm. 4.2] for arbitrary Borel measurable numerical criteria; results related to Theorem 3.3(iii) were given in Feinberg [19, Thm. 2] and [18, Thm. 3.2, Cors. 3.2, 3.3, 3.3.5]. As noted in our introduction, these prior results showed that w.r.t. any given finite measure on \( \mathcal{X} \), there exist (Borel measurable) randomized or nonrandomized, almost-everywhere \( \epsilon \)-optimal policies. This is different from Theorem 3.3(ii) or (iii), in which the policy \( \pi^* \) is \( \epsilon \)-optimal everywhere.

(b) In the prior works just mentioned, \( \Gamma \) is Borel and admits a Borel measurable selection. As discussed earlier in Remark 3.1(b), in this case every \( \mathcal{S}_* \) in the preceding theorem coincides with its subset of strategic measures induced by Borel measurable policies. Moreover, the set \( \mathcal{S}_* \) admits a Borel measurable selection (since the mapping \( x \mapsto \rho_x[\mu] \), where \( \mu \) is any Borel measurable stationary policy, is a Borel measurable selection of \( \mathcal{S}_* \)). However, for the partial minimization problem in the preceding theorem, this ensures only the feasibility of at least one Borel measurable policy \( \pi \), in the sense that \( \rho_x[\pi] \) lies in the feasible set \( \tilde{\mathcal{S}}_x(x) \) for every state \( x \). The level sets of the objective function, \( \{(x,p) \in \tilde{\mathcal{S}}_x \mid \phi(x,p) < r \}, \ r \in \mathbb{R}, \) need not admit Borel measurable selections even if \( \phi \) is Borel measurable. This is the reason that if we only allow Borel measurable policies, in general we can only obtain an almost-everywhere \( \epsilon \)-optimal policy.

As mentioned earlier, the results of this section can be generalized to partially observable problems; see Sections 4.1.3 and 5.3.1. In addition, our proof arguments given in Section 5, which deal with structured policies, can be adapted and applied to decentralized control problems like those studied in [56, 57] to derive similar results in the universal measurability framework.
4 Applications and Extensions

In this section we first specialize the results of Section 3 to risk-neutral and risk-sensitive MDPs, focusing on performance criteria of the average type. We then extend our analysis to partially observable problems (Section 4.1.3) and minimax control problems (Section 4.2), with the latter being the major extension.

4.1 MDPs with Risk-Neutral and Risk-Sensitive Average Criteria

We will consider various average cost and risk criteria in this subsection and apply Theorem 3.3 to derive optimality results for MDPs under these criteria. If \( J^* : \Pi \times \mathcal{X} \to \mathbb{R} \) is a cost or risk criterion, w.r.t. \( J^* \) and a subset \( \Pi_* \) of policies in \( \Pi \), we define the optimal average cost or risk function by
\[
g^*(x) := \inf_{\pi \in \Pi_*} J^*(\pi, x), \quad x \in \mathcal{X},
\]
and we say a policy \( \pi \in \Pi_* \) is optimal for state \( x \) if \( J^*(\pi, x) = g^*(x) \), and \( \epsilon \)-optimal for state \( x \) (where \( \epsilon > 0 \)) if
\[
J^*(\pi, x) \leq \begin{cases} g^*(x) + \epsilon & \text{if } g^*(x) > -\infty, \\ -\epsilon^{-1} & \text{if } g^*(x) = -\infty. \end{cases}
\]

If these relations hold for all states \( x \in \mathcal{X} \), \( \pi \) is called (\( \epsilon \))-optimal w.r.t. \( J^* \) and \( \Pi_* \).

4.1.1 The Risk-Neutral Case

Let \( c^+ \) (\( c^- \)) be the positive (negative) part of the one-stage cost function \( c \).

**Definition 4.1.** We say an MDP is in the model class AC\(^+\)(AC\(^-\)), if for \( c^+ = c^- \) (\( c^+ = c^- \)),
\[
\mathbb{E}_x^\pi \left( \sum_{k=0}^{n-1} c(x_k, a_k) \right) < +\infty, \quad \forall x \in \mathcal{X}, \pi \in \Pi, n \geq 0.
\]

We say an MDP is in the model class \( \tilde{\text{AC}}^+ \)(\( \tilde{\text{AC}}^- \)), if \( c(\cdot) \) is real-valued and bounded below (above) on \( \Gamma \).

For MDPs in the AC\(^+\) or AC\(^-\) class, the long-run expected average cost criteria \( J^{(i)}, 1 \leq i \leq 4 \), given below are all well-defined (i.e., they do not involve \( \infty - \infty \)). For \( \pi \in \Pi, x \in \mathcal{X} \),
\[
J^{(1)}(\pi, x) := \limsup_{n \to \infty} n^{-1} J_n(\pi, x), \quad J^{(2)}(\pi, x) := \liminf_{n \to \infty} n^{-1} J_n(\pi, x),
\]
\[
J^{(3)}(\pi, x) := \limsup_{n \to \infty} n^{-1} J_{n,j}(\pi, x), \quad J^{(4)}(\pi, x) := \liminf_{n \to \infty} n^{-1} J_{n,j}(\pi, x),
\]
where for \( n \geq 1 \) and \( j \geq 0 \),
\[
J_n(\pi, x) := \mathbb{E}_x^\pi \left( \sum_{k=0}^{n-1} c(x_k, a_k) \right), \quad J_{n,j}(\pi, x) := \mathbb{E}_x^\pi \left( \sum_{k=0}^{n-1} c(x_{k+j}, a_{k+j}) \right).
\]

(\( J_n \) is the standard \( n \)-stage cost function; \( J_{n,j} \) instead measures the expected costs incurred from time \( j \) to time \( n - j - 1 \).) For the sub-models \( \tilde{\text{AC}}^+ \) and \( \tilde{\text{AC}}^- \), we consider four more well-defined average cost criteria \( \tilde{J}^{(i)}, 1 \leq i \leq 4 \), which involve average costs along sample paths: for \( \pi \in \Pi, x \in \mathcal{X} \),
\[
\tilde{J}^{(1)}(\pi, x) := \mathbb{E}_x^\pi \left( \limsup_{n \to \infty} \tilde{c}_n \right), \quad \tilde{J}^{(2)}(\pi, x) := \mathbb{E}_x^\pi \left( \liminf_{n \to \infty} \tilde{c}_n \right),
\]
\[
\tilde{J}^{(3)}(\pi, x) := \mathbb{E}_x^\pi \left( \limsup_{n \to \infty} \tilde{c}_{n,j} \right), \quad \tilde{J}^{(4)}(\pi, x) := \mathbb{E}_x^\pi \left( \liminf_{n \to \infty} \tilde{c}_{n,j} \right),
\]
where \( \tilde{c}_n := n^{-1} \sum_{k=0}^{n-1} c(x_k, a_k) \) and \( \tilde{c}_{n,j} := n^{-1} \sum_{k=0}^{n-1} c(x_{k+j}, a_{k+j}) \).
Remark 4.1. The standard average cost criterion is $J^{(3)}$. Our definitions of the criteria $J^{(i)}, i = 3, 4$, were motivated by a theorem about Banach limits [35, Thm. 3.4.1]; see [55, Remark 2.1] for more details. Regarding the criteria $J^{(i)}, i = 1, 2$, it has been shown in [6, 17] that interestingly, if $X$ is finite and $c(\cdot)$ is bounded below (the action space can be arbitrary), then the optimal average cost functions w.r.t. these criteria and $\Pi$ are identical. Moreover, in this case, there exists a nonrandomized Markov policy that is $\epsilon$-optimal simultaneously w.r.t. all these criteria, and under additional compactness/continuity conditions, this policy can be taken to be nonrandomized and stationary. The finiteness of the state space is critical for these results, however. For an infinite state space $X$, the optimal average cost functions w.r.t. these criteria are in general not the same. □

Consider the case $\Pi_s = \Pi, \Pi_{sm},$ or $\Pi_{ss},$ and correspondingly let $S_\ast = S, S_m$ or $S_s,$ respectively. Note that if $\pi$ is a semi-Markov or semi-stationary policy, then for each initial state $x \in X, \rho_x[\pi] \in S_m$ or $S_s,$ respectively. Let $J^\ast$ be any criterion given above. Since $J^\ast(\pi, x)$ can be expressed as a function of $P^\pi_x,$ we can write the optimal average cost function $g^\ast$ w.r.t. $J^\ast$ and $\Pi_s,$ as

$$g^\ast(x) := \inf_{\pi \in \Pi_s} J^\ast(\pi, x) = \inf_{p \in S_s(x)} \phi(p), \quad x \in X. \quad (4.8)$$

Here we define the function $\phi : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ by using the same defining equation for the criterion $J^\ast,$ except that we replace the expectation w.r.t. $P^\pi_x$ by the expectation w.r.t $p.$ For technical convenience, in this definition, we extend $c(\cdot)$ to the entire state-action space by letting $c(\cdot) \equiv +\infty$ outside $\Gamma,$ which makes $c(\cdot)$ a lower semianalytic function on $X \times \mathcal{A}.$ (Recall also that by convention $\infty - \infty = -\infty + \infty = \infty.$)

Lemma 4.1. The function $\phi : \mathcal{P}(\Omega) \to \overline{\mathbb{R}}$ corresponding to any average cost criterion defined in (4.4)-(4.7) is lower semianalytic.

Proof. Consider the case $J^{(3)},$ for example. The corresponding function $\phi$ is given by $\phi(p) := \limsup_{n \to \infty} \sup_{j \geq 0} \sup_{i \geq 0} \sum_{k=0}^{n-1} c(x_{k+j}, a_{k+j}) p(d\omega).$ Since $c(\cdot)$ is lower semianalytic by our model assumption, $\phi$ is lower semianalytic by [5, Lem. 7.30, Cor. 7.48.1]. The same argument applies to the other criteria under consideration. □

Theorem 4.1. Consider these two cases of an average-cost MDP:

1. The MDP is in the class $AC^+$ or $AC^-$ and the criterion $J^\ast \in \{J^{(i)}\}_{1 \leq i \leq 4}.$
2. The MDP is in the class $\tilde{AC}^+$ or $\tilde{AC}^-$ and the criterion $J^\ast \in \{\tilde{J}^{(i)}\}_{1 \leq i \leq 4}.$

In each case, w.r.t. $J^\ast$ and $\Pi_s \in \{\Pi, \Pi_{sm}, \Pi_{ss}\},$ the following hold:

(a) The optimal average cost function $g^\ast$ is lower semianalytic.

(b) For each $\epsilon > 0,$ there exists an $\epsilon$-optimal policy $\pi^\ast \in \Pi_s,$ that is, moreover, optimal for every state that admits an optimal policy.

(c) In Case 1, the optimal average cost function w.r.t. $\Pi$ coincides with that w.r.t. $\Pi_{sm},$ and thus the $\epsilon$-optimal policy $\pi^\ast$ w.r.t. $\Pi$ given in (b) can be taken to be semi-Markov.

(d) If $\Pi_s \in \{\Pi, \Pi_{sm}\},$ then for Case 1 with $J^\ast \in \{J^{(2)}, J^{(4)}\}$ and $c(\cdot)$ being bounded from below, as well as for Case 2, the $\epsilon$-optimal policy $\pi^\ast$ in (b) can be taken to be nonrandomized.

Proof. Parts (a)-(b) follow from (4.8), Lemma 4.1, and Theorem 3.3(i)-(ii). Part (c) follows from the following two facts. First, for each initial distribution $p_0$ and policy $\pi \in \Pi,$ there is a Markov policy $\pi_m \in \Pi_m$ under which the marginal distributions of $(x_n, a_n), n \geq 0,$ coincide with those under $\pi$ (cf. the proof of [48, Prop. 1]). Second, the criteria $J^{(i)}, 1 \leq i \leq 4,$ can be expressed as functions of those marginal distributions of $(x_n, a_n), n \geq 0.$ Under the assumptions of part (d), condition (3.5) in Theorem 3.3(iii) is satisfied by the function $\phi.$ In particular, under the stated assumptions for Case 1, inequality (3.5) holds by Theorem 3.2(ii) and Fatou’s lemma; for Case 2, clearly (3.5) holds with equality by Theorem 3.2(ii). Thus part (d) follows from Theorem 3.3(iii). □
Remark 4.2. We make a few comments on the preceding theorem and related results for average-cost MDPs.
(a) For the standard average cost criterion \( J^{(1)} \), Theorem 4.1(c) asserts the existence of a randomized semi-Markov \( \epsilon \)-optimal policy that is, moreover, optimal for every state that admits an optimal policy. This conclusion is the strongest possible, without extra conditions on the MDP. Indeed it is known that even in an MDP with a countable state space, a finite action space, and bounded one-stage costs, there need not exist a nonrandomized semi-Markov \( \epsilon \)-optimal policy [16, Chap. 7, Example 3] nor a randomized Markov \( \epsilon \)-optimal policy [17, Sec. 5]. In both of these counterexamples, there exists an optimal policy for each state.
(b) For the criteria \( J^{(1)}, J^{(3)}, \tilde{J}^{(i)}, i \leq 4 \), using Theorem 4.1, we can further show that under certain reachability and boundedness conditions, the optimal average cost functions w.r.t. \( \Pi \) are constant almost everywhere w.r.t. certain \( \sigma \)-finite measures (see [55, Lem. 2.4, Cor. 3.2] for more details).
(c) For MDPs with the average cost criterion \( J^{(1)} \) and with Borel measurable policies, there are extensive studies on the existence of nonrandomized stationary optimal policies and the validity of optimality equations or optimality inequalities, under various continuity/compactness or ergodicity conditions, and with a variety of methods. Some of the results have also been extended to MDPs with universally measurable policies. The literature is too vast to list here; we refer the interested reader to the surveys and book chapters [1, 11, 29], [27, Chaps. 5-6], [28, Chaps. 10-12], the recent articles [22, 23, 51, 54], and the references therein.

4.1.2 The Risk-Sensitive Case

The results of Section 3 on strategic measures are applicable to a large class of risk-sensitive MDPs. As concrete examples, we discuss below several criteria related to average risk. For simplicity, in this discussion, we assume that the one-stage cost function \( c \) is real-valued on \( \Gamma \).

Following a general framework in [4], we define the first two average risk criteria through a strictly increasing, nonnegative continuous function \( \psi : \mathbb{R} \to [0, +\infty] \) as follows. For all \( \pi \in \Pi \), let

\[
J_\psi(\pi, x) := \limsup_{n \to \infty} \frac{1}{n} \psi^{-1} \left( E_x \left[ \psi \left( \sum_{k=0}^{n-1} c(x_k, a_k) \right) \right] \right), \quad x \in \mathcal{X},
\]

(4.9)

\[
\tilde{J}_\psi(\pi, x) := \limsup_{n \to \infty} \sup_{j \geq 0} \frac{1}{n} \psi^{-1} \left( E_x \left[ \psi \left( \sum_{k=0}^{n-1} c(x_{j+k}, a_{j+k}) \right) \right] \right), \quad x \in \mathcal{X}.
\]

(4.10)

In the special case where \( \psi(z) := e^{\beta z} \) for some \( \beta > 0 \), (4.9) becomes

\[
J_\psi(\pi, x) = \limsup_{n \to \infty} \frac{1}{\beta n} \log E_x \left[ e^{\beta \sum_{k=0}^{n-1} c(x_k, a_k)} \right], \quad x \in \mathcal{X},
\]

which is related to the standard exponential utility and has been considered in many works on risk-sensitive MDPs under certain continuity/compactness model assumptions (see, e.g., [13, 30]).

We mention that the non-negativity of \( \psi \) is assumed for simplicity. As in [4], we can also treat a larger class of criteria without this restriction, using the same arguments given below, if we introduce additional conditions to ensure that the expectations in (4.9)-(4.10) are well-defined.

The next two criteria are based on Conditional Value-at-Risk (CVaR) and Value-at-Risk (VaR), two popular criteria for risk-averse control. In particular, for \( \omega \in \Omega \), define the average cost along a sample path by \( \bar{c}(\omega) := \limsup_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} c(x_k, a_k) \) (one can also define \( \bar{c}(\omega) \) differently; the idea is the same). For \( r \in \mathbb{R} \), let \( \{r\}_+ := \max\{r, 0\} \). Define the risk criteria \( R_1, R_2 \) to be the CVaR\(\alpha\) and VaR\(\alpha\) of the random variable \( \bar{c}(\omega) \) under a policy \( \pi \in \Pi \), respectively, for some parameter \( \alpha \in (0, 1] \) that represents the level of risk-aversion:

\[
R_1(\pi, x) := \inf_{z \in \mathbb{R}} \left\{ z + \frac{1}{E_x^\pi} \left[ (\bar{c}(\omega) - z)_+ \right] \right\}, \quad x \in \mathcal{X},
\]

(4.11)
R_2(\pi, x) := \inf \{ z \in \mathbb{R} \mid \mathbb{P}_x^\pi \{ \bar{c}(\omega) \leq z \} \geq 1 - \alpha \}, \quad x \in X. \quad (4.12)

(As a side note, if for some constant \ell, \bar{c}(\omega) = \ell, \mathbb{P}_x^\pi-almost surely, then \ R_1(\pi, x) = R_2(\pi, x) = \ell. This “degenerate” case can happen if \pi is stationary and satisfies certain ergodicity conditions.) For the probabilistic interpretations of CVaR \alpha and VaR \alpha, and their use in stochastic programming/control, we refer the interested reader to [46, Chap. 6], [52], and the references therein.

Every risk criterion given above can be expressed as a function of \mathbb{P}_x^\pi. So, as in the risk-neutral case, we can express the optimal risk function \mathcal{G}^* w.r.t. the criterion and a certain subset of policies as a partial minimization problem of the form (4.8) on strategic measures. The function \phi(p) in (4.8) is defined in the same way as before, by extending \psi(-) to \mathbb{R} \times \mathbb{R} and by substituting \alpha for \alpha in the defining equation for the criterion. To apply Theorem 3.3, let us show that \phi is lower semianalytic.

**Lemma 4.2.** Let \( X \) be a Borel space, \( f : X \rightarrow \mathbb{R} \), and \( s : \mathbb{R} \rightarrow \mathbb{R} \). If \( f \) is lower semianalytic and \( s \) is non-decreasing, then \( s \circ f \) is lower semianalytic.

**Proof.** By the definition of lower semianalytic functions, we need to show that for any \( r \in \mathbb{R} \), the set \( B := \{ x \in X \mid s(f(x)) > r \} \) is analytic. Since \( s \) is non-decreasing, \( E := \{ a \in \mathbb{R} \mid s(a) < r \} \) is either the empty set or a nonempty interval of the form \( (-\infty, r) \) or \( [r, \infty) \) for some \( r \in \mathbb{R} \). In the first case, \( B = \emptyset \) and is thus analytic. In the second case, we have \( B = \{ x \in X \mid f(x) > r \} \) or \( \{ x \in X \mid f(x) \leq r \} \), which is analytic by [5, Lem. 7.30(1)] since \( f \) is lower semianalytic.

**Lemma 4.3.** Let \( c(\cdot) \) be real-valued on \( \Gamma \). The function \( \phi : \mathcal{P}(\Omega) \rightarrow \mathbb{R} \) corresponding to any risk criterion defined in (4.9)-(4.12) is lower semianalytic.

**Proof.** Consider first the criteria \( J_\psi, \bar{J}_\psi \) in (4.9)-(10.10). Since \( \psi \) is strictly increasing, we can extend \( \psi^{-1} \) to \( \mathbb{R} \) in such a way that the extension is a nondecreasing function on \( \mathbb{R} \). To simplify notation, we use the same symbol \( \psi^{-1} \) to refer to this extension in the proof. Since \( c(\cdot) \) is lower semianalytic, for all \( n \geq 1, j \geq 0 \), the function \( v_{n,j}(\omega) := \sum_{k=0}^{n-1} c(x_{j+k}, a_{j+k}) \) is lower semianalytic [5, Lem. 7.30(4)], and hence the function \( \psi(v_{n,j}(\omega)) \) is lower semianalytic by the monotonicity of \( \psi \) and Lemma 4.2.

Then by [5, Prop. 7.48], the function \( \psi^{-1}(f_{n,j}(p)) := \psi(v_{n,j}(\omega))p(dw) \) is lower semianalytic on \( \mathcal{P}(\Omega) \), so by Lemma 4.2 \( \psi^{-1}(f_{n,j}(p)) \) is lower semianalytic on \( \mathcal{P}(\Omega) \) as well. Finally, it follows from [5, Lem. 7.30(2) and (4)] that the function \( \phi(p) = \limsup_{n \rightarrow \infty} \psi^{-1}(f_{n,j}(p)) \) for the criterion \( J_\psi \) and the function \( \phi(p) = \limsup_{n \rightarrow \infty} \sup_{j \geq 0} \frac{1}{\pi} \psi^{-1}(f_{n,j}(p)) \) for the criterion \( \bar{J}_\psi \) are both lower semianalytic.

