Existence of Markov equilibrium control in discrete time

Erhan Bayraktar [*]   Bingyan Han [†]

December 8, 2023

Abstract

For time-inconsistent stochastic controls in discrete time and finite horizon, an open problem in Björk and Murgoci (Finance Stoch, 2014) is the existence of an equilibrium control. A nonrandomized Borel measurable Markov equilibrium policy exists if the objective is inf-compact in every time step. We provide a sufficient condition for the inf-compactness and thus existence, with costs that are lower semicontinuous (l.s.c.) and bounded from below and transition kernels that are continuous in controls under given states. The control spaces need not to be compact.

Keywords: Time inconsistency, Borel measurable selection, inf-compact functions.

MSC codes: 49L20, 28B20, 91A80.

1 Introduction

In this short note, we revisit discrete-time stochastic controls with time-inconsistent objectives studied in Björk and Murgoci (2014). Unlike standard stochastic optimal control problems, time inconsistency reflects the phenomenon that agent preferences change in time. Bellman optimality principle fails. The optimal plan at the present moment is not followed in the future. Classical examples include non-exponential discounting (Strotz, 1955), prospect theory (Kahneman and Tversky, 1979), mean-variance portfolio selection (Basak and Chabakauri, 2010), and optimal stopping (Bayraktar et al., 2021). See the recent monograph Björk et al. (2021) and reference therein for a comprehensive review.

As proposed in Strotz (1955), agents should only consider the policy that they can actually follow. By the consistent planning framework in Strotz (1955), agents acknowledging time inconsistency view themselves at different future times as distinct planners. In this context, an equilibrium policy $\pi^* = (u_1^*, \ldots, u_{T-1}^*)$ should satisfy that, if all future selves at time $t+1, \ldots, T-1$ use $u_{t+1}^*, \ldots, u_{T-1}^*$, then it is optimal for the planner at time t to use u_t^*. Thus, there is no incentive for the planner to deviate from π^* at any time t. A formal definition is given in Björk and Murgoci (2014, Definition 2.5).

If equilibrium policies exist, Björk and Murgoci (2014, Theorem 3.14) characterizes them with the extended Bellman equation. However, the existence of an equilibrium policy remains open (Björk and Murgoci, 2014, Section 6). Recall the Borel space framework in Bertsekas and Shreve (1978, Chapters 7 and 8) for the classical stochastic optimal control. The existence of optimal policies is established through measurable selection theorems. The optimal policy obtained usually has weaker properties compared with the objective function. For instance, if the objective is l.s.c.,

then an optimal policy, if it exists, is Borel measurable (Bertsekas and Shreve, 1978, Proposition 7.33). If the objective is lower semianalytic, then an ε-optimal policy is universally measurable (Bertsekas and Shreve, 1978, Proposition 7.50). However, the extended Bellman equation introduces auxiliary functions (2.4), tied to the optimal policy, and thus inheriting the weaker properties. Consequently, demonstrating the existence of an equilibrium policy via a recursive application of measurable selection becomes challenging.

The key observation is, if the transition kernels of states have nice properties, then it may still be possible to apply certain measurable selection theorem repeatedly. The weak continuity of transition kernels is often inadequate. But a stronger continuity condition, given in Assumption 3.3 (4), is sufficient for the existence of an equilibrium policy. Notably, Assumption 3.3 (4) is not new and has been considered in Hernández-Lerma (1991) on average cost Markov decision processes. It is relevant to setwise convergence and also used in Hernández-Lerma and Lasserre (1999, Assumption 8.3.1 (c)) and Saldi et al. (2018, Assumption 3.1 (b)). In conjunction with a version of the Borel measurable selection theorem (Rieder, 1978, Theorem 4.1 and Corollary 4.3) under the inf-compact condition, we managed to prove the existence of a nonrandomized Borel measurable Markov equilibrium policy.

In the literature, Jaśkiewicz and Nowak (2021) explored infinite-horizon Markov decision processes with quasi-hyperbolic discounting in their equation (2.2). The existence of a randomized stationary Markov perfect equilibrium is proved with continuous costs and transition kernels with densities; see Jaśkiewicz and Nowak (2021, Assumptions C3.1 – C3.2). While Jaśkiewicz and Nowak (2021, Assumption C3.2) closely resembles Assumption 3.3 (4), their proof differs since they assumed compactness instead of inf-compactness and relied on the structure of quasi-hyperbolic discounting to apply the fixed point theorem. Other existence results are available for different formulations. Huang and Zhou (2021) considered weak and strong equilibria in the continuous time under certain convexity (concavity) conditions. Bayraktar et al. (2021) and reference therein studied time-inconsistent stopping problems. Bayraktar et al. (2023) examined relaxed equilibria with regularization. Bayraktar and Han (2023) considered optimal transport with time-inconsistency and provided the existence of the so-called equilibrium transport.

The rest of the paper is organized as follows. Section 2 formulates the general problem, while readers need to refer to Björk and Murgoci (2014) for the derivation of the extended Bellman equation. Section 3 establishes the sufficient condition for the existence. Illustrative examples are in Section 4.

2 Problem formulation

In this work, a topological space $\mathcal{X}$ will always be endowed with the Borel σ-algebra $\mathcal{B}(\mathcal{X})$. A Borel subset of a Polish (complete separable metric) space is called a Borel space. Consider a nonstationary discrete-time controlled Markov process with the following specifications:

(1) The finite horizon T: A positive integer.