Consider now the criterion \( R_1 \) given in (4.11). Since \( c(\cdot) \) is lower semianalytic, by [5, Lem. 7.30(2) and (4)], the function \( \bar{c}(\cdot) \) is lower semianalytic on \( \Omega \), and then the function \( f(\omega, z) := \psi^{-1}(\bar{c}(\omega) - z) \) is lower semianalytic on \( \Omega \times \mathbb{R} \) by [5, Lem. 7.30(2) and (4)]. Consequently, the function \( (p, z) \mapsto \int \bar{f}(\omega, z)p(dw) \) is lower semianalytic on \( \mathcal{P}(\Omega) \times \mathbb{R} \) by [5, Prop. 7.48]. It then follows from [5, Lem. 7.30(4), Prop. 7.47] that the result of the partial minimization, \( \phi(p) = \inf_{z \in \mathbb{R}} \{ z + \int f(\omega, z)p(dw) \} \), is a lower semianalytic function on \( \mathcal{P}(\Omega) \).

Finally, consider the criterion \( R_2 \) given in (4.12). Since \( \bar{c}(\cdot) \) is lower semianalytic, \( (\omega, z) \mapsto \bar{c}(\omega) - z \) is a lower semianalytic function on \( \Omega \times \mathbb{R} \) by [5, Lem. 7.30(4)], so the set \( E := \{ (\omega, z) \in \Omega \times \mathbb{R} \mid \bar{c}(\omega) - z \leq 0 \} \) is analytic. Then by [5, Prop. 7.48], the function \( (p, z) \mapsto \int 1_E(\omega, z)p(dw) = p(\bar{c}(\omega) \leq z) \) is upper semianalytic on \( \mathcal{P}(\Omega) \times \mathbb{R} \) and hence the set \( D := \{ (p, z) \in \mathcal{P}(\Omega) \times \mathbb{R} \mid p(\bar{c}(\omega) \leq z) \geq 1 - \alpha \} \) is analytic. By the definition of \( \phi \) in this case, we have \( \phi(p) = \inf_{z \in D_p} z, p \in \mathcal{P}(\Omega) \), where \( D_p \) is the \( p \)-section of \( D \). Therefore \( \phi \) is lower semianalytic by [5, Prop. 7.47].

**Remark 4.3.** If \( \Gamma \) and \( c(\cdot) \) are Borel measurable, the function \( \phi \) for any criterion defined in (4.9)-(4.12) is Borel measurable. This is obvious for (4.9)-(10.10). For (4.11)-(4.12), this can be seen from an alternative expression of \( \phi \) given in (4.15) and (4.16), respectively, in Section 4.2.1.

From Lemma 4.3 and the proof of Theorem 3.3, we obtain the following:

**Theorem 4.2.** Let \( c(\cdot) \) be real-valued on \( \Gamma \). Let \( J^* \) be any risk criterion defined in (4.9)-(4.12), and let \( \Pi_c \in \{ \Pi, \Pi', \Pi_{ss}, \Pi_{ss}', \Pi_{ss}' \} \). Then w.r.t. \( J^* \) and \( \Pi_c \), the optimal risk function \( g^* \) is lower semianalytic, and for each \( \epsilon > 0 \), there exists an \( \epsilon \)-optimal policy \( \pi^* \in \Pi_c \), that is, moreover, optimal for every state that admits an optimal policy.
4.1.3 Partially Observable Problems

We consider a general non-stationary model for stochastic control with imperfect state information. It is similar to the one defined in [5, Def. 10.3] and can be described as follows:

- The state space \( \mathcal{X} \), the action space \( \mathcal{A} \), and the state transition stochastic kernel \( q(dy | x, a) \) are the same as defined in Section 2.2.
- The observation space \( \mathcal{Z} \) is a Borel space. There is a Borel measurable mapping \( f : \mathcal{X} \to \mathcal{Z} \), which we call the observation function. For \( n \geq 0 \), observable to the controller at the nth stage is not \( x_n \) but the value of \( z_n = f(x_n) \). We call \( i_n := (z_0, a_0, z_1, a_1, \ldots, z_n) \) an nth information vector and denote its space \((\mathcal{Z} \times \mathcal{A})^n \times \mathcal{Z} \) by \( I_n \).
- The control constraint at the nth stage depends on the nth information vector realized in the system. Specifically, there is a set-valued mapping \( U_n : i_n \mapsto U_n(i_n) \) on \( I_n \), where \( U_n(i_n) \subset \mathcal{A} \) is a nonempty set of admissible actions when \( i_n \) is the realized nth information vector. We assume that the graph of \( U_n \), \( \Gamma_n := \{(i_n, a_n) \mid i_n \in I_n, a_n \in U_n(i_n)\} \), is analytic.
- The distribution \( p_0 \) of the initial state \( x_0 \) is known to the controller.

In this model we let \( \Omega := (\mathcal{X} \times \mathcal{Z} \times \mathcal{A})^\infty \) and define a policy to be a sequence of stochastic kernels \( \mu_n, n \geq 0 \), where \( \mu_n \) is a universally measurable stochastic kernel on \( \mathcal{A} \) given \( I_n \times \mathcal{P}(\mathcal{X}) \) such that

\[
\mu_n(U_n(i_n) \mid i_n; p_0) = 1, \quad \forall i_n \in I_n, p_0 \in \mathcal{P}(\mathcal{X}).
\]

The space of all policies is denoted by \( \Pi_I \), and the subset of all nonrandomized policies is denoted by \( \Pi'_I \). Since \( \Gamma_n, n \geq 0 \) are analytic, by the Jankov-von Neumann selection theorem [5, Prop. 7.49], there exists a sequence of analytically measurable functions \( f_n : I_n \to \mathcal{A} \) with the graph of \( f_n \) contained in \( \Gamma_n \). Such a sequence corresponds to a nonrandomized policy in \( \Pi'_I \). The sets \( \Pi'_I \) and \( \Pi_I \) are therefore nonempty.

Let \( J^*(\pi, p_0) \) be an extended real-valued function on \( \Pi_I \times \mathcal{P}(\mathcal{X}) \) that measures the performance of \( \pi \) for an initial distribution \( p_0 \). For the criterion \( J^* \) and \( \Pi_* \subseteq \{\Pi_I, \Pi'_I\} \), the optimal value function \( g^* \) and the (\( \epsilon \))-optimal policies are defined as in the case of MDPs with initial distributions \( p_0 \in \mathcal{P}(\Omega) \) in place of initial states \( x \in \mathcal{X} \) (cf. (4.1)-(4.2)).

Let \( S_I \) denote the set of strategic measures \( \rho_{p_0}[\pi] \) on \( \mathcal{B}(\Omega) \) induced by all policies \( \pi \in \Pi_I \) and initial distributions \( p_0 \in \mathcal{P}(\mathcal{X}) \). Let \( S'_I \) denote the subset of strategic measures corresponding to those nonrandomized policies in \( \Pi'_I \). Similarly to the case of MDPs, let \( \hat{S}_I := \{(p_0, p) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\Omega) \mid p \in S_I, p_0(p) = p_0\} \) and \( \hat{S}'_I := \{(p_0, p) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\Omega) \mid p \in S'_I, p_0(p) = p_0\} \), where \( p_0(p) \) is the initial distribution of \( x_0 \) w.r.t. \( p \). For each \( p_0 \in \mathcal{P}(\mathcal{X}) \), \( \hat{S}_I(p_0) \) and \( \hat{S}'_I(p_0) \) denote the \( p_0 \)-section of \( S_I \) and of \( S'_I \), respectively.

**Proposition 4.1.** For the partially observable control problem defined above, the following hold:

(i) The sets \( S_I, S'_I, \hat{S}_I, \) and \( \hat{S}'_I \) are analytic; they are Borel if \( \Gamma_n, n \geq 0 \), are Borel.

(ii) Let \( \zeta : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\Omega) \) be a universally measurable mapping such that \( \zeta(p_0) \in \hat{S}_I(p_0) \) for all \( p_0 \in \mathcal{P}(\mathcal{X}) \). Then there exists a universally measurable policy \( \pi \in \Pi_I \) such that \( \rho_{p_0}[\pi] = \zeta(p_0) \) for all \( p_0 \in \mathcal{P}(\mathcal{X}) \). If, in addition, \( \zeta(p_0) \in \hat{S}'_I(p_0) \) for all \( p_0 \in \mathcal{P}(\mathcal{X}) \), then \( \pi \) can be taken to be nonrandomized.

**Theorem 4.3.** Consider the partially observable control problem defined above. Suppose that the one-stage cost function \( c(\cdot) \) is real-valued and bounded above or below on \( \mathcal{X} \times \mathcal{A} \). Define \( J^* \) according to any average cost or risk criterion given in (4.4)-(4.7) or (4.9)-(4.12) by extending those definitions to the space of initial distributions. Then w.r.t. \( J^* \) and \( \Pi_* \subseteq \{\Pi_I, \Pi'_I\} \), the following hold:

(i) the optimal average cost/risk function \( g^* \) is lower semianalytic; and

(ii) for each \( \epsilon > 0 \), there exists an \( \epsilon \)-optimal policy \( \pi^* \in \Pi_* \) that is, moreover, optimal for every initial distribution that admits an optimal policy.
The proofs of Prop. 4.1 and Theorem 4.3, being similar to their counterparts in the case of MDPs, will be outlined in Section 5.3.1.

4.2 Minimax Control

We now extend our results to a class of minimax control problems.

4.2.1 Model and Assumptions

We first specify the minimax control model:

(a) The state space \( \mathcal{X} \), the action space \( \mathcal{A} \), and the state transition stochastic kernel \( q(dy \mid x, a) \) are the same as defined in Section 2.2.

(b) The action space \( \mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \), where \( \mathcal{A}_1 \) and \( \mathcal{A}_2 \) are Borel spaces and correspond to the action spaces for player 1 (the controller) and player 2 (the opponent or nature), respectively. In terms of its components, we write an action \( a = (a_1, a_2) \).

(c) For \( i = 1, 2 \), the control constraint for player \( i \) is given by a set-valued mapping \( A_i \) on \( \mathcal{X} \), \( A_i : x \mapsto A_i(x) \subset \mathcal{A}_i \), where \( A_i(x) \) is a nonempty subset of admissible actions for player \( i \) at state \( x \). The graphs of \( A_1 \) and \( A_2 \), denoted by \( \Gamma_1 \) and \( \Gamma_2 \), respectively, are analytic.

(d) The two players take actions simultaneously at each stage with player 1 never observing player 2’s actions. Specifically, for \( n \geq 0 \), let \( a_n := (a_{1n}, a_{2n}) \) where \( a_{in} \) is the action of player \( i \) at the \( n \)th stage. Prior to making these actions, player 2 has the knowledge of the history \( h_n := (x_0, a_0, \ldots, x_n) \) up to stage \( n \), whereas player 1 has access only to \( i_n := (x_0, a_0, x_1, a_1, \ldots, x_n) \) and never observes the actions taken by player 2. The space of \( i_n \) is denoted by \( I_n \).

Let us remark on the role of assumption (d). With this restriction on the information available to the controller, the problem becomes a game against “nature.” As will be seen later, while we can readily characterize strategic measures induced by cooperative or noncooperative players in partial-information settings that are much more general than the setting in (d), we need this assumption in order to reformulate minimax control problems as minimax problems on strategic measures.

We define a policy for player \( i \) to be a sequence \( \pi_i \) := \( \{\mu^i_n\}_{n \geq 0} \), where \( \mu^i_n \) is a universally measurable stochastic kernel on \( \mathcal{A}_i \), given the space of the variables observable to that player at the \( n \)th stage—namely, \( I_n \) for player 1 and \( H_n \) for player 2, such that

\[
\mu^i_n(A_1(x_n) \mid i_n) = 1 \quad \forall i_n \in I_n, \quad \mu^2_n(A_2(x_n) \mid h_n) = 1 \quad \forall h_n \in H_n.
\]

We denote the set of all policies for player \( i \) by \( \Pi_i \). Since \( \Gamma_1 \) and \( \Gamma_2 \) are analytic, \( \Pi_1 \) and \( \Pi_2 \) are nonempty by [5, Prop. 7.49].

Denote the probability measure on \( \mathcal{U}(\Omega) \) induced by a policy pair \((\pi_1, \pi_2)\) \( \in \Pi_1 \times \Pi_2 \) and initial distribution \( p_0 \in \mathcal{P}(\mathcal{X}) \) by \( \mathcal{P}^{\pi_1, \pi_2}_{p_0} \). The associated expectation operator is denoted by \( \mathbb{E}^{\pi_1, \pi_2}_{p_0} \), and the restriction of \( \mathbb{E}^{\pi_1, \pi_2}_{p_0} \) to \( \mathcal{B}(\Omega) \) is denoted by \( \rho_{p_0}[\pi_1, \pi_2] \). As before, if \( p_0 = \delta_\xi \), we write \( \xi \) instead of \( \delta_\xi \) in the subscripts.

Let \( J^* : \Pi_1 \times \Pi_2 \times \mathcal{X} \rightarrow \bar{\mathbb{R}} \) be some given function. We consider the minimax control problem for player 1 w.r.t. the criterion \( J^* \), where the optimal value function is given by

\[
g^*(x) := \inf_{\pi_1 \in \Pi_1} \sup_{\pi_2 \in \Pi_2} J^*(\pi_1, \pi_2, x), \quad x \in \mathcal{X}.
\]

A policy \( \pi_1 \in \Pi_1 \) is called optimal for state \( x \) if \( \sup_{\pi_2 \in \Pi_2} J^*(\pi_1, \pi_2, x) = g^*(x) \), and \( \epsilon \)-optimal for state \( x \) (where \( \epsilon > 0 \)) if

\[
\sup_{\pi_2 \in \Pi_2} J^*(\pi_1, \pi_2, x) \leq \begin{cases} 
g^*(x) + \epsilon & \text{if } g^*(x) > -\infty, \\
-1/\epsilon & \text{if } g^*(x) = -\infty. \end{cases}
\]
If these relations hold for all states $x \in \mathcal{X}$, $\pi_1$ is called an ($\epsilon$-)optimal policy of player 1 w.r.t. $J^*$.  

We now introduce two assumptions. The first one is an absolute continuity condition on the probability measures induced on $I_n$ when player 1 applies the same policy. It will allow us to rewrite the minimax control problem as a minimax optimization problem on strategic measures (cf. Prop. 4.3(ii) and the proof of Theorem 4.4).

**Assumption 4.1.** For any $\pi_1 \in \Pi_1$, $\pi_2, \tilde{\pi}_2 \in \Pi_2$, $x \in \mathcal{X}$, and $n \geq 0$, the marginal distributions of $\mathbb{P}_{x, \pi_2}^{\pi_1}$ and $\mathbb{P}_{x, \tilde{\pi}_2}^{\pi_1}$ on $I_n$ are absolutely continuous w.r.t. each other.

Note that if in the above we replace “for any $x \in \mathcal{X}$” by “for any initial distribution $p_0 \in \mathcal{P}(\mathcal{X})”, we obtain an equivalent assumption. We now describe a class of problems that satisfy Assumption 4.1.

**Example 4.1.** Suppose the state transition stochastic kernel $q(dy \mid x, a) = f(y, x, a) \eta(dy \mid x, a^1)$, where:

(a) $f$ is a strictly positive Borel measurable function on $\mathcal{X}^2 \times \mathcal{A}$; and

(b) $\eta$ is a nonnegative kernel on $\mathcal{X}$ given $\mathcal{X} \times \mathcal{A}_1$ such that (i) for each $B \in \mathcal{B}(\mathcal{X})$, $\eta(B \mid x, a^1)$ is Borel measurable in $(x, a^1)$, and (ii) for each $(x, a^1) \in \mathcal{X} \times \mathcal{A}_1$, $\eta(\cdot \mid x, a^1)$ is a nonnegative measure on $\mathcal{B}(\mathcal{X})$ and furthermore, $\eta(\cdot \mid x, a^1)$ is $\sigma$-finite uniformly in $(x, a^1)$ in the sense that for some sequence of Borel sets $B_k$ with $\mathcal{X} = \bigcup B_k$, we have $\sup_{x \in \mathcal{X}, a^1 \in \mathcal{A}_1} \eta(B_k \mid x, a^1) < \infty$ for all $k$.

Then Assumption 4.1 holds. Indeed, let $n \geq 0, p_0 \in \mathcal{P}(\mathcal{X})$, and let $\pi_1 := \{\mu_{i}^{0}\} \in \Pi_1$, and $\pi_2 := \{\tilde{\mu}_i^{0}\}, \tilde{\pi}_2 := \{\tilde{\mu}_i^{0}\} \in \Pi_2$. Consider a set $B \in \mathcal{U}(I_n)$ with $\mathbb{P}_{p_0}^{\pi_1, \pi_2}\{i_n \in B\} = 0$, and let us show that $\mathbb{P}_{p_0}^{\pi_1, \tilde{\pi}_2}\{i_n \in B\} = 0$, which will imply the desired conclusion since $\pi_2, \tilde{\pi}_2$ are arbitrary. To this end, let $\gamma_n$ be the unique nonnegative $\sigma$-finite measure on $\mathcal{U}(I_n)$ determined by the product of $p_0(dx_0)$ with the other stochastic kernels and nonnegative kernels given below:

$$p_0(dx_0), \mu_0^1(da_0^1 \mid x_0), \mu_2^1(da_0^2 \mid x_0), \eta(dx_1 \mid x_0, a_0^1), \ldots,$$

$$\ldots, \mu_{n-1}^1(da_{n-1}^1 \mid i_{n-1}), \mu_{n-1}^2(da_{n-1}^2 \mid h_{n-1}), \eta(dx_n \mid x_{n-1}, a_{n-1}^1).$$

Define the measure $\hat{\gamma}_n$ on $\mathcal{U}(I_n)$ similarly, with $\hat{\mu}_k^2$ replacing $\mu_k^2$ in the above, for all $k < n$.

The existence of $\gamma_n$ and $\hat{\gamma}_n$ follows from a repeated application of the product measure theorem given in [2, Thm. 2.6.2] together with the proof arguments of [5, Prop. 7.45] (the latter extend the result of the product measure theorem to $\mathcal{U}(I_n)$, which is not a product $\sigma$-algebra, by using the properties of universally measurable stochastic kernels). By these arguments, one can also show that $\mathbb{P}_{p_0}^{\pi_1, \pi_2}\{i_n \in B\} = \int_{H_n} \mathbb{I}_B(i_n) \prod_{k=1}^{n} f(x_k, x_{k-1}, a_{k-1}) \gamma_n(dh_n)$ and $\mathbb{P}_{p_0}^{\pi_1, \tilde{\pi}_2}\{i_n \in B\} = \int_{H_n} \mathbb{I}_B(i_n) \prod_{k=1}^{n} f(x_k, x_{k-1}, a_{k-1}) \hat{\gamma}_n(dh_n)$. Since $\mathbb{P}_{p_0}^{\pi_1, \pi_2}\{i_n \in B\} = 0$ and the function $f$ is strictly positive, we have $\int_{H_n} \mathbb{I}_B(i_n) \gamma_n(dh_n) = 0$. As can be seen from (4.14), the marginal measures of $\gamma_n$ and $\hat{\gamma}_n$ on $I_n$ must coincide, since $\eta(dx_{k+1} \mid x_k, a_k^1)$ and $\hat{\mu}_k^1(da_k^1 \mid i_k)$, $k < n$, do not depend on the actions of player 2. Therefore, $\int_{H_n} \mathbb{I}_B(i_n) \hat{\gamma}_n(dh_n) = 0$, implying $\mathbb{P}_{p_0}^{\pi_1, \tilde{\pi}_2}\{i_n \in B\} = 0$.

The second assumption places a measurability requirement on the criterion $J^*$. It is commonly satisfied in applications, as we will discuss shortly, using examples.

**Assumption 4.2.** Associated with the criterion $J^*$, there is an upper semianalytic function $\phi : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ such that $J^*(\pi_1, \pi_2, x) = \phi(\rho_x[\pi_1, \pi_2])$ for all $x \in \mathcal{X}, \pi_1 \in \Pi_1, \pi_2 \in \Pi_2$.

**Example 4.2.** Let $\Gamma := \{(x, a) \in \mathcal{X} \times \mathcal{A} \mid (x, a^2) \in \mathcal{A}_1, (x, a^2) \in \mathcal{A}_2\}$. Note that $\Gamma$ is analytic (Borel) if $\Gamma_1$ and $\Gamma_2$ are analytic (Borel); see Section 5.3.2 for a proof. Let $c : \Gamma \to \mathbb{R}$ be universally measurable. Consider any average cost or risk criterion $J^*$ discussed in Section 4.1, with $\pi_1, \pi_2$ in place of $\pi$ and with $\mathbb{P}_{x, \pi_2}^{\pi_1}$, $\mathbb{E}_{x, \pi_2}^{\pi_1}$ in place of $\mathbb{P}_{x, \pi_2}^{\pi_1}, \mathbb{E}_{x, \pi_2}^{\pi_1}$, respectively. As before, since $J^*$ can be expressed as a function of $\mathbb{P}_{x}^{\pi_1, \pi_2}$, by replacing $\mathbb{P}_{x}^{\pi_1, \pi_2}$ with $p \in \mathcal{P}(\Omega)$, we can define an associated function $\phi(p)$ on $\mathcal{P}(\Omega)$ that satisfies the equality condition in Assumption 4.2. Let us show that under suitable conditions on $c(\cdot)$ and the sets $\Gamma_1, \Gamma_2$, the function $\phi$ is upper semianalytic and hence those average cost or risk criteria discussed earlier all satisfy Assumption 4.2.
Consider first any average cost criterion $J^*$ defined in (4.4)-(4.7). For that criterion, assume a model condition analogous to one of the model classes given in Def. 4.1 for MDPs, so that for all $x \in \mathcal{X}, \pi_1 \in \Pi_1,$ and $\pi_2 \in \Pi_2,$ $J^*(\pi_1, \pi_2, x)$ is well-defined and does not involve the summation $-\infty - \infty$ or $-\infty + \infty$. Now suppose that $c(\cdot)$ is upper semianalytic on $\Gamma$. Since this means the function $-c(\cdot)$ is lower semianalytic on $\Gamma$, after taking care of some technical subtlety, we can apply arguments symmetric to those given in the proof of Lemma 4.1 to obtain that the function $\phi(p)$ corresponding to $J^*$ is upper semianalytic.

The technical subtlety just mentioned is as follows. In defining $\phi(p)$ on $\mathcal{P}(\Omega)$, we extend $c$ from $\Gamma$ to $\mathcal{X} \times \mathcal{A}$ and also allow summations involving both $\infty$ and $-\infty$. For an upper semianalytic $c(\cdot)$, we extend $c$ to $\mathcal{X} \times \mathcal{A}$ by letting $c(\cdot) \equiv -\infty$ outside $\Gamma$, and we take $\infty - \infty = -\infty + \infty = -\infty$ in defining $\phi(p)$. These are opposite to the conventions we use when dealing with lower semianalytic functions. (They are needed solely for technical convenience; in particular, for making the extension of $c$ and $\phi$ upper semianalytic on the entire spaces. They do not affect the values of $\phi(p)$ for strategic measures $p$ due to the model condition mentioned earlier.)

Consider now any risk criterion $J^*$ defined in (4.9)-(4.11). Assume that $c(\cdot)$ is real-valued and upper semianalytic on $\Gamma$. Then the function $\phi(p)$ associated with $J^*$ is upper semianalytic. The proof of this is entirely symmetric to the proofs of Lemmas 4.2 and 4.3 in the case of (4.9) or (4.10). For the CVaR-based criterion (4.11), the proof is slightly different. First, by an argument symmetric to that given in the proof of Lemma 4.3, the function $(p,z) \mapsto \frac{1}{\alpha} \int_{\Omega} (\bar{c}(\omega) - z)_+ + p(d\omega)$ is upper semianalytic on $\mathcal{P}(\Omega) \times \mathbb{R}$, and hence it is upper semianalytic in $p$ for each $z$. Using this and the fact that, with $Q$ being the countable set of rational numbers,

$$\phi(p) := \inf_{z \in \mathbb{Q}} \left\{ z + \alpha^{-1} \int_{\Omega} (\bar{c}(\omega) - z)_+ p(d\omega) \right\}$$

(which can be verified directly), it follows from [5, Lem. 7.30(2)] that $\phi$ is upper semianalytic.

Finally, for the VaR-based criterion (4.12), let us assume that the sets $\Gamma_1$ and $\Gamma_2$ are Borel and the function $c(\cdot)$ is real-valued and Borel measurable on $\Gamma$. Then by [5, Prop. 7.29], the set $D := \{(p,z) \in \mathcal{P}(\Omega) \times \mathbb{R} | p\{\bar{c}(\omega) \leq z\} \geq 1 - \alpha\}$ is Borel. It can be verified directly that the function $\phi(p)$ in this case satisfies

$$\phi(p) := \inf \left\{ \int_{\Omega} p\{\bar{c}(\omega) \leq z\} \right\} \geq 1 - \alpha$$

$$\phi(p) = \inf \left\{ z \in \mathbb{Q} | p\{\bar{c}(\omega) \leq z\} \geq 1 - \alpha \right\}, \quad p \in \mathcal{P}(\Omega).$$

Thus, if $\{z_1, z_2, \ldots\}$ is an enumeration of $Q$, we can express $\phi$ as $\phi(p) = \inf_{i \geq 1} f_i(p)$, for the Borel measurable functions $f_i$ defined by $f_i(p) := z_i$ if $(p, z_i) \in D$, and $f_i(p) = +\infty$ if $(p, z_i) \notin D$. This shows that $\phi$ is Borel measurable and hence upper semianalytic.

### 4.2.2 Properties of Strategic Measures and Optimality Results

Let $\mathcal{S}_M$ be the set of all strategic measures:

$$\mathcal{S}_M := \{ p \in \mathcal{P}(\Omega) | p = \rho_{p_0}[\pi_1, \pi_2], \pi_1 \in \Pi_1, \pi_2 \in \Pi_2, p_0 \in \mathcal{P}(\mathcal{X}) \}.$$

Define $\tilde{\mathcal{S}}_M := \{(x,p) \in \mathcal{X} \times \mathcal{P}(\Omega) | p \in \mathcal{S}_M, p_0(p) = \delta_x \}$ and denote the $x$-section of $\tilde{\mathcal{S}}_M$ by $\tilde{\mathcal{S}}_M(x)$. The proofs of the next two propositions are given in Section 5.3.2.

**Proposition 4.2.** For the minimax control problem defined above:

(i) The sets $\mathcal{S}_M$ and $\tilde{\mathcal{S}}_M$ are analytic; they are Borel if $\Gamma_1$ and $\Gamma_2$ are Borel.
Let \( \zeta : X \to \mathcal{P}(\Omega) \) be a universally measurable mapping such that \( \zeta(x) \in \hat{S}_M(x) \) for all \( x \in X \).

Then there exist universally measurable policies \( \pi_1 \in \Pi_1, \pi_2 \in \Pi_2 \) such that \( \rho_x[\pi_1, \pi_2] = \zeta(x) \) for all \( x \in X \).

For \( n \geq 0 \), let \( \hat{\nu}_n^i(da_i^n | i_n; p) \) be a Borel measurable stochastic kernel on \( A_1 \) given \( I_n \times \mathcal{P}(\Omega) \) such that for each \( p \in \mathcal{P}(\Omega) \), it is a conditional distribution of \( a_i^n \) given \( i_n \) w.r.t. \( p \). The existence of these stochastic kernels is ensured by \([5, \text{Cor. 7.27.1}]\) (cf. the discussion at the beginning of Section 5.1).

Let \( \hat{\nu}_n(p) \) denote the marginal of \( p \) on \( I_n \). Define

\[
\hat{S}_M := \left\{ (x, p, p') \in X \times \mathcal{P}(\Omega)^2 \mid (x, p) \in \hat{S}_M, (x, p') \in \hat{S}_M, p' \text{ satisfies (4.17)-(4.18)} \right\}
\]

where the relations (4.17) and (4.18) are as follows:

\[
p' \left\{ \hat{\nu}_n^i( \cdot | i_n; p') = \hat{\nu}_n^i( \cdot | i_n; p) \right\} = 1, \quad \forall n \geq 0, \tag{4.17}
\]

\[
\hat{\mu}_n(p') \text{ is absolutely continuous w.r.t. } \hat{\mu}_n(p), \quad \forall n \geq 0. \tag{4.18}
\]

We denote the \((x, p)\)-section of \( \hat{S}_M \) by \( \hat{S}_M(x, p) \).

**Proposition 4.3.** For the minimax control problem defined above:

(i) The set \( \hat{S}_M \) is analytic; it is Borel if \( \Gamma_1 \) and \( \Gamma_2 \) are Borel.

(ii) Under Assumption 4.1, if \( p = \rho_x[\pi_1, \pi_2] \) for some \( x \in X \) and \( (\pi_1, \pi_2) \in \Pi_1 \times \Pi_2 \), then the set \( \hat{S}_M(x, p) = \{ \rho_x[\pi_1, \pi_2] | \pi_2 \in \Pi_2 \} \).

We are now almost ready to present our optimality result. In order to fully settle the measurability questions involved, we still need an extension of those results listed in (d) near the end of Section 2.1. This requires an additional axiomatic assumption and is given by the following proposition under the axioms of ZFC and analytic determinacy. Its proof relies on a uniformization theorem of Kondô [33, Cor. 38.7] (proved under ZFC) and will be given in Section 5.3.3 for completeness. While this proposition is sufficient for our purpose, its conclusions actually hold for a larger class of sets \( D \) and functions \( f \), as will be explained in Remark 5.1 after the proof.

**Proposition 4.4.** Assume the axiom of analytic determinacy. Let \( X \) and \( Y \) be Borel spaces, let \( D \subset X \times Y \) be analytic, and let \( f : D \to \bar{\mathbb{R}} \) be upper semianalytic. Define \( f^* : \text{proj}_X(D) \to \bar{\mathbb{R}} \) by \( f^*(x) = \inf_{y \in D_x} f(x, y) \), where \( D_x \) is the \( x \)-section of \( D \). Then the following hold:

(i) The function \( f^* \) is universally measurable.

(ii) The set \( E^* := \{ x \in \text{proj}_X(D) \mid \arg \min_{y \in E} f(x, y) \neq \emptyset \} \) is universally measurable. For each \( \epsilon > 0 \), there exists a universally measurable function \( \psi : \text{proj}_X(D) \to Y \) such that \( \psi(x) \in D_x \) for all \( x \in \text{proj}_X(D) \) and

\[
f(x, \psi(x)) = f^*(x), \quad \forall x \in E^*; \quad \tag{4.19}
\]

\[
f(x, \psi(x)) \leq \begin{cases} f^*(x) + \epsilon & \text{if } f^*(x) > -\infty, \\ -1/\epsilon & \text{if } f^*(x) = -\infty, \end{cases} \quad \forall x \in \text{proj}_X(D) \setminus E^*. \quad \tag{4.20}
\]

**Remark 4.4.** Regarding analytic determinacy and the standard ZFC axioms:

(a) Roughly speaking, the axiom of analytic determinacy concerns a certain class of infinite-stage two-player games, called analytic games, in which the two players take turns to choose a natural number, thereby producing an infinite sequence of such numbers, and the set of winning outcomes is a member of the algebra generated by analytic sets in the space of all such sequences. The axiom asserts that all these games are determined (i.e., one of the two players has a winning strategy). For the precise definition and discussions on the mathematical evidence for this axiom and the related large cardinal axioms in set theory, see the book [33, Sec. 26.B] and the articles [39, p. 6582] and [34].
(b) The following assertion is known to be independent of ZFC (namely, ZFC cannot prove nor disprove it; cf. [36, p. 546] and [43, p. 677]): there exists a coanalytic set \( B \) in \( X \times Y := [0,1]^2 \) whose projection on \( X \) is not universally measurable. If such a set \( B \) exists, a counterexample to Prop. 4.4 can be constructed by letting \( D = [0,1]^2 \) and \( f(x, y) = \mathbb{1}_{B'}(x, y) \), and similar counterexamples to the universal measurability of the optimal value functions in MDPs and in minimax optimization problems have been discussed in [9, Ex. (48)] and [41, Ex. 1, Remark 4], respectively. Under ZFC and analytic determinacy, such sets \( B \) do not exist [33, Thm. 36.20]; see Section 5.3.3 for more details.

\[ \square \]

**Theorem 4.4.** Assume the axiom of analytic determinacy. Consider the minimax control problem defined above, and let Assumptions 4.1-4.2 hold. Then w.r.t. \( J^* \), we have the following:

(i) The optimal value function \( g^* \) for player 1 is universally measurable.

(ii) For each \( \varepsilon > 0 \), player 1 has a universally measurable \( \varepsilon \)-optimal policy \( \pi_1^\varepsilon \in \Pi_1 \) that is, moreover, optimal for every state that admits an optimal policy.

**Proof.** By Assumptions 4.1-4.2 and Prop. 4.3(ii), for all \( \pi_1 \in \Pi_1 \) and \( x \in \mathcal{X} \), we have

\[
\sup_{\pi_2 \in \Pi_2} J^*(\pi_1, \pi_2, x) = \sup_{p' \in \mathcal{S}_M(x, p)} \phi(p'), \quad \text{where} \quad p = \rho_x[\pi_1, \hat{\pi}_2] \text{ for some } \hat{\pi}_2 \in \Pi_2. \tag{4.21}
\]

In view of the definitions of \( \hat{S}_M \) and \( \tilde{S}_M \), this implies that for all \( x \in \mathcal{X} \),

\[
g^*(x) = \inf_{p \in \hat{S}_M(x)} f(x, p), \quad \text{where} \quad f : \hat{S}_M \to \mathbb{R} \ \text{with} \ f(x, p) := \sup_{p' \in \tilde{S}_M(x, p)} \phi(p').
\]

Since \( \phi \) is upper semianalytic by Assumption 4.2 and \( \hat{S}_M \) is analytic by Prop. 4.3(i), the function \( f \) is upper semianalytic on \( \hat{S}_M \) by [5, Prop. 7.47]. Recall that \( \hat{S}_M \) is analytic by Prop. 4.2(i). It then follows from Prop. 4.4(i) that \( g^* \) is universally measurable.

Let \( \varepsilon > 0 \). By Prop. 4.4(ii), there is a universally measurable mapping \( \zeta : \mathcal{X} \to \mathcal{P}(\Omega) \) with \( \zeta(x) \in \hat{S}_M(x) \) for all \( x \in \mathcal{X} \) such that

\[
f(x, \zeta(x)) = g^*(x), \quad \forall x \in E^* := \left\{ x \in \mathcal{X} \ \big| \ \arg\min_{p \in \hat{S}_M(x)} f(x, p) \neq \emptyset \right\} \tag{4.22}
\]

\[
f(x, \zeta(x)) \leq \begin{cases} g^*(x) + \varepsilon & \text{if } g^*(x) > -\infty, \\ -1/\varepsilon & \text{if } g^*(x) = -\infty, \end{cases} \quad \forall x \notin E^*. \tag{4.23}
\]

Let \( (\pi_1^*, \pi_2^*) \) be the pair of policies given by Prop. 4.2(ii) for this mapping \( \zeta \). By (4.21) and Prop. 4.2(ii), \( \sup_{\pi_2 \in \Pi_2} J^*(\pi_1^*, \pi_2, x) = f(x, \rho_x[\pi_1^*, \pi_2^*]) = f(x, \zeta(x)) \), so, in view of (4.22)-(4.23), the policy \( \pi_1^* \) is \( \varepsilon \)-optimal and moreover, optimal for every state that admits an optimal policy. \( \square \)

**Remark 4.5.** We mention two direct extensions of the preceding theorem.

(a) One can also define the optimal value function \( g^* \) and the \( (\varepsilon) \)-optimal policies w.r.t. the subset of nonrandomized policies, or the subset of randomized or nonrandomized, semi-Markov or semi-stationary policies. By similar arguments one can show that Theorem 4.4 remains valid if \( \Pi_1 \) is replaced by such a subset of policies of player 1.

(b) Another extension concerns the definition of \( g^* \). Suppose that we have a sequence of criteria \( J_n^* \), \( n \geq 0 \), all satisfying Assumption 4.2, and we define player 1’s optimal value function \( g^* \) as

\[
g^*(x) := \inf_{\pi_1 \in \Pi_1} \limsup_{n \to \infty} \sup_{\pi_2 \in \Pi_2} J_n^*(\pi_1, \pi_2, x), \quad x \in \mathcal{X}.
\]

Then, by redefining the function \( f(x, p) \) and following the same reasoning in the above proof, one can show that Theorem 4.4 remains valid. An example of this type of \( g^* \) is the function considered in [31] for robust MDPs: \( g^*(x) := \inf_{\pi_1 \in \Pi_1} \limsup_{n \to \infty} \sup_{\pi_2 \in \Pi_2} \rho_x \left[ \frac{1}{n-1} \sum_{k=0}^{n-1} \epsilon(x_k, a_k) \right] \).

\[ \square \]
Remark 4.6. Suppose that $\Gamma_1$ and $\Gamma_2$ admit Borel measurable selections so that both players have Borel measurable policies. Then $S_M$ coincides with the set of strategic measures induced by Borel measurable policy pairs and moreover, because of the absolute continuity condition in Assumption 4.1, one can show that for each initial state, player 1 has a Borel measurable $\epsilon$-optimal policy. This is similar to the MDP case (cf. Remark 3.1(b)). But this need not be true for general minimax problems without Assumption 4.1 (such as the stochastic games studied in [40]). The reason, roughly speaking, is that although one can always modify an $\epsilon$-optimal universally measurable policy of player 1 into a Borel measurable policy by “sacrificing” its $\epsilon$-optimality in some situations, unlike in MDPs, the probability of these situations happening can be made non-negligible by player 2’s adversarial responses. Thus for minimax control, Borel measurable policies, even if they exist, are in general inadequate. \hfill \Box

4.2.3 Some Implications on Optimality Equations/Inequalities

So far we have not discussed optimality equations or inequalities for the optimal value function $g^*$. Basic optimality results like the ones given above can be useful in deriving such relations for some performance criteria. We close this section with a few results and examples in this regard.

First, let us extend Prop. 4.2(ii) to parametrized policies (the proof is given in Section 5.3.2):

**Proposition 4.5.** For the minimax control problem defined above, suppose that $\Theta$ is a Borel space and $\zeta : \mathcal{X} \times \Theta \to \mathcal{P}(\Omega)$ is a universally measurable mapping with $\zeta(x, \theta) \in \tilde{S}_M(x)$ for all $x \in \mathcal{X}$ and $\theta \in \Theta$. Then there exist policies $\pi_1(\theta) \in \Pi_1$, $\pi_2(\theta) \in \Pi_2$, parametrized by $\theta$, such that

(i) $\mu_n[\pi_1(\theta), \pi_2(\theta)] = \zeta(x, \theta)$ for all $x \in \mathcal{X}$ and $\theta \in \Theta$; and

(ii) with $\pi_1(\theta) = \{\mu_n[\theta]\}_{n \geq 0}$ and $\pi_2(\theta) = \{\mu_n[\theta]\}_{n \geq 0}$, the stochastic kernels $\mu_n[\theta](da_n | i_n)$ and $\mu_n[\theta](da_n^2 | h_n)$ are universally measurable in $(i_n, \theta)$ and $(h_n, \theta)$, respectively, for each $n \geq 0$.

We will use this proposition in proving the next result, which is about the existence of policies of player 2 that can respond to player 1’s strategy in a nearly optimal way from time 1 onwards. These policies will then be used to derive an inequality on $g^*$ for some cost criteria.

Let $\pi_1 := \{\mu_n^1\}_{n \geq 0} \in \Pi_1$ and $\pi_2 := \{\mu_n^2\}_{n \geq 0} \in \Pi_2$. Using their stochastic kernels for time $n \geq 1$, we define their “shifted” policies $\pi_n^1(\tilde{i}_1) := \{\mu_n^{1}[\tilde{i}_1]\}_{n \geq 0} \in \Pi_1$ and $\pi_n^2(\tilde{h}_1) := \{\mu_n^{2}[\tilde{h}_1]\}_{n \geq 0} \in \Pi_2$ for each $\tilde{i}_1 := (\tilde{x}_0, \tilde{a}_0^1) \in \mathcal{X} \times \mathcal{A}_1$ and $\tilde{h}_1 := (\tilde{x}_0, \tilde{a}_0^2) \in \mathcal{X} \times \mathcal{A}$ as follows: for $n \geq 0, i_n \in I_n, h_n \in H_n$,

$$\mu_n^{1}[\tilde{i}_1](\cdot | i_n) := \mu_n^{1+1}(\cdot | \tilde{i}_1, i_n), \quad \mu_n^{2}[\tilde{h}_1](\cdot | h_n) := \mu_n^{2+1}(\cdot | \tilde{h}_1, h_n).$$ (4.24)

**Proposition 4.6.** Consider the minimax control problem defined above. Under Assumptions 4.1-4.2, for each $\tilde{i}_1 \in \Pi_1$ and $\epsilon > 0$, player 2 has a policy $\pi_2 \in \Pi_2$ such that

$$J^*(\tilde{i}_1^*, \pi_1^2(\tilde{h}_1), x) \geq f_\epsilon(\tilde{i}_1^*, \pi_1^2(\tilde{h}_1), x), \quad \forall \tilde{i}_1 \in \mathcal{X} \times \mathcal{A}_1, \tilde{h}_1 = (\tilde{i}_1, \tilde{a}_0^2) \in \mathcal{X} \times \mathcal{A},$$ (4.25)

where $f_\epsilon : \Pi_1 \times \mathcal{X} \to [0, \infty]$ is given by $f_\epsilon(\pi^*_1(\cdot), x) := \sup_{\pi_2^* \in \Pi_2} J^*(\pi^*_1, \pi_2^*, x) - \epsilon$ if $\sup_{\pi_2^* \in \Pi_2} J^*(\pi^*_1, \pi_2^*, x) < +\infty$ and $f_\epsilon(\pi^*_1(\cdot), x) := 1/\epsilon$ otherwise.

**Remark 4.7.** A corollary of Prop. 4.6 is that for any $\pi_1 \in \Pi_1$, $\sup_{\pi_2 \in \Pi_2} J^*(\pi_1, \pi_2, x)$ is universally measurable in $x$. To see this, let $\tilde{\pi}_1(\cdot) \equiv \pi_1$ and for $k \geq 1$, let $\pi_{2,k}$ be a policy given by Prop. 4.6 for $\epsilon = 1/k$. Then for a fixed $\tilde{h}_1$, observe that $J^*(\pi_1, \pi_{2,k}(\tilde{h}_1), x)$ is a universally measurable function of $x$ and, by (4.25), converges pointwise to $\sup_{\pi_2 \in \Pi_2} J^*(\pi_1, \pi_2, x)$ as $k \to \infty$. \hfill \Box

**Proof of Prop. 4.6.** Similarly to the proof of Theorem 4.4, we first consider the partial maximization problem

$$f(x, p) := \sup_{p' \in \tilde{S}_M(x, p)} \phi(p'), \quad (x, p) \in \text{proj}_{\mathcal{X} \times \mathcal{P}(\Omega)}(\tilde{S}_M) = \tilde{S}_M,$$
where $\phi$, the function corresponding to $J^*$, is upper semianalytic by Assumption 4.2 and the set $\hat{S}_M$ is analytic by Prop. 4.3(i). Then by [5, Prop. 7.50], for each $\epsilon > 0$, there is a universally measurable mapping $\xi : \hat{S}_M \rightarrow \mathcal{P}(\Omega)$ with $\xi(x,p) \in \hat{S}_M(x,p)$ for all $(x,p) \in \hat{S}_M$ such that

$$\phi(\xi(x,p)) \geq \begin{cases} f(x,p) - \epsilon & \text{if } f(x,p) < +\infty, \\ 1/\epsilon & \text{if } f(x,p) = +\infty. \end{cases} \quad (4.26)$$

Let $\pi_1 \in \Pi_1$ be any policy of player 1. To prove the proposition, we will construct below a policy $\pi_2$ that satisfies the requirement (4.25) for $\pi_1$ and a given $\epsilon > 0$. To this end, first, let $\pi_2 \in \Pi_2$ be any policy of player 2 that never uses the information about $a_0^\dagger$, so that we can simply write its “shifted” policies $\pi_2^\dagger(i_1)$ as $\pi_2(i_1)$ instead, for $i_1 \in X \times A_1$. Note that the “shifted” policies $\pi_1^\dagger(i_1), \pi_2(i_1)$ comprise stochastic kernels that are universally measurable jointly in the variable $i_1$ and their respective observable variables (cf. (4.24)). Then, by using [5, Lem. 7.28, Prop. 7.46] and the Dynkin system theorem [5, Prop. 7.24], it is not hard to show that the mapping $\sigma$ defined as

$$\sigma : (x, \tilde{i}_1) \mapsto \rho_x[\pi_1^\dagger(\tilde{i}_1), \pi_2^\dagger(\tilde{i}_1)] \in \hat{S}_M(x)$$

is universally measurable from $X \times (X \times A_1)$ into $\hat{S}_M$.

Since the compositions of universally measurable mappings on Borel spaces are universally measurable [5, 7.44], it follows that the mapping $\zeta(x,i_1) := \xi(x, \sigma(x, i_1))$ is universally measurable from $X \times (X \times A_1)$ into $\mathcal{P}(\Omega)$. Since $\xi(x,p) \in \hat{S}_M(x,p) \subset \hat{S}_M(x)$ for all $(x,p) \in \hat{S}_M$, we have $\zeta(x,i_1) \in \hat{S}_M(x)$ for all $x \in X$ and $i_1 \in X \times A_1$. By applying Prop. 4.5 to this function $\zeta$ with $\theta = i_1$ and $\Theta = X \times A_1$, we obtain policies $\pi_1^\dagger(i_1), \pi_2(i_1)$ that are parametrized by $i_1$ and have those properties given in Prop. 4.5(i)-(ii). Let us write $\pi_2(i_1)$ as $\pi_2(i_1) = \{\mu_n^\dagger[i_1]\}_{n \geq 0}$ and define a policy $\pi_2 := \{\mu_n^\dagger\}_{n \geq 0}$ for player 2 by letting $\mu_n^\dagger$ be any analytically measurable selection of $\Gamma_2$ and letting

$$\mu_n^\dagger(\cdot | h_n) := \mu_{n-1}[x_0, a_0^\dagger(\cdot | x_1, a_1, \ldots, x_n)], \quad h_n \in H_n, \ n \geq 1.$$  

Since $\pi_2(i_1)$ has the property given in Prop. 4.5(ii), $\mu_n^\dagger$, $n \geq 1$, are universally measurable stochastic kernels, so $\pi_2$ as defined is a valid policy in $\Pi_2$. Note that $\pi_2$ never uses the information about $a_0^\dagger$ and denoting its “shifted” policies by $\pi_2^\dagger(x_0, a_0^\dagger)$, the above definition amounts to setting $\pi_2^\dagger(x_0, a_0^\dagger) = \pi_2(x_0, a_0^\dagger)$ for all $(x_0, a_0^\dagger) \in X \times A_1$.

Now let us show that the policy $\pi_2$ satisfies the requirement (4.25). For each $\tilde{i}_1 \in X \times A_1$, using the property of $\pi_1(i_1)$ and $\pi_2(i_1)$ given by Prop. 4.5(i), we have

$$\rho_x[\pi_1^\dagger(i_1), \pi_2^\dagger(i_1)] = \rho_x[\pi_1^\dagger(i_1), \pi_2^\dagger(i_1)] = \zeta(x, \tilde{i}_1) = \xi(x, \sigma(x, i_1)) \in \hat{S}_M(x, \sigma(x, \tilde{i}_1)). \quad (4.27)$$

By Prop. 4.3(ii),

$$\hat{S}_M(x, \sigma(x, \tilde{i}_1)) = \{ \rho_x[\pi_1^\dagger(i_1), \pi_2^\dagger] | \pi_2^\dagger \in \Pi_2 \}. \quad (4.28)$$

Thus $\rho_x[\pi_1^\dagger(i_1), \pi_2^\dagger] \in \{ \rho_x[\pi_1^\dagger(i_1), \pi_2^\dagger] | \pi_2^\dagger \in \Pi_2 \}$, implying $\rho_x[\pi_1^\dagger(i_1), \pi_2^\dagger(i_1)] = \rho_x[\pi_1^\dagger(i_1), \pi_2^\dagger(i_1)]$ by the definition of $\rho_x[\cdot, \cdot]$. Using Assumption 4.2 and (4.26)-(4.27), we then have that

$$J^*(\pi_1^\dagger(i_1), \pi_2^\dagger(i_1), x) = \phi(\rho_x[\pi_1^\dagger(i_1), \pi_2^\dagger(i_1)]) \leq \begin{cases} f(x, \sigma(x, \tilde{i}_1)) - \epsilon & \text{if } f(x, \sigma(x, \tilde{i}_1)) < +\infty, \\ 1/\epsilon & \text{if } f(x, \sigma(x, \tilde{i}_1)) = +\infty. \end{cases} \quad (4.29)$$

By the definition of $J^*$, (4.28) and Prop. 4.3(ii), $f(x, \sigma(x, i_1)) = \sup_{\pi_2 \in \Pi_2} J^*(\pi_1^\dagger(i_1), \pi_2, x)$. This together with (4.29) establishes the desired relation (4.25) for $\pi_2$. The proof is now complete. \qed
We now use the preceding results to derive optimality equations or inequalities for several risk-neutral criteria. In the two examples below we assume the axiom of analytic determinacy and Assumption 4.1. For simplicity, we also assume that \( c(\cdot) \) is a bounded upper semianalytic function.

**Example 4.3 (Optimality equations).** First, consider the average cost criterion \( \tilde{J}^{(1)} \) given in (4.6). Since \( c(\cdot) \) is assumed to be bounded, for any \( \pi_1 \in \Pi_1, \pi_2 \in \Pi_2, x \in \mathcal{X} \), we have \( \tilde{J}^{(1)}(\pi_1, \pi_2, x) := \mathbb{E}^{\pi_1, \pi_2} \left[ \limsup_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} c(x_k, a_k) \right] = \mathbb{E}^{\pi_1, \pi_2} \left[ \limsup_{n \to \infty} n^{-1} \sum_{k=1}^{\infty} c(x_k, a_k) \right] \). Inside the last term above, take conditional expectation given \((x_0, a_0, x_1)\) to obtain

\[
\tilde{J}^{(1)}(\pi_1, \pi_2, x) = \mathbb{E}^{\pi_1, \pi_2} \left[ \tilde{J}^{(1)}(\pi_1^s(x_0, a_0^1), \pi_2^s(x_0, a_0), x_1) \right].
\]

(4.30)

Now for \( \epsilon > 0 \), let \( \pi_1^\epsilon \in \Pi_1 \) be an \( \epsilon \)-optimal policy of player 1 given by Theorem 4.4. Correspondingly, let \( \pi_2^\epsilon \in \Pi_2 \) be a policy of player 2 given by Prop. 4.6 for \( \pi_1 = \pi_1^\epsilon \), so that we have

\[
\tilde{J}^{(1)}(\pi_1^s(x_0, a_0^1), \pi_2^s(x_0, a_0), x_1) \geq \sup_{\pi_2^\epsilon \in \Pi_2} \tilde{J}^{(1)}(\pi_1^s(x_0, a_0^1), \pi_2^\epsilon, x_1) - \epsilon \geq g^\epsilon(x_1) - \epsilon
\]

(4.31)

for all \((x_0, a_0, x_1) \in \mathcal{X} \times \mathbb{A} \times \mathcal{X} \). Then by the \( \epsilon \)-optimality of \( \pi_1^\epsilon \) and (4.30)-(4.31), for all \( x \in \mathcal{X} \),

\[
g^\epsilon(x) + \epsilon \geq \tilde{J}^{(1)}(\pi_1^\epsilon, \pi_2^\epsilon, x) = \mathbb{E}^{\pi_1^\epsilon, \pi_2^\epsilon} \left[ \tilde{J}^{(1)}(\pi_1^s(x_0, a_0^1), \pi_2^s(x_0, a_0), x_1) \right] \geq \mathbb{E}^{\pi_1^\epsilon, \pi_2^\epsilon} \left[ g^\epsilon(x_1) \right] - \epsilon.
\]

Since \( \epsilon \) is arbitrary and the initial decision rule of \( \pi_2^\epsilon \) can be chosen arbitrarily, this implies

\[
g^\epsilon(x) \geq \inf_{\nu \in \mathcal{P}(\mathcal{A}_1(x))} \sup_{a^2 \in \mathcal{A}_2(x)} \int_{\mathcal{A}_1} \int_{\mathcal{X}} g^\epsilon(y) q(dy \mid x, a^1, a^2) \nu(da^1 \mid x).
\]

(4.32)

To obtain the opposite inequality, let \( \pi_1 \) be a policy of player 1 that applies the policy \( \pi_1^\epsilon \) from time 1 onwards; in other words, its shifted policies \( \pi_1^\epsilon(x_0, a_0^1) \) all coincide with \( \pi_1^s \). Then by (4.30) and the \( \epsilon \)-optimality of \( \pi_1^\epsilon \), we have that

\[
g^\epsilon(x) \leq \sup_{\pi_2 \in \Pi_2} \tilde{J}^{(1)}(\pi_1, \pi_2, x) \leq \sup_{\pi_2^\epsilon \in \Pi_2} \mathbb{E}^{\pi_1^\epsilon, \pi_2^\epsilon} \left[ g^\epsilon(x_1) \right] + \epsilon, \quad \forall x \in \mathcal{X}.
\]

(4.33)

Since \( \epsilon \) is arbitrary and the initial decision rule of \( \pi_1 \) can be chosen arbitrarily, we obtain the opposite inequality to (4.32) and hence the optimality equation

\[
g^\epsilon(x) = \inf_{\nu \in \mathcal{P}(\mathcal{A}_1(x))} \sup_{a^2 \in \mathcal{A}_2(x)} \int_{\mathcal{A}_1} \int_{\mathcal{X}} g^\epsilon(y) q(dy \mid x, a^1, a^2) \nu(da^1 \mid x), \quad \forall x \in \mathcal{X}.
\]

(4.34)

Similarly, (4.34) also holds for the average cost criteria \( \tilde{J}^{(i)} \), \( 2 \leq i \leq 4 \), discussed in Section 4.1.1.

As another example, for \( \beta \in [0, 1) \), consider the \( \beta \)-discounted cost criterion \( J^\beta(\pi_1, \pi_2, x) := \mathbb{E}^{\pi_1, \pi_2} \left[ \sum_{k=0}^\infty \beta^k c(x_k, a_k) \right] \) (which satisfies Assumption 4.2 for upper semianalytic \( c(\cdot) \)). Denote the optimal value function in the minimax control problem by \( V^\beta \). By arguments similar to the above, we have the optimality equation

\[
V^\beta(x) = \inf_{\nu \in \mathcal{P}(\mathcal{A}_1(x))} \sup_{a^2 \in \mathcal{A}_2(x)} \int_{\mathcal{A}_1} \int_{\mathcal{X}} \left\{ c(x, a^1, a^2) + \beta V^\beta(y) \right\} q(dy \mid x, a^1, a^2) \nu(da^1 \mid x), \quad x \in \mathcal{X}
\]

(4.35)

and also the optimality equations for the finite-horizon setting. From these equations it follows easily that within the class of bounded universally measurable functions on \( \mathcal{X} \), \( V^\beta \) is the unique solution to (4.35), and that starting from the constant zero function, value iteration (with the operator defined by the right-hand side of (4.35)) converges to \( V^\beta \) uniformly over \( \mathcal{X} \).
Example 4.4 (Optimality inequality). Consider the standard average cost criterion \( J^{(1)} \) (cf. (4.4)). For bounded \( c(\cdot) \), using Fatou’s lemma, we have that for any \( \pi_1 \in \Pi_1, \pi_2 \in \Pi_2, x \in \mathcal{X} \),

\[
J^{(1)}(\pi_1, \pi_2, x) := \lim_{n \to \infty} \sup_{n} n^{-1} E_{x}^{\pi_1, \pi_2} \left[ \sum_{k=0}^{n-1} c(x_k, a_k) \right] \leq E_{x}^{\pi_1, \pi_2} \left[ J^{(1)}(\pi_1^{s}(x_0, a_0^{1}), \pi_2^{s}(x_0, a_0), x_1) \right].
\]

Applying the argument used to derive (4.33) in the previous example with the above inequality in place of (4.30), we obtain the optimality inequality

\[
g^*(x) \leq \inf_{\nu \in \mathcal{P}(\mathcal{A}_1(x))} \sup_{a \in \mathcal{A}_2(x)} \int_{\mathcal{X}} \int_{\mathcal{X}} g^*(y) q(dy \mid x, a^1, a^2) \nu(da^1 \mid x), \quad \forall x \in \mathcal{X}. \tag{4.36}
\]

If \( q(dy \mid x, a^1) \) only so that player 2 cannot affect the state transitions, (4.36) is simplified to \( g^*(x) \leq \inf_{a \in \mathcal{A}_1(x)} \int_{\mathcal{X}} g^*(y) q(dy \mid x, a^1) \) for all \( x \in \mathcal{X} \). In this special case one can further analyze the structure of \( g^* \) by using the submartingale property of the process \( \{g^*(x_n)\}_{n \geq 0} \) under any policy pair \( (\pi_1, \pi_2) \), similarly to our previous work [55] on average-cost MDPs.

5 Proofs

This section collects some of our proofs for Sections 3-4. To analyze strategic measures, we will use an approach discussed in Feinberg [21, Remark 3.3], which is to work with particular versions of conditional probability distributions that are jointly measurable as functions of both the conditioning variables and the probability measures themselves. According to [21], this approach was suggested by Maitra and Sudderth. To prepare the ground for the analysis, we first introduce some definitions and lemmas in Section 5.1. We then give the proofs for Sections 3 and 4 in Sections 5.2 and 5.3, respectively.

5.1 Some Definitions and Useful Lemmas

First, let us specify several stochastic kernels that will be used in our proofs to characterize the various sets of strategic measures. Recall that for \( n \geq 0 \), \( h_n := (x_0, a_0, x_1, a_1, \ldots, x_n) \) and \( h'_n := (x_0, a_0, \ldots, x_n, a_n) \); their spaces are denoted by \( H_n \) and \( H'_n \), respectively.

- Let \( \nu_n(da_n \mid h_n; p) \) be a Borel measurable stochastic kernel on \( \mathcal{A} \) given \( H_n \times \mathcal{P}(\Omega) \) such that for each \( p \in \mathcal{P}(\Omega) \), it is a conditional distribution of \( a_n \) given \( h_n \) w.r.t. \( p \).
- Let \( Q_n(dx_{n+1} \mid h'_n; p) \) be a Borel measurable stochastic kernel on \( \mathcal{X} \) given \( H'_n \times \mathcal{P}(\Omega) \) such that for each \( p \in \mathcal{P}(\Omega) \), it is a conditional distribution of \( x_{n+1} \) given \( h'_n \) w.r.t. \( p \).
- Let \( \kappa(da \mid x; \gamma) \) be a Borel measurable stochastic kernel on \( \mathcal{A} \) given \( \mathcal{X} \) and \( \mathcal{P}(\mathcal{X} \times \mathcal{A}) \) such that for each \( \gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{A}) \), it is a conditional distribution of \( a \) given \( x \) w.r.t. \( \gamma \).

The existence of these stochastic kernels follows from [5, Cor. 7.27.1], which “decomposes” a Borel measurable stochastic kernel on a product space into the product of two Borel measurable stochastic kernels. Specifically, by [5, Cor. 7.27.1], if \( X, Y, \) and \( Z \) are Borel spaces and \( \psi(d(y, z) \mid x) \) is a Borel measurable stochastic kernel on \( Y \times Z \) given \( X \), then there exist Borel measurable stochastic kernels \( \psi_1(dy \mid x) \) and \( \psi_2(dz \mid x, y) \) such that for all \( x \in X \),

\[
\psi(B \mid x) = \int_{Y} \int_{Z} 1_{B}(y, z) \psi_2(dz \mid x, y) \psi_1(dy \mid x), \quad B \in \mathcal{B}(Y \times Z).
\]

To obtain \( \nu_n(da_n \mid h_n; p) \) (resp. \( Q_n(dx_{n+1} \mid h'_n; p) \)), we apply [5, Cor. 7.27.1] with \( X = \mathcal{P}(\Omega) \) and with \( \psi \) being the mapping that maps \( p \in \mathcal{P}(\Omega) \) to its marginal on the space \( H_n \) (resp. \( H_{n+1} \)), which is a Borel measurable stochastic kernel on \( H'_n = H_n \times \mathcal{A} \) (resp. \( H_{n+1} = H'_n \times \mathcal{X} \)) given \( \mathcal{P}(\Omega) \). To
Proof of Lemma 5.1. Let $p \in \mathcal{P}(\Omega)$ and $\pi := \{\mu_n\}_{n \geq 0} \in \Pi$, and let $p_0$ be the marginal distribution of $x_0$ w.r.t. $p$. Then $p = \rho_{p_0}[\pi]$ if and only if for all $n \geq 0$,

$$Q_n(dx_{n+1} \mid h_n'; p) = q(dx_{n+1} \mid x_n, a_n) \quad \text{and} \quad \nu_n(da_n \mid h_n; p) = \mu_n(da_n \mid h_n), \quad p\text{-almost surely.}$$

Let us clarify that in (5.1) the random variables are $h_n', (x_n, a_n)$, and $h_n$. By (5.1) we mean that for $p$-almost all $\omega \in \Omega$, these two probability measures on the space of $x_{n+1}$ are equal:

$$Q_n(\cdot \mid h_n'(\omega); p) = q(\cdot \mid x_n(\omega), a_n(\omega)),$$

and so are these two probability measures on the space of $a_n$:

$$\nu_n(\cdot \mid h_n(\omega); p) = \mu_n(\cdot \mid h_n(\omega)).$$

Similar notations will be used for other stochastic kernels involved in the proofs.

**Proof of Lemma 5.1.** If $p = \rho_{p_0}[\pi]$, then clearly (5.1) holds for all $n \geq 0$. Conversely, using (5.1), it can be verified by induction on $n$ that the marginals of $p$ and $\mathbb{P}^n_{p_0}$ on $\mathcal{B}(H_n')$ are the same for all $n \geq 0$. Since $\cup_{n \geq 0} \mathcal{B}(H_n')$ is an algebra generating $\mathcal{B}(\Omega)$, it follows that $p = \rho_{p_0}[\pi]$. \hfill $\square$

The next lemma will be used in proving the measurability of the sets of strategic measures induced by policies with various structures. We need some notation and definitions before stating the lemma. If $X$ and $Y$ are Borel spaces, let $\mathcal{P}(X) \otimes \mathcal{P}(Y)$ be the set of all product probability measures on $\mathcal{B}(X \times Y)$ (that is, $\mathcal{P}(X) \otimes \mathcal{P}(Y) := \{\nu \times \nu' \mid \nu \in \mathcal{P}(X), \nu' \in \mathcal{P}(Y)\}$, where $\nu \times \nu'$ is the product measure associated with $\nu$ and $\nu'$ on the product $\sigma$-algebra $\mathcal{B}(X) \otimes \mathcal{B}(Y) = \mathcal{B}(X \times Y)$. Recall also that $\mathcal{D}(X)$ is the set of all Dirac measures on $X$. We write $\omega \in \Omega$ in terms of its components also as $\omega = (\omega^1, \omega^2, \ldots)$. If $I$ is a subset of positive integers, let $\omega^{(I)}$ consist of those $i$th coordinates of $\omega$ for $i \in I$ (i.e., $\omega^{(I)} := \{\omega^i\}_{i \in I}$), and let $\Omega^{(I)}$ be the space of $\omega^{(I)}$.

**Lemma 5.2.** Let $Z = \Omega^{(I)}$, $Z' = \Omega^{(I')}$, $Z'_1 = \Omega^{(I'_1)}$, and $Z'_2 = \Omega^{(I'_2)}$, where $I$ and $I'$ are two disjoint subsets of positive integers, and $I'_1$ and $I'_2$ form a partition of $I'$. Let $\Psi_1(dx' \mid z; p)$ and $\Psi_2(dx' \mid z; p)$ be Borel measurable stochastic kernels on $Z'$ given $Z \times \mathcal{P}(\Omega)$. Then these subsets of $\mathcal{P}(\Omega)$ are Borel:

(i) $B_1 := \{p \in \mathcal{P}(\Omega) \mid p(\Psi_1(\cdot \mid \omega^{(I)}); p) = \Psi_2(\cdot \mid \omega^{(I)}); p) = 1\}$;

(ii) $B_2 := \{p \in \mathcal{P}(\Omega) \mid p(\Psi_1(\cdot \mid \omega^{(I')}); p) = \mathcal{D}(Z') = 1\}$;

(iii) $B_3 := \{p \in \mathcal{P}(\Omega) \mid p(\Psi_1(\cdot \mid \omega^{(I'_1)}); p) \in \mathcal{P}(Z'_1) \otimes \mathcal{P}(Z'_1) = 1\}$.

Suppose that $\Theta$ is a Borel space and $\Psi_1$ and $\Psi_2$ depend also on a parameter $\theta \in \Theta$, that is, $\Psi_1(dx' \mid z; p, \theta)$ and $\Psi_2(dx' \mid z; p, \theta)$ are two Borel measurable stochastic kernels on $Z'$ given $Z \times \mathcal{P}(\Omega) \times \Theta$. Then the following subset of $\mathcal{P}(\Omega) \times \Theta$ is Borel:

(iv) $B_4 := \{(p, \theta) \in \mathcal{P}(\Omega) \times \Theta \mid p(\Psi_1(\cdot \mid \omega^{(I)}); p, \theta) = \Psi_2(\cdot \mid \omega^{(I')}; p, \theta) = 1\}$.

**Proof.** Since $B_1$ is a special case of $B_4$ with $\Theta = \emptyset$, to prove that these two sets are Borel, it suffices to prove that $B_4$ is Borel. Let $\{f_k\}_{k \geq 0}$ be a countable family of bounded Borel measurable functions on $Z'$ that form a measure determining class for $\mathcal{P}(Z')$. The set $E := \{(p, \theta, z) \in \mathcal{P}(\Omega) \times \Theta \times Z \mid \Psi_1(\cdot \mid z; p, \theta) = \Psi_2(\cdot \mid z; p, \theta)\}$ consists of those $(p, \theta, z)$ such that $\int_{Z'} f_k(z') \Psi_1(dx' \mid z; p, \theta) = \int_{Z'} f_k(z') \Psi_2(dx' \mid z; p, \theta)$ for all $k \geq 0$. These integrals are Borel measurable in $(p, \theta, z)$ by [5, Prop. 7.29] (since the functions $f_k$ and the stochastic kernels $\Psi_1, \Psi_2$ are Borel measurable). Therefore, the set $E$ is Borel. Let $E(p, \theta)$ denote the $(p, \theta)$-section of $E$. Then $\phi(p, \theta) := p(\omega^{(I)} \in \mathcal{P}(\Omega) \mid \phi(p, \theta) = 1)$ is a Borel measurable function of $(p, \theta)$ by [5, Cor. 7.26.1]. Since $B_4 = \{(p, \theta) \in \mathcal{P}(\Omega) \mid \phi(p, \theta) = 1\}$, $B_4$ is Borel and hence $B_1$ is also Borel, as discussed earlier.

Consider now $B_2$. Since $\mathcal{D}(Z')$ is the image of $Z'$ under the homeomorphism $z' \mapsto z'$, it is a Borel subset of $\mathcal{P}(Z')$ by [42, Sec. I.3, Cor. 3.3]. Let $E := \{(p, z) \in \mathcal{P}(\Omega) \times Z \mid \Psi_1(\cdot \mid z; p) \in \mathcal{D}(Z')\}$. Then
The preimage of the Borel set $\mathcal{D}(Z')$ under the Borel measurable mapping $(z, p) \mapsto \Psi_1(dz \mid z; p)$ and therefore, $E$ is Borel. The rest of the proof is similar to the above proof for $B_4$: with $E(p)$ being the $p$-section of $E$ and with $\phi(p) := p[\omega^{(1)} \in E(p)]$, we have that $\phi$ is Borel measurable by [5, Cor. 7.26.1], and hence the set $B_2 = \{p \in \mathcal{P}(\Omega) \mid \phi(p) = 1\}$ is Borel.

For $B_3$, note that $\mathcal{P}(Z_1') \times \mathcal{P}(Z_2')$ is the image of $\mathcal{P}(Z_1') \times \mathcal{P}(Z_2')$ under the homeomorphism $(\nu, \nu') \mapsto \nu \times \nu'$ from $\mathcal{P}(Z_1') \times \mathcal{P}(Z_2')$ into $\mathcal{P}(Z')$, and therefore, $\mathcal{P}(Z_1') \times \mathcal{P}(Z_2')$ is a Borel subset of $\mathcal{P}(Z')$ by [42, Sec. I.3, Cor. 3.3]. The rest of the proof is the same as that for $B_2$ with $\mathcal{P}(Z_1') \times \mathcal{P}(Z_2')$ in place of $\mathcal{D}(Z')$.

We will use Lemma 5.2(ii) to characterize strategic measures induced by nonrandomized policies. Blackwell [8] used this technique for a similar purpose and attributed it to Sudderth [50]. We will use Lemma 5.2(iii)-(iv) in the proofs for the minimax control problems studied in Section 4.

### 5.2 Proofs for Section 3

#### 5.2.1 Proof of Theorem 3.1

We will prove a more general result than Theorem 3.1. The result and the proof arguments will be applicable not only to MDPs but also to the partially observable problems and the minimax control problems studied in Section 4.

Any strategic measure $p \in \mathcal{S}$ must satisfy two conditions:

$\begin{align*}
  p\{(x_n, a_n) \in \Gamma\} &= 1, \quad \forall n \geq 0, \\
  Q_n(dx_{n+1} \mid h_n', p) &= q(dx_{n+1} \mid x_n, a_n), \quad \forall n \geq 0, \quad p\text{-almost surely},
\end{align*}$

in accordance with the control constraint and the state transition dynamics in the MDP.

Now consider the strategic measures of policies with additional structural constraints. In particular, suppose that for each $n \geq 0$, a subset of the variables from $(x_0, a_0, \ldots, x_{n-1}, a_{n-1})$ is chosen and, together with $x_n$, they form the variable $h_n$, whose space we denote by $H_n$. Suppose also that we have a collection of constraints on the stochastic kernels $\{\mu_n\}_{n \geq 0}$ that form a policy. These constraints, one for each $\mu_n$, consist of either requirement (i) or both requirements given below:

(i) $\mu_n$ is a universally measurable stochastic kernel on $\mathcal{A}$ given $H_n$;

(ii) $\mu_n$ is nonrandomized; that is, $\mu_n(da_n \mid h_n) \in \mathcal{D}(\mathcal{A})$ for all $h_n \in H_n$.

If $\mathcal{C}$ is such a collection of constraints, we write $\Pi_\mathcal{C}$ for the set of all policies that satisfy $\mathcal{C}$, and we write $\mathcal{S}_\mathcal{C}$ for the set of strategic measures induced by those policies. Note that $\Pi, \Pi', \Pi_{sm}, \Pi_{sm}'$, as well as $\Pi_n$ and $\Pi'_n$, can all be viewed as $\Pi_\mathcal{S}$ for some $\mathcal{C}$.

Let $\nu_n(da_n \mid h_n; p)$ be a Borel measurable stochastic kernel on $\mathcal{A}$ given $H_n \times \mathcal{P}(\Omega)$ such that for each $p \in \mathcal{P}(\Omega)$, it is a conditional distribution of $a_n$ given $h_n$ w.r.t. $p$. (As discussed at the start of Section 5.1, such a stochastic kernel exists.) A strategic measure $p \in \mathcal{S}_\mathcal{C}$ clearly satisfies that

$\nu_n(da_n \mid h_n; p) = \hat{\nu}_n(da_n \mid h_n; p)$

and

$\nu_n(da_n \mid h_n; p) \in \mathcal{D}(\mathcal{A}), \quad \forall n \geq 0, \quad p\text{-almost surely}$

where by “and $\mathcal{C}$” we mean that the condition “$\hat{\nu}_n(da_n \mid h_n; p) \in \mathcal{D}(\mathcal{A})$” is present if and only if according to $\mathcal{C}$, $\mu_n$ is nonrandomized.

**Lemma 5.3.** The set $\mathcal{S}_\mathcal{C} = \{p \in \mathcal{P}(\Omega) \mid p$ satisfies (5.2)-(5.4)\}.

**Proof.** As discussed above, every $p \in \mathcal{S}_\mathcal{C}$ satisfies (5.2)-(5.4). Consider now any $p \in \mathcal{P}(\Omega)$ that satisfies these constraints. By a repeated application of [5, Cor. 7.27.2] to decompose the marginals of $p$ on $H_n' \times H_{n+1}$, $n \geq 0$, and taking into account (5.3) and (5.4), we can represent $p$ as the composition of its marginal $p_0(dx_0)$ on $H_0$ with a sequence of Borel measurable stochastic kernels:

$p_0(dx_0), \ldots, \hat{\nu}(da_0 \mid h_0), q(dx_1 \mid x_0, a_0), \ldots, \hat{\nu}(da_n \mid h_n), q(dx_{n+1} \mid x_n, a_n), \ldots$
where \( \hat{\nu}_n(da_n \mid \hat{h}_n) = \hat{\nu}_n(da_n \mid \hat{h}_n; p) \). (In other words, \( p \) coincides with the unique probability measure on \( \mathcal{B}(\Omega) \) determined by the above sequence.) The stochastic kernels \( \{\hat{\nu}_n\} \) need not satisfy the control constraint in the MDP; i.e., the sets \( D_n := \{\hat{h}_n \in \hat{H}_n \mid \hat{\nu}_n(A(x_n) \mid \hat{h}_n) < 1\} \), \( n \geq 0 \), need not be empty. They also need not fully satisfy the structural constraints specified by \( C \). In particular, the following sets \( E_n, n \geq 0 \), need not be empty: \( E_n := \{\hat{h}_n \in \hat{H}_n \mid \hat{\nu}_n(\cdot \mid \hat{h}_n) \notin \mathcal{D}(\hat{\mathfrak{A}})\} \) if according to \( C, \mu_n \) must be nonrandomized; and \( E_n := \emptyset \) otherwise.

We now modify \( \hat{\nu}_n \) on \( D_n \cup E_n \) for each \( n \geq 0 \) to construct a policy in \( \Pi_c \). Since the stochastic kernel \( \hat{\nu}_n \) is Borel measurable and the graph \( \Gamma(\cdot) \) is analytic, by [5, Prop. 7.46] \( \hat{\nu}_n(A(x_n) \mid \hat{h}_n) = \int_\hat{\mathfrak{A}} \chi_{\Gamma}(x_n,a_n) \hat{\nu}_n(da_n \mid \hat{h}_n) \) is a universally measurable function of \( \hat{h}_n \), and hence the set \( D_n \) is universally measurable. The proof for Lemma 5.2(ii) shows that the set \( E_n \) is Borel measurable. Therefore, \( D_n \cup E_n \) is universally measurable. Since \( p \) satisfies (5.2) and (5.4), we must have

\[
p\{\hat{h}_n \in D_n \cup E_n\} = 0, \quad \forall n \geq 0.
\]

Now define a policy \( \pi := \{\mu_n\}_{n \geq 0} \in \Pi_c \) as follows: for each \( n \geq 0 \), let

\[
\mu_n(\cdot \mid \hat{h}_n) := \begin{cases} 
\hat{\nu}_n(\cdot \mid \hat{h}_n), & \text{if } \hat{h}_n \notin D_n \cup E_n; \\
\mu^o(\cdot \mid x_n), & \text{if } \hat{h}_n \in D_n \cup E_n,
\end{cases}
\]

where \( \mu^o \) is some fixed nonrandomized stationary policy in \( \Pi'_s \). In view of (5.4)-(5.5), \( p \)-almost surely, \( \mu_n(\cdot \mid \hat{h}_n) = \nu_n(\cdot \mid \hat{h}_n; p) \) for all \( n \geq 0 \). Then by Lemma 5.1 \( p = \rho_{p_0}[\pi] \) and hence \( p \in S_c \). \( \square \)

**Proposition 5.1.** The set \( S_c \) is analytic, and it is Borel if \( \Gamma \) is Borel.

**Proof.** For each \( n \geq 0 \), let \( E_n := \{p \in \mathcal{P}(\Omega) \mid p \text{ satisfies (5.2) for the given } n\} \), \( F_n := \{p \in \mathcal{P}(\Omega) \mid p \text{ satisfies (5.3) for the given } n\} \), and \( G_n := \{p \in \mathcal{P}(\Omega) \mid p \text{ satisfies (5.4) for the given } n\} \). The set \( F_n \) is Borel by Lemma 5.2(i), which is applied with \( \Psi_1 \) and \( \Psi_2 \) corresponding to \( Q_n(dx_{n+1} \mid h_n; p) \) and \( q(dx_{n+1} \mid x_n,a_n) \), respectively. The set \( G_n \) is Borel by Lemma 5.2(ii), which is applied with \( \Psi_1 \) and \( \Psi_2 \) corresponding to \( \hat{\nu}(da_n \mid \hat{h}_n; p) \) and \( \nu_n(da_n \mid h_n; p) \), respectively. Consider now the set \( E_n \). Condition (5.2) is the same as that \( p(D) = 1 \) for the set \( D := (\mathfrak{X} \times \hat{\mathfrak{A}})^n \times \Gamma \times (\mathfrak{X} \times \hat{\mathfrak{A}})^\infty \). Since \( \Gamma \) is analytic, \( D \) is an analytic subset of \( \Omega \) [5, Prop. 7.38]. Then, by [5, Prop. 7.43] \( E_n = \{p \in \mathcal{P}(\Omega) \mid p(D) = 1\} \) is analytic. If \( \Gamma \) is Borel, then \( D \) is Borel and hence by [5, Prop. 7.25] \( E_n \) is also Borel. Since \( S_c = \bigcap_{n \geq 0}(E_n \cap F_n \cap G_n) \) by Lemma 5.3, the desired conclusion follows. \( \square \)

For \( n \geq 0 \), let \( \gamma_n : \mathcal{P}(\Omega) \to \mathcal{P}(\mathfrak{X} \times \hat{\mathfrak{A}}) \) with \( \gamma_n(p) \) being the marginal of \( p \) on the space of \( (x_n,a_n) \). Recall the stochastic kernel \( \kappa \) defined in Section 5.1. In proving Lemma 5.3 and Prop. 5.1, we have proved the following corollary, which establishes Theorem 3.1 for \( S_\ast \in \{S,S_m\} \):

**Corollary 5.1.** The sets \( S, S', S_m, \) and \( S'_m \) can be expressed as:

(i) \( S = \{p \in \mathcal{P}(\Omega) \mid p \text{ satisfies (5.2)-(5.3)}\} \);

(ii) \( S_m = S \cap M, \) where \( M \) is the Borel set defined by

\[
M := \{p \in \mathcal{P}(\Omega) \mid p\{\nu_n(da_n \mid h_n; p) = \kappa(da_n \mid x_n; \gamma_n(p)), \forall n \geq 0\} = 1\};
\]

(iii) \( S' = S \cap M' \) and \( S'_m = S_m \cap M', \) where \( M' \) is the Borel set defined by

\[
M' := \{p \in \mathcal{P}(\Omega) \mid p\{\nu_n(da_n \mid h_n; p) \in \mathcal{D}(\hat{\mathfrak{A}}), \forall n \geq 0\} = 1\}.
\]

These sets of strategic measures are analytic; they are Borel if \( \Gamma \) is Borel.

We now consider the set \( S_\ast \). As in Feinberg [21, Proof of Lem. 3.1], we consider the mapping \( \hat{\gamma} : \mathcal{P}(\Omega) \to \mathcal{P}(\mathfrak{X} \times \hat{\mathfrak{A}}) \) defined by \( \hat{\gamma}(p) := \sum_{n=0}^{\infty} 2^{-n-1} \gamma_n(p) \). Since \( \gamma_n \) is Borel measurable for all \( n \geq 0 \), the mapping \( \hat{\gamma} \) is Borel measurable. Consequently, the mapping \( (x,p) \to \kappa(da_n \mid x; \hat{\gamma}(p)) \) is also Borel measurable (since the stochastic kernel \( \kappa \), defined in Section 5.1, is Borel measurable, and compositions of Borel mappings are Borel measurable); in other words, it is a Borel measurable stochastic kernel on \( \hat{\mathfrak{A}} \) given \( \mathfrak{X} \times \mathcal{P}(\Omega) \).
Lemma 5.4. We have \( S_s = S \cap \tilde{M} \) and \( S'_s = S_s \cap M' \), where the set
\[
\tilde{M} := \{ p \in \mathcal{P}(\Omega) \mid p(\nu_n(da_n \mid h_n; p) = \kappa(da_n \mid x_n; \hat{\gamma}(p)), \forall n \geq 0 \} = 1 \}
\]
is Borel, and \( M' \) is the Borel set given in Cor. 5.1. Both \( S_s \) and \( S'_s \) are analytic; they are Borel if \( \Gamma \) is Borel.

Proof. Let \( B_n := \{ p \in \mathcal{P}(\Omega) \mid p(\nu_n(da_n \mid h_n; p) = \kappa(da_n \mid x_n; \hat{\gamma}(p)) = 1 \} \) for all \( n \geq 0 \). Then the set \( \tilde{M} = \cap_{n \geq 0} B_n \). By applying Lemma 5.2(i) to each set \( B_n \), with \( \Psi_1 \) and \( \Psi_2 \) corresponding to \( \nu_n(da_n \mid h_n; p) \) and \( \kappa(da_n \mid x_n; \hat{\gamma}(p)) \), respectively, we obtain that \( B_n \) is Borel. Hence the set \( \tilde{M} \) is Borel.

Clearly \( S_s \subset S \cap \tilde{M} \) and \( S'_s \subset S \cap \tilde{M} \cap M' \). The proof of the opposite inclusions is similar to that of Lemma 5.3. If \( p \in S \cap \tilde{M} \), we define a stationary policy \( \mu \in \Pi_s \) by
\[
\mu(da \mid x) := \begin{cases} 
\kappa(da \mid x; \hat{\gamma}(p)) & \text{if } x \notin D; \\
\mu^o(da \mid x) & \text{if } x \in D,
\end{cases}
\]
where \( \mu^o \in \Pi_s \) and the set \( D \) is defined by \( D := \{ x \in \mathcal{X} \mid \kappa(A(x) \mid x; \hat{\gamma}(p)) < 1 \} \). If \( p \in S \cap \tilde{M} \cap M' \), then we define a nonrandomized stationary policy \( \mu \in \Pi'_s \) by (5.6) with \( \mu^o \in \Pi'_s \) and with the set \( D \cup E \) in place of \( D \), where the set \( E := \{ x \in \mathcal{X} \mid \kappa(da \mid x; \hat{\gamma}(p)) \notin D(\mathcal{X}) \} \). The set \( \tilde{M} \in \mathcal{U}(\mathcal{X}) \) by [5, Prop. 7.46] (because the mapping \( (x, p) \mapsto \kappa(da \mid x; \hat{\gamma}(p)) \) is Borel measurable and the graph \( \Gamma \) of \( A(\cdot) \) is analytic), and the set \( E \) is Borel by the proof of Lemma 5.2(ii). So \( \mu \) as defined is a valid universally measurable policy. In view of Cor. 5.1(i) and (5.2), if \( p \in S \cap \tilde{M} \), then \( p(x_n \in D, \forall n \geq 0) = 0 \), and if \( p \in S \cap \tilde{M} \cap M' \), then \( p(x_n \in D \cup E, \forall n \geq 0) = 0 \). It follows that in either case, \( p \)-almost surely, \( \mu(da_n \mid x_n) = \nu_n(da_n \mid h_n; p) \) for all \( n \geq 0 \), so by Lemma 5.1, \( p = \rho_0[p] \), where \( \rho_0 \) is the distribution of \( x_0 \) w.r.t. \( p \). This proves \( S_s = S \cap \tilde{M} \) and \( S'_s = S_s \cap M' \). The last statement of the lemma then follows from Cor. 5.1 and the Borel measurability of \( \tilde{M} \) and \( M' \).

Theorem 3.1 has now been proved.

5.2.2 Proof of Proposition 3.1

First, we prove a result that entails Prop. 3.1 except for the cases \( S_s \) and \( S'_s \). Consider again the set \( \Pi_C \) of structured policies and the associated set \( \tilde{S}_C \) of strategic measures introduced in Section 5.2.1. Let \( \tilde{S}_C := \{ (x, p) \in \mathcal{X} \times \mathcal{P}(\Omega) \mid p \in \tilde{S}_C, \rho_0(p) = \delta_x \} \) (recall that \( \rho_0(p) \) is the initial distribution of \( x_0 \) w.r.t. \( p \)). By Prop. 5.1 and the proof of Lemma 3.1, \( \tilde{S}_C \) is analytic, and it is Borel if \( \Gamma \) is Borel. Denote the \( x \)-section of \( \tilde{S}_C \) by \( \tilde{S}_C(x) \).

Also, corresponding to \( C \) and the associated spaces \( \tilde{H}_n \), \( n \geq 0 \), we derive a collection \( C_s \) of constraints by replacing the condition that \( \mu_n \) depends on \( h_n \in \tilde{H}_n \) by the condition that \( \mu_n \) depends on both \( h_n \) and the initial state \( x_0 \), if \( h_n \) does not already include \( x_0 \). The collection of variables in \( h_n \) and \( \{ x_0 \} \) is denoted by \( h_n^* \), and their space is denoted by \( \tilde{H}_n^* \). The set \( \Pi_{C_s} \) is the set of all policies satisfying the constraints in \( C_s \).

Proposition 5.2. Let \( \zeta : \mathcal{X} \rightarrow \mathcal{P}(\Omega) \) be a universally measurable mapping such that \( \zeta(x) \in \tilde{S}_C(x) \) for all \( x \in \mathcal{X} \). Then there exists a policy \( \pi \in \Pi_{C_s} \) with \( \zeta(x) = \rho_x[\pi] \) for all \( x \in \mathcal{X} \).

Proof. The proof is similar to that of Lemma 5.3, except that here the parametric dependence on the initial state \( x \) needs to be taken into account in the construction of the policy \( \pi \).

To simplify notation, let \( \nu_n(da_n \mid h_n; x) := \nu_n(da_n \mid h_n; \zeta(x)) \) for \( n \geq 0 \). Note that as a \( \mathcal{P}(\mathcal{A}) \)-valued function of \( (\hat{h}_n, x) \), \( \nu_n(da_n \mid \hat{h}_n; x) \) is the composition of the mapping \( (\hat{h}_n, x) \mapsto (\hat{h}_n, \zeta(x)) \) with the mapping \( (\hat{h}_n, p) \mapsto \nu_n(da_n \mid \hat{h}_n; p) \). By [5, Prop. 7.44] compositions of universally measurable
functions on Borel spaces are universally measurable. Since $\zeta$ is universally measurable and the stochastic kernel $\tilde{\nu}_n(da_n \mid \hat{h}_n; \mu)$ is Borel measurable, it then follows from [5, Prop. 7.44] that the stochastic kernel $\tilde{\nu}_n(da_n \mid \hat{h}_n; x)$ is universally measurable.

For each $n \geq 0$, let $D_n$ be the set of points $(\hat{h}_n, x)$ at which $\tilde{\nu}_n$ violates the control constraint:

$$D_n := \{ (\hat{h}_n, x) \in \hat{H}_n \times \mathcal{X} \mid \tilde{\nu}_n(A(x_n) \mid \hat{h}_n; x) < 1 \}.$$ 

If the structural constraints specified by $\mathcal{C}$ require $\mu_n$ to be nonrandomized, let

$$E_n := \{ (\hat{h}_n, x) \in \hat{H}_n \times \mathcal{X} \mid \tilde{\nu}_n(\cdot \mid \hat{h}_n; x) \notin \mathcal{D}(\mathcal{A}) \};$$

otherwise, let $E_n := \emptyset$. As discussed earlier, $\tilde{\nu}_n$ is universally measurable and the set $\mathcal{D}(\mathcal{A})$ is Borel, while the graph $\Gamma$ of $A(\cdot)$ is analytic by assumption. Consequently, the sets $D_n$ and $E_n$ are universally measurable by [5, Prop. 7.46] and by [5, Cor. 7.44.1], respectively. Moreover, since $\zeta(x) \in \mathcal{S}_C(x)$, $\zeta(x)\{x_0 = x\} = 1$ and $\zeta(x) \in \mathcal{S}_C$. Then by Lemma 5.3, (5.2), and (5.4), $\zeta(x)\{\tilde{\nu}_n(A(x_n) \mid \hat{h}_n; \zeta(x)) = 1 \} = 1$, and if $\mu_n$ must be nonrandomized according to $\mathcal{C}$, we also have $\zeta(x)\{\tilde{\nu}_n(\cdot \mid \hat{h}_n; \zeta(x)) \in \mathcal{D}(\mathcal{A}) \} = 1$. In view of the definition of $\tilde{\nu}_n$, these relations imply

$$\zeta(x)\{ (\hat{h}_n, x_0) \in D_n \cup E_n \} = 0, \quad \forall n \geq 0, \; x \in \mathcal{X}. \quad (5.7)$$

Now define a policy $\pi := \{ \mu_n \}_{n \geq 0} \in \Pi_\mathcal{C}$ as follows: for $n \geq 0$ and $\hat{h}_n \in \hat{H}_n^*$,

$$\mu_n(da_n \mid \hat{h}_n^*) := \begin{cases} \tilde{\nu}_n(da_n \mid \hat{h}_n; x_0), & \text{if } (\hat{h}_n, x_0) \notin D_n \cup E_n; \\ \mu^o(da_n \mid x_n), & \text{if } (\hat{h}_n, x_0) \in D_n \cup E_n, \end{cases} \quad (5.8)$$

where $\mu^o$ is some fixed policy in $\Pi^o$. Then by (5.4), (5.7), and the definition of $\tilde{\nu}_n$, for each $x \in \mathcal{X}$, it holds $\zeta(x)$-almost surely that $\mu_n(da_n \mid \hat{h}_n^*) = \nu_n(da_n \mid \hat{h}_n; \zeta(x))$ for all $n \geq 0$. By Lemma 5.1 this implies $\zeta(x) = \rho_c[\pi]$ for all $x \in \mathcal{X}$. So $\pi$ is the desired policy in $\Pi_\mathcal{C}$. \hfill $\square$

Applying Prop. 5.2 with $\mathcal{S}_C \in \{ \mathcal{S}, \mathcal{S}', \mathcal{S}_m, \mathcal{S}'_m \}$, we obtain the conclusions of Prop. 3.1 for those four cases.

The remaining two cases in Prop. 3.1 are $\mathcal{S}_* = \mathcal{S}$ and $\mathcal{S}_* = \mathcal{S}'$, with the graph of $\zeta$ contained in $\mathcal{S}_s$. For these two cases, we construct the desired semi-stationary policy from the stochastic kernel $\kappa(da \mid x'; \gamma(\zeta(x)))$, where the function $\gamma$ is as defined before Lemma 5.4. The construction is similar to those in the proofs of Lemma 5.4 and Prop. 5.2. In particular, let $\tilde{\kappa}(da \mid x'; x'') := \kappa(da \mid x'; \gamma(\zeta(x)))$. Since $\zeta$ is universally measurable and the mapping $\gamma$ and the stochastic kernel $\kappa$ are Borel measurable, the stochastic kernel $\tilde{\kappa}$ is universally measurable by [5, Prop. 7.44]. Let

$$D := \{ (x', x) \in \mathcal{X}^2 \mid \tilde{\kappa}(A(x') \mid x'; x) < 1 \}. \quad (5.9)$$

In the case $\mathcal{S}_* = \mathcal{S}_s$, let $E := \emptyset$, and in the case $\mathcal{S}_* = \mathcal{S}'_s$, let

$$E := \{ (x', x) \in \mathcal{X}^2 \mid \tilde{\kappa}(da \mid x'; x) \notin \mathcal{D}(\mathcal{A}) \}. \quad (5.10)$$

Then $D, E \in \mathcal{U}(\mathcal{X}^2)$ by [5, Prop. 7.46, Cor. 7.44.1] (since $\tilde{\kappa}$ is universally measurable, the graph $\Gamma$ of $A(\cdot)$ is analytic, and the set $\mathcal{D}(\mathcal{A})$ is Borel). Moreover, for all $x \in \mathcal{X}$, since $\zeta(x) \in \mathcal{S}_s(x)$, we have $\zeta(x)\{x_0 = x\} = 1$ and $\zeta(x)\{\tilde{\kappa}(A(x_n) \mid x_n; x) = 1, \\forall n \geq 0 \} = 1$ by (5.2) and Lemma 5.4. In the case $\zeta(x) \in \mathcal{S}_s(x)$, we also have $\zeta(x)\{\tilde{\kappa}(da_n \mid x_n; x) \in \mathcal{D}(\mathcal{A}), \forall n \geq 0 \} = 1$ by Lemma 5.4. Therefore,

$$\zeta(x)\{ (x_n, x_0) \in D \cup E \} = 0, \quad \forall n \geq 0, \; x \in \mathcal{X}. \quad (5.9)$$

Now define a semi-stationary policy $\mu$ as follows: for $(x_0, x') \in \mathcal{X}^2$,

$$\mu(da \mid x_0, x') := \begin{cases} \tilde{\kappa}(da \mid x'; x_0), & \text{if } (x', x_0) \notin D \cup E; \\ \mu^o(da \mid x'), & \text{if } (x', x_0) \in D \cup E, \end{cases} \quad (5.10)$$
for some fixed $\mu^c \in \Pi'$. In the case $S_* = S'_*$, this policy $\mu$ is nonrandomized. By (5.9), the definition of $\bar{\kappa}$, and Lemma 5.4, we have that for each $x \in X$, it holds $\zeta(x)$-almost surely that $\mu(da_n \mid x_0, x_n) = \nu_n(da_n \mid h_n; \zeta(x))$ for all $n \geq 0$. By Lemma 5.1 this implies $\zeta(x) = \rho_x[\mu]$ for all $x \in X$, so $\mu$ is the desired semi-stationary policy in the case $S_* = S_*$ or $S'_*$.

The proof of Prop. 3.1 is now complete.

### 5.2.3 Proof of Theorem 3.2

In what follows we prove a more general version of Theorem 3.2 with $\Pi$ replaced by the set $\Pi_C$ of structured policies introduced in Section 5.2.1. As mentioned earlier, our proof is similar to that of Feinberg [19, Thm. 1] and makes use of an important lemma in Gikhman and Skorokhod [24, Lem. 1.2], in addition to the properties of universally measurable sets and functions.

Let $\lambda$ denote the Lebesgue measure on $[0, 1]$.

**Lemma 5.5.** For each $\pi := \{\mu_n\}_{n \geq 0} \in \Pi_C$, there exist universally measurable functions $f_n : [0, 1] \times \hat{H}_n \to \mathbb{A}$, $n \geq 0$, such that for all $\hat{h}_n \in \hat{H}_n$,

$$f_n(\theta, \hat{h}_n) \in A(x_n) \quad \forall \theta \in [0, 1],$$

and $\mu_n(da_n \mid \hat{h}_n) = (\lambda \circ \phi^{-1}_{\hat{h}_n})(da_n)$, where $\lambda \circ \phi^{-1}_{\hat{h}_n}$ is the image measure of $\lambda$ under the mapping $\phi_{\hat{h}_n} : [0, 1] \to \mathbb{A}$ given by $\phi_{\hat{h}_n} := f_n(\cdot, \hat{h}_n)$.

**Proof.** For each $n \geq 0$, by [24, Lem. 1.2], there exists $B([0, 1]) \times \mathcal{U}(\hat{H}_n)$-measurable function $f_n : [0, 1] \times \hat{H}_n \to \mathbb{A}$ that meets the requirements except that the set $E := \{(\theta, \hat{h}_n) \in [0, 1] \times \hat{H}_n \mid f_n(\theta, \hat{h}_n) \notin A(x_n)\}$ need not be empty. We modify $f_n$ on $E$ to define another function $\tilde{f}_n$ as follows. Let $\tilde{f}_n := f_n$ on $E^c$ and $\tilde{f}_n(\theta, \hat{h}_n) := \mu^c(x_n)$ on $E$, for some analytically measurable selection $\mu^c$ of $\Gamma$. Then $\tilde{f}_n(\theta, \hat{h}_n) \in A(x_n)$ for all $\theta \in [0, 1], \hat{h}_n \in \hat{H}_n$. Let us show that $\tilde{f}_n$ satisfies the rest of the requirements, namely, the universal measurability of $\tilde{f}_n$ and the relation $\mu_n(da_n \mid \hat{h}_n) = (\lambda \circ \phi^{-1}_{\hat{h}_n})(da_n)$ for $\phi_{\hat{h}_n} := \tilde{f}_n(\cdot, \hat{h}_n)$.

The set $E^c$ can be equivalently expressed as

$$E^c = \{(\theta, \hat{h}_n) \in [0, 1] \times \hat{H}_n \mid (x_n, f_n(\theta, \hat{h}_n)) \in \Gamma\} = \psi^{-1}(\Gamma),$$

where $\psi$ denotes the mapping $(\theta, \hat{h}_n) \mapsto (x_n, f_n(\theta, \hat{h}_n))$. Since $\psi$ is universally measurable by the measurability of $f_n$ and the set $\Gamma$ is analytic by our model assumption, $E^c$ is universally measurable [5, Cor. 7.44.1]. This, together with the measurability of $f_n$, shows that $\tilde{f}_n$ is universally measurable.

For all $\hat{h}_n \in \hat{H}_n$, since $\mu_n(da_n \mid \hat{h}_n) = (\lambda \circ \phi^{-1}_{\hat{h}_n})(da_n)$, we have

$$\lambda(\phi^{-1}_{\hat{h}_n}(A(x_n))) = \mu_n(A(x_n) \mid \hat{h}_n) = 1.$$  \hspace{1cm} (5.11)

In the first equality, for the validity of the first equality, we also used the fact that the set $\phi^{-1}_{\hat{h}_n}(A(x_n))$ is universally measurable. To see this, note that by the measurability of $f_n$, $\phi_{\hat{h}_n}$ is universally measurable, and since $\Gamma$ is analytic, its $x_n$-section $A(x_n)$ is analytic by [5, Prop. 7.40]. Then by [5, Cor. 7.44.1] $\phi^{-1}_{\hat{h}_n}(A(x_n))$ is universally measurable. Now from (5.11) it follows that

$$\lambda(E_{\hat{h}_n}) = \lambda(\{\theta \in [0, 1] \mid f_n(\theta, \hat{h}_n) \notin A(x_n)\}) = 0,$$

where $E_{\hat{h}_n}$ denotes the $\hat{h}_n$-section of the set $E$. Since $\tilde{f}_n(\cdot, \hat{h}_n)$ can only differ from $f_n(\cdot, \hat{h}_n)$ on $E_{\hat{h}_n}$, this implies that

$$\lambda(\{\theta \in [0, 1] \mid \tilde{f}_n(\theta, \hat{h}_n) \in B\}) = \lambda(\{\theta \in [0, 1] \mid f_n(\theta, \hat{h}_n) \in B\}), \quad \forall B \in \mathcal{B}(\mathbb{A}),$$
or equivalently, expressed in terms of image measures, \( \lambda \circ \tilde{\phi}_{h_n}^{-1} = \lambda \circ \phi_{h_n}^{-1} \), where \( \tilde{\phi}_{h_n} := \tilde{f}_n(\cdot,\hat{h}_n) \).
Since \( \mu_n(da_n \mid \hat{h}_n) = (\lambda \circ \phi_{h_n}^{-1})(da_n) \), we have \( \mu_n(da_n \mid \hat{h}_n) = \lambda \circ \phi_{h_n}^{-1}(da_n) \). Thus the function \( \tilde{f}_n \) satisfies all the requirements.

For each \( \pi \in \Pi_C \), consider the family of functions \( \{f_n\} \) given in Lemma 5.5. Corresponding to each point \( \hat{\theta} := (\theta_0, \theta_1, \ldots) \in [0,1]^{\infty} \), the sequence of functions \( \tilde{f}(\tilde{\theta}) := \{f_n(\theta_n, \cdot)\}_{n \geq 0} \) defines a nonrandomized universally measurable policy in \( \Pi_C \). To simplify notation, we denote this policy also by \( \tilde{f}(\hat{\theta}) \).

**Lemma 5.6.** Let \( \pi \in \Pi_C \), and consider the associated family of nonrandomized policies \( \tilde{f}(\hat{\theta}) \in \Pi_C \), \( \hat{\theta} \in [0,1]^{\infty} \), given by Lemma 5.5. The mapping \( \psi : \mathcal{P}(\mathcal{X}) \times [0,1]^{\infty} \to \mathcal{P}(\Omega) \) with \( \psi(p_0, \hat{\theta}) := \rho_{p_0}[\tilde{f}(\hat{\theta})] \) is universally measurable.

**Proof.** By [5, Lem. 7.28(b)], it suffices to show that for each \( B \in \mathcal{B}(\Omega) \), \( \rho_{p_0}[\tilde{f}(\hat{\theta})](B) \) is universally measurable in \( (p_0, \hat{\theta}) \). In turn, since those Borel sets of \( \Omega \) with this property form a Dynkin system, by the Dynkin system theorem [5, Prop. 7.24], it suffices to consider the collection \( \cup_{n \geq 0} \mathcal{B}(\mathcal{H}^n) \) of sets (which correspond to those cylindrical sets that generate \( \mathcal{B}(\Omega) \)) and to show that for each \( B_n \in \mathcal{B}(\mathcal{H}^n) \), \( \rho_{p_0}[\tilde{f}(\hat{\theta})](\mathcal{H}^n \cap B_n) \) is universally measurable in \( (p_0, \hat{\theta}) \). Now by [5, Prop. 7.45] we can express \( \rho_{p_0}[f(\hat{\theta})](\mathcal{H}^n \cap B_n) \) as the iterated integral

\[
\int_{\mathcal{X}} \cdots \int_{\mathcal{X}} 1_{B_n}(h_n, f_n(\theta_n, \hat{h}_n)) q(dx_n \mid x_{n-1}, f_{n-1}(\theta_{n-1}, \hat{h}_{n-1})) \cdots q(dx_1 \mid x_0, f_0(\theta_0, x_0)) p_0(dx_0).
\]

The functions \( f_k, k \leq n \), are universally measurable by Lemma 5.5, the stochastic kernel \( q \) is Borel measurable by our model assumption, and the identity mapping on \( \mathcal{P}(\mathcal{X}) \), \( p_0 \mapsto p_0 \), is also Borel measurable. It then follows from [5, Cor. 7.44.3, Prop. 7.46] that the above integral is universally measurable in \( (p_0, \theta_0, \theta_1, \ldots, \theta_n) \) and therefore universally measurable in \( (p_0, \hat{\theta}) \).

Recall that \( \bar{\lambda} \) is the countable product of Lebesgue measures on \([0,1]\).

**Lemma 5.7.** Let \( \pi \in \Pi_C \) and \( \tilde{f}(\hat{\theta}) \in \Pi_C, \hat{\theta} \in [0,1]^{\infty} \), be as in Lemma 5.6. Then for all \( p_0 \in \mathcal{P}(\mathcal{X}) \),

\[
\rho_{p_0}[\pi](B) = \int_{[0,1]^{\infty}} \rho_{p_0}[(\tilde{f}(\hat{\theta}))(B)] \bar{\lambda}(d\hat{\theta}), \quad \forall B \in \mathcal{B}(\Omega).
\]

**Proof.** By Lemma 5.6 and [5, Lem. 7.28(b) or Prop. 7.46], the integral on the right-hand side of (5.13) is well-defined for each \( B \in \mathcal{B}(\Omega) \). Thus the right-hand side of (5.13) defines a probability measure on \( \mathcal{B}(\Omega) \). To show that this measure coincides with \( \rho_{p_0}[\pi] \), it suffices to show that (5.13) holds for all subsets of \( \Omega \) of the form \( \{h'_n \in B_n\} \) with \( B_n \in \mathcal{B}(\mathcal{H}^n) \) for some \( n \geq 0 \) (since these subsets form an algebra that generates \( \mathcal{B}(\Omega) \)). This can be verified directly by using the iterated-integral expression (5.12) for \( \rho_{p_0}[f(\hat{\theta})](\mathcal{H}^n \cap B_n) \), and by successively integrating out the variables \( \theta_0, \theta_1, \ldots, \theta_n \) and applying the Fubini theorem together with the relations \( \mu_k(da_k \mid \hat{h}_k) = (\lambda \circ \phi_{h_k}^{-1})(da_k) \), \( k \leq n \), provided by Lemma 5.5.

Note that the relation (5.13) is equivalent to (3.1) in Theorem 3.2(ii) (since a probability measure on \( \mathcal{U}(\Omega) \) is uniquely determined by its restriction on \( \mathcal{B}(\Omega) \)). With Lemmas 5.6-5.7, we have thus proved Theorem 3.2(i)-(ii) with \( \Pi_C \) in place of \( \Pi \). Letting \( \Pi_C = \Pi \) yields Theorem 3.2(i)-(ii), and letting \( \Pi_C = \Pi_m \) or \( \Pi_{sm} \) yields the remaining assertions of Theorem 3.2 regarding Markov or semi-Markov policies.
5.3 Proofs for Section 4

5.3.1 Proofs of Proposition 4.1 and Theorem 4.3

In this subsection we consider the partially observable control problem defined in Section 4.1.3 and prove Prop. 4.1 and Theorem 4.3. Because the proofs are almost identical to those given earlier for MDPs, we will only outline the main arguments.

For the partially observable problem, let $h_n := (x_0, z_0, a_0, \ldots, x_n, z_n)$, $h_n := (x_0, z_0, a_0, \ldots, x_n, z_n)$, and $h'_n := (x_0, z_0, a_0, \ldots, x_n, z_n, a_n)$ for $n \geq 0$, and denote their spaces by $H_n, H_n,$ and $H'_n$, respectively. For each $n \geq 0$, the stochastic kernels $\nu_n(da_n | h_n; p)$ and $Q_n(dx_{n+1} | h'_n; p)$ are as defined in Section 5.1 with the spaces $\Omega$, $H_n$, and $H'_n$ corresponding to those in the partially observable problem. In addition:

- Let $O_n(dz_n | h_n'; p)$ be a Borel measurable stochastic kernel on $\mathbb{Z}$ given $H_n' \times \mathcal{P}(\Omega)$ such that for each $p \in \mathcal{P}(\Omega)$, it is a conditional distribution of $z_n$ given $h_n'$ w.r.t. $p$.
- Let $\tilde{\nu}_n(da_n | i_n; p)$ be a Borel measurable stochastic kernel on $\mathbb{A}$ given $I_n \times \mathcal{P}(\Omega)$ such that for each $p \in \mathcal{P}(\Omega)$, it is a conditional distribution of $a_n$ given $i_n$ w.r.t. $p$.

The existence of these stochastic kernels is ensured by [5, Cor. 7.27.1], as explained in Section 5.1.

For $p \in \mathcal{P}(\Omega)$, consider these constraints:

\[
\begin{align*}
p\{ (i_n, a_n) \in \Gamma_n \} &= 1, \quad \forall n \geq 0; \quad (5.14) \\
Q_n(dx_{n+1} | h'_n; p) &= q(dx_{n+1} | x_n, a_n), \quad \forall n \geq 0, \quad p\text{-almost surely}; \quad (5.15) \\
O_n(dz_n | h_n'; p) &= \delta_f(z_n)(dz_n), \quad \forall n \geq 0, \quad p\text{-almost surely}; \quad (5.16) \\
\nu_n(da_n | h_n; p) &= \tilde{\nu}_n(da_n | i_n; p), \quad \forall n \geq 0, \quad p\text{-almost surely}; \quad (5.17) \\
\nu_n(da_n | i_n; p) &\in \mathcal{D}(\mathbb{A}), \quad \forall n \geq 0, \quad p\text{-almost surely}. \quad (5.18)
\end{align*}
\]

Clearly, any strategic measure $p \in S_I$ satisfies (5.14)-(5.17), and any $p \in S'_I$, being induced by a nonrandomized policy, satisfies (5.14)-(5.18). The lemma below follows from essentially the same line of argument for Lemma 5.3.

**Lemma 5.8.** The set $S_I = \{ p \in \mathcal{P}(\Omega) \mid p$ satisfies (5.14)-(5.17) $\}$. The set $S'_I = \{ p \in \mathcal{P}(\Omega) \mid p$ satisfies (5.14)-(5.18) $\} = S_I \cap M'$, where the set $M'$ is as given in Cor. 5.1.

Using Lemma 5.8 and the same reasoning given in the proofs of Prop. 5.1, Cor. 5.1, and Lemma 3.1, we obtain that the sets $S_I, S'_I, S_I'$, and $S'_I$ are analytic, and that they are Borel if $\Gamma_n, n \geq 0$, are Borel. This establishes Prop. 4.1(i). Proposition 4.1(ii) then follows from Lemma 5.8 and essentially the same proof arguments for Prop. 5.2. Finally, similarly to the proof of Theorem 3.3 or 4.1, Theorem 4.3 is a consequence of Prop. 4.1, Lemmas 4.1, 4.3, and [5, Props. 7.47, 7.50].

5.3.2 Proofs of Propositions 4.2-4.3 and 4.5

In order to reuse some of the results in Section 5.2 and avoid repeating similar arguments, we view the strategic measures in the minimax control problem as the strategic measures in an MDP induced by policies with particular structures. Specifically, define a set-valued mapping $A(\cdot)$ on $\mathbb{X}$ by $A(x) := \{ a = (a^1, a^2) \in \mathbb{A}_1 \times \mathbb{A}_2 \mid a^1 \in A_1(x), a^2 \in A_2(x) \}, x \in \mathbb{X}$. Let $\Gamma$ be the graph of $A(\cdot)$; i.e., $\Gamma := \{ (x, a) \in \mathbb{X} \times \mathbb{A} \mid (x, a^1) \in \Gamma_1, (x, a^2) \in \Gamma_2 \}$. Note that $\Gamma$ is the preimage of $\Gamma_1 \times \Gamma_2$ under the homeomorphism $(x, a^1, a^2) \mapsto (x, a^1, a^2, a^2)$ from $\mathbb{X} \times \mathbb{A}$ into $\mathbb{X} \times \mathbb{A}_1 \times \mathbb{X} \times \mathbb{A}_2$. Therefore, $\Gamma$ is analytic (Borel) if $\Gamma_1$ and $\Gamma_2$ are analytic (Borel) by [5, Props. 7.38, 7.40]. The strategic measures in the minimax control problem can also be viewed as the strategic measures of a subset $\Pi$ of policies in the MDP whose control constraints are specified by $\Gamma$, where $\Pi$ consists of those policies $\pi := \{ \mu_n \}_{n \geq 0}$ such that each $\mu_n$ has the product form

\[
\mu_n(da_n | h_n) = \mu_n^1(da_n^1 | i_n) \mu_n^2(da_n^2 | h_n), \quad (5.19)
\]
for some universally measurable stochastic kernel \( \mu_1^n \) on \( \mathcal{A}_1 \) given \( I_n \) and some universally measurable stochastic kernel \( \mu_2^n \) on \( \mathcal{A}_2 \) given \( H_n \).

Recall that \( \mathcal{P}(\mathcal{A}_1) \otimes \mathcal{P}(\mathcal{A}_2) \) denotes the set of all product probability measures on \( \mathcal{B}(\mathcal{A}_1 \times \mathcal{A}_2) \).

Recall also that for \( n \geq 0 \), \( \nu^n_1(\{a^n_1 | i_n; p\}) \) is a Borel measurable stochastic kernel on \( \mathcal{A}_1 \) given \( h_n \times \mathcal{P}(\Omega) \) such that for each \( p \in \mathcal{P}(\Omega) \), it is a conditional distribution of \( a^n_1 \) given \( i_n \) w.r.t. \( p \). For \( i = 1, 2 \), let \( \nu^n_i(\{a^n_i | h_n; p\}) \) be a Borel measurable stochastic kernel on \( \mathcal{A}_i \) given \( H_n \times \mathcal{P}(\Omega) \) such that for each \( p \in \mathcal{P}(\Omega) \), it is a conditional distribution of \( a^n_i \) given \( h_n \) w.r.t. \( p \). For \( p \in \mathcal{P}(\Omega) \), consider these constraints:

\[
\nu^n_1(\{a^n_1 | h_n; p\}) \in \mathcal{P}(\mathcal{A}_1) \otimes \mathcal{P}(\mathcal{A}_2), \quad \forall n \geq 0, \quad p\text{-almost surely;}
\]

\[
\nu^n_i(\{a^n_i | h_n; p\}) = \hat{\nu}_i^n(\{a^n_i | i_n; p\}), \quad \forall n \geq 0, \quad p\text{-almost surely.} \quad (5.20)
\]

We now use them to characterize the set \( S_M \) of strategic measures in the minimax control problem.

**Lemma 5.9.** The set \( S_M = \{p \in \mathcal{P}(\Omega) \mid p \text{ satisfies (5.2)-(5.3) and (5.20)-(5.21)}\} \).

**Proof (outline).** Clearly, every \( p \in S_M \) satisfies the constraints (5.2)-(5.3) and (5.20)-(5.21). The converse follows from the same line of proof for Lemma 5.3. Specifically, if \( p \) satisfies these constraints, we define a pair of policies \( \pi_i := \{\mu_i^n\}_{n \geq 0} \in \Pi_i, \ i = 1, 2 \), for the minimax control problem as follows. For each \( n \geq 0 \), we let \( \mu_1^n(\{a^n_1 | i_n\}) := \nu_i^n(\{a^n_1 | i_n; p\}) \) except on a set \( D_1^n \in \mathcal{U}(I_n) \) with \( p\{i_n \in D_1^n\} = 0 \), and we let \( \mu_2^n(\{a^n_2 | h_n\}) := \nu_i^n(\{a^n_2 | h_n; p\}) \) except on a set \( D_2^n \in \mathcal{U}(H_n) \) with \( p\{h_n \in D_2^n\} = 0 \). These sets \( D_1^n \) and \( D_2^n \) consist of points where the control constraints are violated:

\[
D_1^n := \{i_n \in I_n \mid \nu_1^n(A_1(x_n) | i_n; p) < 1\}, \quad D_2^n := \{h_n \in H_n \mid \nu_2^n(A_2(x_n) | h_n; p) < 1\}.
\]

Then, with \( f_i : \mathbb{X} \rightarrow \mathcal{A}_i \) being an analytically measurable selection of \( \Gamma_i, \ i = 1, 2 \), we let \( \mu_1^n(\{a^n_1 | i_n\}) := f_1(x_n) \) on \( D_1^n \), and we let \( \mu_2^n(\{a^n_2 | h_n\}) := f_2(x_n) \) on \( D_2^n \). Finally, by viewing \( (\pi_1, \pi_2) \) as a single policy of the form (5.19) in an MDP and noting that (5.20) implies \( \nu^n_1(\{a^n_1 | h_n; p\}) = \nu^n_1(\{a^n_1 | i_n; p\}) \), \( p\)-almost surely, we can apply Lemma 5.1 to obtain that \( p = \rho_x[\pi_1, \pi_2] \) and hence \( p \in S_M \). \( \square \)

**Proof of Props. 4.2 and 4.5 (outline).** As Lemma 5.2(i) and the proof of Prop. 5.1 showed, the set \( \{p \in \mathcal{P}(\Omega) \mid p \text{ satisfies (5.2)-(5.3) and (5.21)}\} \) is analytic (Borel) if \( \Gamma \) is analytic (Borel). For each \( n \geq 0 \), the set \( \{p \in \mathcal{P}(\Omega) \mid p \text{ satisfies (5.20) for the given } n\} \) is Borel by Lemma 5.2(iii). Combining these results with Lemma 5.9 yields that the set \( S_M \) is analytic, and that it is Borel if \( \Gamma_1 \) and \( \Gamma_2 \) are Borel. The desired conclusion for the set \( S_M \) follows from the same proof for Lemma 3.1. This proves Props. 4.2(ii).

Proposition 4.2(ii) follows from arguments similar to those given in the proofs of Props. 5.2 and Lemma 5.9. In particular, we construct the desired policy pair \( \pi_i := \{\mu_i^n\}_{n \geq 0} \in \Pi_i, \ i = 1, 2 \) by modifying the stochastic kernels \( \nu_i^n(\{a^n_i | i_n; \zeta(x)\}) \) and \( \nu_i^n(\{a^n_i | h_n; \zeta(x)\}) \) to satisfy the control constraints. This modification is similar to the one in the proof of Prop. 5.2 except that we also take into account the dependence of \( \zeta(x) \) on the initial state \( x \) as in the proof of Lemma 5.9.

The proof of Prop. 4.5 is almost the same as the above proof for Prop. 4.2(ii). The only extension we need is to replace \( \zeta(x) \) by \( \zeta(x, \theta) \) in the proof of Lemma 5.9 and in the preceding arguments, when defining the stochastic kernels \( \{\mu_i^n[\theta]\}_{n \geq 0} \in \Pi_i, \ i = 1, 2 \), that form the desired parametrized policies \( \pi^i(\theta) \), for \( \theta \in \Theta \).

**Proof of Prop. 4.3.** (i) For \( p, p' \in \mathcal{P}(\Omega) \), recall that condition (4.17) requires that \( \nu_i^n(\{a^n_i | i_n; p\}) = \nu_i^n(\{a^n_i | i_n; p'\}) \) \( p' \text{-almost surely and condition (4.18) requires } \hat{\pi}_i(p') \text{ to be absolutely continuous w.r.t. } \hat{\pi}_i(p) \), for all \( n \geq 0 \). Let \( E_n := \{(p, p') \in \mathcal{P}(\Omega)^2 \mid \text{ (p, p') satisfies (4.17) for the given } n\} \) and \( F_n := \{(p, p') \in \mathcal{P}(\Omega)^2 \mid \text{ (p, p') satisfies (4.18) for the given } n\} \). We first prove that these sets are Borel. For the set \( E_n \), this follows from applying Lemma 5.2(iv) with \( (\theta, p) \) in the lemma corresponding to \( (p, p') \) here and with \( \Psi_1 \) and \( \Psi_2 \) corresponding to \( \nu_i^n(\{a^n_i | i_n; p\}) \) and \( \nu_i^n(\{a^n_i | i_n; p\}) \), respectively.
Consider the set $F_n$. Note that the marginal $\hat{\mu}_n(p)$ of $p$ on $I_n$ is a Borel measurable function from $\mathcal{P}(\Omega)$ into $\mathcal{P}(I_n)$. For $\eta_1, \eta_2 \in \mathcal{P}(I_n)$, we write $\eta_1 \ll \eta_2$ if $\eta_1$ is absolutely continuous w.r.t. $\eta_2$, and we denote by $s(\eta_1, \eta_2)$ the singular part of $\eta_1$ w.r.t. $\eta_2$. By [14, Thm. 2.10], $s(\cdot, \cdot)$ is a Borel measurable function from $\mathcal{P}(I_n)^2$ into the space of nonnegative finite measures on $I_n$ (equipped with the topology of weak convergence), and therefore, the set

$$B := \{(\eta_1, \eta_2) \in \mathcal{P}(I_n)^2 \mid \eta_1 \ll \eta_2\} = \{(\eta_1, \eta_2) \in \mathcal{P}(I_n)^2 \mid s(\eta_1, \eta_2)(I_n) = 0\}$$

is Borel by [14, Thms. 2.1, 2.10]. Since $F_n$ is the preimage of $B$ under the Borel measurable mapping $(p, p') \mapsto (\hat{\mu}_n(p'), \hat{\mu}_n(p))$, it follows that $F_n$ is Borel.

Now $\hat{S}_M = D \cap (\mathbb{X} \times \cap_{n \geq 0}(E_n \cap F_n))$, where $D := \{(x, p, p') \in \mathbb{X} \times \mathcal{P}(\Omega)^2 \mid (x, p) \in \hat{S}_M, (x, p') \in \hat{S}_M\}$. Since $D$ is the preimage of the set $\hat{S}_M \times \hat{S}_M$ under the homeomorphism $(x, p, p') \mapsto (x, p, x, p')$ from $\mathbb{X} \times \mathcal{P}(\Omega)^2$ into $\mathbb{X} \times \mathcal{P}(\Omega) \times \mathbb{X} \times \mathcal{P}(\Omega)$, $D$ is analytic (Borel) if $\hat{S}_M$ is analytic (Borel) by [5, Props. 7.38, 7.40]. Combining this with Prop. 4.2(i) and the Borel measurability of $E_n$ and $F_n$, it follows that $\hat{S}_M$ is analytic (Borel) if $\Gamma_1$ and $\Gamma_2$ are analytic (Borel). This proves part (i) of the proposition.

(ii) Let $p = \rho_x[\pi_1, \pi_2]$ for some $x \in \mathbb{X}$, $\pi_1 := \{\mu_n^1\}_{n \geq 0} \in \Pi_1$, and $\pi_2 \in \Pi_2$. It is clear that under Assumption 4.1, $\rho_x[\pi_1, \pi_2] \in \hat{S}_M(x, p)$ for all $\pi_2 \in \Pi_2$. Now consider an arbitrary $p' \in \hat{S}_M(x, p)$. Then $p' = \rho_x[\tilde{\pi}_1, \tilde{\pi}_2]$ for some $\tilde{\pi}_i = \{\hat{\mu}_n^i\}_{n \geq 0} \in \Pi_i$, $i = 1, 2$, and furthermore, in view of (4.17), for all $n \geq 0$,

$$p'\{\hat{\mu}_n^1(da_n^1 \mid i_n; p) = \hat{\mu}_n^1(da_n^1 \mid i_n)\} = 1, \quad p'\{\hat{\mu}_n^2(da_n^2 \mid i_n; p) = \hat{\mu}_n^2(da_n^2 \mid i_n)\} = 1,$$

which implies $p'\{\hat{\mu}_n^1(da_n^1 \mid i_n) = \hat{\mu}_n^1(da_n^1 \mid i_n; p)\} = 1$. Combining this with the relations

$$\hat{\mu}_n(p') \ll \hat{\mu}_n(p) \quad \text{and} \quad p'\{\hat{\mu}_n^1(da_n^1 \mid i_n) = \hat{\mu}_n^1(da_n^1 \mid i_n)\} = 1, \quad \forall n \geq 0,$$

(where the first relation is from (4.18)), we obtain that $p'\{\hat{\mu}_n^1(da_n^1 \mid i_n) = \hat{\mu}_n^1(da_n^1 \mid i_n)\} = 1$ for all $n \geq 0$. Since $\nu_n(da_n \mid h_n; p') = \hat{\mu}_n^1(da_n^1 \mid i_n) \hat{\mu}_n^2(da_n^2 \mid h_n)$ $p'$-almost surely for all $n \geq 0$, this implies

$$\nu_n(da_n \mid h_n; p') = \hat{\mu}_n^1(da_n^1 \mid i_n) \hat{\mu}_n^2(da_n^2 \mid h_n), \quad \forall n \geq 0, \quad p'$-almost surely.$$

Then by viewing $(\pi_1, \tilde{\pi}_2)$ as a single policy of the form (5.19) in an MDP and by applying Lemma 5.1, we have that $p' = \rho_x[\pi_1, \tilde{\pi}_2]$. This proves part (ii) of the proposition.

### 5.3.3 Projective Class $\Sigma^1_2$ and Proof of Proposition 4.4

First, we briefly explain some concepts and results from descriptive set theory [33] that will be needed in proving Prop. 4.4. In descriptive set theory, the projective class $\Sigma^1_2$ consists of all sets $B$ such that for some Polish spaces $X$ and $Y$, $B \subset X \times Y$ and $B$ is the image of some coanalytic set in $Y$ under a continuous mapping from $Y$ into $X$ [33, Sec. 37.A]. The class $\Sigma^1_2$ is closed under countable intersections and unions, and continuous preimages and images (in particular, projections) [33, Prop. 37.1]. It is also closed under the Souslin operation and contains all analytically measurable sets, as well as all limit-measurable sets (cf. [33, Sec. 37.A] and [5, Prop. B.10]).

We call a set in the class $\Sigma^1_2$ a $\Sigma^1_2$ set. Besides the aforementioned basic facts, the following results about $\Sigma^1_2$ sets are important for our purpose. Let $X$ and $Y$ be Polish spaces. For a function $f$, let grh $f$ denote its graph.

(a) A uniformization theorem of Kondó [33, Cor. 38.7] asserts that if $B$ is a $\Sigma^1_2$ set in $X \times Y$, then there exists a function $f : \text{proj}_X(B) \to Y$ with grh $f \subset B$ such that grh $f$ is a $\Sigma^1_2$ set.

(b) Under the axiom of analytic determinacy, $\Sigma^1_2$ sets are universally measurable [33, Thm. 36.20].

(c) Under the axiom of analytic determinacy, the function $f$ in (a) is universally measurable. (This follows from (a) and (b), since for $E \in \mathcal{B}(Y)$, $f^{-1}(E) = \text{proj}_X(\text{grh} f \cap X \times E)$ is a $\Sigma^1_2$ set.)
We are now ready to prove Prop. 4.4. The line of argument is almost the same as that of [5, Props. 7.47, 7.50] and similar to that of [12, Cor. 1, Thm. 2], except that we use the above results for \( \Sigma_1^1 \) sets instead of the properties of analytic or Borel sets used in those cited references.

**Proof of Prop. 4.4.** (i) For each \( r \in \mathbb{R} \), the set \( \{(x, y) \in D \mid f(x, y) < r \} \) is a \( \Sigma^1_2 \) set. To see this, write it as \( D \setminus F_r \), where \( F_r := \{(x, y) \in D \mid f(x, y) \geq r\} \). Since \( f \) is upper semianalytic, \( F_r \) is analytic [5, Lem. 7.30(1)]; by assumption \( D \) is analytic. Therefore, \( D \setminus F_r \) is a \( \Sigma^1_2 \) set. Then the set

\[
E_r := \{ x \in \text{proj}_X(D) \mid f^*(x) < r \} = \text{proj}_X(\{(x, y) \in D \mid f(x, y) < r\})
\]

is also a \( \Sigma^1_2 \) set [33, Prop. 37.1], so by [33, Thm. 36.20], under the axiom of analytic determinacy, \( E_r \) is universally measurable for all \( r \in \mathbb{R} \). By [15, Thm. 4.1.6], this implies that \( f^* \) is universally measurable.

(ii) To prove the universal measurability of \( E^* \), consider first the set

\[
G := \{(x, y, b) \in X \times Y \times \bar{\mathbb{R}} \mid (x, y) \in D, f(x, y) \leq b\} = \bigcap_{k \geq 1} \bigcup_{\epsilon \in \mathbb{Q}} \{(x, y, b) \in X \times Y \times \bar{\mathbb{R}} \mid (x, y) \in D, f(x, y) \leq r, r \leq b + k^{-1}\},
\]

where \( \mathbb{Q} := \mathbb{Q} \cup \{-\infty, +\infty\} \) (recall that \( \mathbb{Q} \) is the set of rational numbers). The preceding proof for part (i) showed also that \( \{(x, y) \in D \mid f(x, y) \leq r\} \) is a \( \Sigma^1_2 \) set for \( r \in \mathbb{R} \) (since it is \( D \) for \( r = +\infty \), and for \( r < +\infty \), it can be expressed as a countable intersection of sets of the form \( \{(x, y) \in D \mid f(x, y) < r'\}, r' \in \mathbb{R} \). This implies that for each \( k \) and \( r \), the set in the right-hand side of (5.22) is a \( \Sigma^1_2 \) set. Consequently, \( G \) and \( \text{proj}_{X \times \bar{\mathbb{R}}}(G) \) are \( \Sigma^1_2 \) sets and therefore, universally measurable under the axiom of analytic determinacy [33, Thm. 36.20].

Next, define a mapping \( T : \text{proj}_X(D) \to X \times \bar{\mathbb{R}} \) by \( T(x) := (x, f^*(x)) \). The set \( \text{proj}_X(D) \) is analytic since \( D \) is analytic [5, Prop. 7.39]. As proved in part (i), \( f^* \) is universally measurable. Therefore, \( T \) is universally measurable. Since the set \( E^* \) can be expressed as

\[
E^* = \{ x \in \text{proj}_X(D) \mid \exists y \in D_x \text{ s.t. } f(x, y) = f^*(x) \} = T^{-1}(\text{proj}_{X \times \bar{\mathbb{R}}}(G)),
\]

in view of the universal measurability of \( \text{proj}_{X \times \bar{\mathbb{R}}}(G) \) proved earlier, we have, by [5, Cor. 7.44.1], that \( E^* \) is universally measurable.

We now prove the rest of the proposition, by constructing a function \( \psi \) that satisfies the requirements for an arbitrary \( \epsilon > 0 \). To this end, we first define two other functions \( \tilde{\psi} \) and \( \psi^* \) that will be needed in the construction.

Since \( D \) is analytic, by the Jankov-von Neumann selection theorem [5, Prop. 7.49], there exists an analytically measurable function \( \psi : \text{proj}_X(D) \to Y \) with \( \text{grh} \, \psi \subseteq D \).

As proved above, \( G \) is a \( \Sigma^1_2 \) set, so by a uniformization theorem of Kondô [33, Cor. 38.7], there exists a function \( \phi : \text{proj}_{X \times \bar{\mathbb{R}}}(G) \to Y \) with \( \text{grh} \, \phi \subset G \) such that \( \text{grh} \, \phi \) is a \( \Sigma^1_2 \) set. Then \( \phi \) is universally measurable under the axiom of analytic determinacy, as discussed before the proof. Consequently, if \( E^* \) is nonempty, the function \( \psi^* : E^* \to Y \) with \( \psi^*(x) := \phi(x, f^*(x)) \) is also universally measurable [5, Prop. 7.44]. Note that by the definition of \( \psi^* \),

\[
\psi^*(x) \in D_x, \quad f(x, \psi^*(x)) = f^*(x), \quad \forall x \in E^*.
\]

We now proceed to construct the desired function \( \psi \). For integers \( k \), define sets

\[
F(k) := \{(x, y) \in D \mid f(x, y) < k\}, \quad B(k) := \{ x \in \text{proj}_X(D) \mid (k-1)\epsilon \leq f^*(x) < k\},
\]

\[
B(-\infty) := \{ x \in \text{proj}_X(D) \mid f^*(x) = -\infty \}, \quad B(+\infty) := \{ x \in \text{proj}_X(D) \mid f^*(x) = +\infty \}.
\]

As proved earlier in part (i), \( F(k) \) is a \( \Sigma^1_2 \) set, and \( f^* \) is universally measurable. So, under the axiom of analytic determinacy, all these sets are universally measurable.
For each $k$, by a uniformization theorem of Kondô [33, Cor. 38.7], there exists a function $\psi_k : \text{proj}_X(F(k)) \to Y$ with $\text{grh} \psi_k \subset F(k)$ such that $\text{grh} \psi_k$ is a $\Sigma^1_2$ set. As discussed before the proof, under the axiom of analytic determinacy, $\psi_k$ is universally measurable. Let $k^*$ be an integer such that $k^* < -\epsilon^2$. Now define $\psi : \text{proj}_X(D) \to Y$ as follows. Let $\psi(x) := \psi_k^*(x)$ for $x \in E$. For $x \in \text{proj}_X(D) \setminus E^*$, let

$$
\psi(x) := \begin{cases}
\psi_k(x) & \text{if } x \in B(k), \\
\tilde{\psi}(x) & \text{if } x \in B(+\infty), \\
\psi_k^*(x) & \text{if } x \in B(-\infty).
\end{cases}
$$

Then $\psi$ is universally measurable and $\text{grh} \psi \subset D$, and moreover, in view of (5.23) and the definitions of $B(k), B(\infty)$, and $B(-\infty)$, $\psi$ satisfies (4.19)-(4.20), as required. This completes the proof. \hfill \Box

Remark 5.1. As can be seen from the preceding proof, for the conclusions of Prop. 4.4 to hold, it suffices that $D$ is a $\Sigma^1_2$ set and $f$ is such that $\{(x, y) \in D \mid f(x, y) < r\}$ is a $\Sigma^1_2$ set for all $r \in \mathbb{R}$. (Essentially we only need to make a minor change in the preceding proof, which is to apply Kondô’s uniformization theorem to $D$ in the definition of the function $\psi.$) One can draw a parallel between this generalization of Prop. 4.4 and the result in [5, Props. 7.47, 7.50], in which the above two sets are assumed to be analytic. Although this is beyond our scope, we note that the similarity between these two results reflects the parallelism between $\Sigma^1_2$ sets and analytic sets (see [33, Sec. 37] on the hierarchy of projective classes for more details). \hfill \Box

Acknowledgements

The author is grateful to Professor Eugene Feinberg for pointing to several important references on strategic measures and stochastic games and valuable feedback on earlier versions of this work. The author also thanks Professor William Sudderth, for mentioning the early work [36] on Borel gambling problems, which also used Kondô’s uniformization theorem in its analysis; Professor Serdar Yüksel, for helpful discussion on minimax control problems; and an anonymous reviewer, whose critical comments helped improve the paper. This research was supported, during different periods of time, by grants from DeepMind, Alberta Machine Intelligence Institute (AMII), and Alberta Innovates—Technology Futures (AITF).

References

[1] Arapostathis, A., Borkar, V. S., Fernández-Gaucherand, E., Ghosh, M. K., and Marcus, S. I. (1993). Discrete-time controlled Markov processes with average cost criterion: A survey. SIAM J. Control Optim., 31(2):282–344.
[2] Ash, R. (1972). Real Analysis and Probability. Academic Press, New York.
[3] Balder, E. J. (1989). On compactness of the space of policies in stochastic dynamic programming. Stoch. Proc. Appl., 32:141–150.
[4] Bäuerle, N. and Rieder, U. (2014). More risk-sensitive Markov decision processes. Math. Oper. Res., 39:105–120.
[5] Bertsekas, D. P. and Shreve, S. E. (1978). Stochastic Optimal Control: The Discrete Time Case. Academic Press, New York.
[6] Bierth, K. J. (1987). An expected average reward criterion. Stoch. Proc. Appl., 26:123–140.
[7] Blackwell, D. (1968). A Borel set not containing a graph. Ann. Math. Statist., 39:1345–1347.
[8] Blackwell, D. (1976). The stochastic processes of Borel gambling and dynamic programming. Ann. Statist., 4:370–374.
[9] Blackwell, D., Freedman, D., and Orkin, M. (1974). The optimal reward operator in dynamic programming. Ann. Probability, 2(5):926–941.
[10] Blackwell, D. and Ryll-Nardzewski, C. (1963). Non-existence of everywhere proper conditional distributions. Ann. Math. Statist., 34:223–225.
[11] Borkar, V. S. (2002). Convex analytic methods in Markov decision processes. In Feinberg, E. A. and Shwartz, A., editors, Handbook of Markov Decision Processes: Methods and Applications, pages 347–375. Springer Science+Business Media, New York.

[12] Brown, L. D. and Purves, R. (1973). Measurable selections of extrema. Ann. Statist., 1:902–912.

[13] Cavazos-Cadena, R. and Salem-Siva, F. (2010). The discounted method and equivalence of average criteria for risk-sensitive Markov decision processes on Borel spaces. Appl. Math. Optim., 61:167–190.

[14] Dubins, L. E. and Freedman, D. (1964). Measurable sets of measures. Pacific J. Math., 14:1211–1222.

[15] Dudley, R. M. (2002). Real Analysis and Probability. Cambridge University Press, Cambridge.

[17] Feinberg, E. A. (1980). An $\varepsilon$-optimality control of a finite Markov chain with an average reward criterion. Theory Probab. Appl., 25(1):70–81.

[18] Feinberg, E. A. (1982a). Controlled Markov processes with arbitrary numerical criteria. Theory Probab. Appl., 27(3):486–503.

[19] Feinberg, E. A. (1982b). Non-randomized Markov and semi-Markov strategies in dynamic programming. Theory Probab. Appl., 27(1):116–126.

[20] Feinberg, E. A. (1991). Non-randomized strategies in stochastic decision processes. Ann. Oper. Res., 29:315–332.

[21] Feinberg, E. A. (1996). On measurability and representation of strategic measures in Markov decision processes. In Statistics, Probability and Game Theory, volume 30 of Lecture Notes – Monograph Series, pages 29–43. IMS.

[22] Feinberg, E. A. and Kasyanov, P. O. (2021). MDPs with setwise continuous transition probabilities. Oper. Res. Lett., 49:734–740.

[23] Feinberg, E. A., Kasyanov, P. O., and Liang, Y. (2020). Fatou’s lemma in its classic form and Lebesgue’s convergence theorems for varying measures with applications to MDPs. Theory Probab. Appl., 65:270–291.

[24] Gikhman, I. I. and Skorokhod, A. V. (1979). Controlled Stochastic Processes. Springer, New York.

[25] Gode, K. (1938). The consistency of the axiom of choice and of the generalized continuum hypothesis. Proc. Natl. Acad. Sci. USA, 24:556–557.

[26] González-Trejo, J. I., Hernández-Lerma, O., and Hoyos-Reyes, L. F. (2002). Minimax control of discrete-time stochastic systems. SIAM J. Control Optim., 41(5):1626–1659.

[27] Hernández-Lerma, O. and Lasserre, J. B. (1996). Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer, New York.

[28] Hernández-Lerma, O. and Lasserre, J. B. (1999). Further Topics on Discrete-Time Markov Control Processes. Springer, New York.

[29] Hernández-Lerma, O. and Lasserre, J. B. (2002). The linear programming approach. In Feinberg, E. A. and Shwartz, A., editors, Handbook of Markov Decision Processes: Methods and Applications, chapter 12. Springer Science+Business Media, New York.

[30] Jaśkiewicz, A. (2007). Average optimality for risk-sensitive control with general state space. Ann. Appl. Probab., 17:654–675.

[31] Jaśkiewicz, A. and Nowak, A. S. (2014). Robust Markov control processes. J. Math. Anal. Appl., 420:1337–1353.

[32] Jaśkiewicz, A. and Nowak, A. S. (2018). Zero-sum stochastic games. In Başar, T. and Zaccour, G., editors, Handbook of Dynamic Game Theory, pages 215–279. Springer International Publishing AG, Switzerland.

[33] Kechriss, A. S. (1995). Classical Descriptive Set Theory. Springer-Verlag, New York.

[34] Koestner, P. (2014). Large Cardinals and Determinacy. In Zalta, E. N., editor, The Stanford Encyclopedia of Philosophy, Metaphysics Research Lab, Stanford University, Spring 2014 edition. https://plato.stanford.edu/archives/spr2014/entries/large-cardinals-determinacy/.

[35] Krengel, U. (1985). Ergodic Theorems. Walter de Gruyter, Berlin.

[36] Maitra, A., Purves, R., and Sudderth, W. (1990). Leavable gambling problems with unbounded utilities. Trans. Amer. Math. Soc., 320:543–567.

[37] Maitra, A. and Sudderth, W. (1993). Borel stochastic games with limsup payoffs. Ann. Probab., 21:861–885.

[38] Maitra, A. and Sudderth, W. (1998). Finitely additive stochastic games with Borel measurable payoffs. Int. J. Game Theory, 27:257–267.

[39] Martin, D. A. and Steel, J. R. (1988). Projective determinacy. Proc. Natl. Acad. Sci., 85:6582–6586.
Nowak, A. S. (1985). Universally measurable strategies in zero-sum stochastic games. *Ann. Probab.*, 13(1):269–287.

Nowak, A. S. (2010). On measurable minimax selectors. *J. Math. Anal. Appl.*, 366:385–388.

Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.

Prikry, K. and Sudderth, W. D. (2016). Measurability of the value of a parametrized game. *Int. J. Game Theory*, 45:675–683.

Rieder, U. (1991). Non-cooperative dynamic games with general utility functions. In Raghavan, T. E. S., Ferguson, T. S., Parthasarathy, T., and Vrieze, O. J., editors, *Stochastic Games and Related Topics*, pages 161–174. Kluwer, Dordrecht, The Netherlands.

Schäl, M. (1975). On dynamic programming: Compactness of the space of policies. *Stoch. Proc. Appl.*, 3:345–364.

Shapiro, A., Dentcheva, D., and Ruszcyński, A. (2021). *Lectures on Stochastic Programming: Modeling and Theory*. Society for Industrial and Applied Mathematics and Mathematical Optimization Society, Philadelphia, 3rd edition.

Shreve, S. E. and Bertsekas, D. P. (1978). Alternative theoretical frameworks for finite horizon discrete-time stochastic optimal control. *SIAM J. Control Optim.*, 16(6):953–978.

Shreve, S. E. and Bertsekas, D. P. (1979). Universally measurable policies in dynamic programming. *Math. Oper. Res.*, 4(1):15–30.

Strauch, R. E. (1966). Negative dynamic programming. *Ann. Math. Statist.*, 37:871–890.

Sudderth, W. (1969). On the existence of good stationary strategies. *Trans. Amer. Math. Soc.*, 135:399–414.

Vega-Amaya, O. (2018). Solutions of the average cost optimality equation for Markov decision processes with weakly continuous kernel: The fixed-point approach revisited. *J. Math. Anal. Appl.*, 464:152–163.

Wei, C., Fauß, M., and Chapman, M. P. (2022). CVaR-based safety analysis in the infinite time horizon setting. In *Proc. American Control Conference*, pages 2863–2870, Piscataway, NJ. IEEE.

Yu, H. (2015). On convergence of value iteration for a class of total cost Markov decision processes. *SIAM J. Control Optim.*, 53(4):1982–2016.

Yu, H. (2020). Average cost optimality inequality for Markov decision processes with Borel spaces and universally measurable policies. *SIAM J. Control Optim.*, 58(4):2469–2502.

Yu, H. (2022). On structural properties of optimal average cost functions in Markov decision processes with Borel spaces and universally measurable policies. *J. Math. Anal. Appl.*, 509:125954.

Yüksel, S. (2020). A universal dynamic program and refined existence results for decentralized stochastic control. *SIAM J. Control Optim.*, 58(5):2711–2739.

Yüksel, S. and Saldi, N. (2017). Convex analysis in decentralized stochastic control, strategic measures, and optimal solutions. *SIAM J. Control Optim.*, 55(1):1–28.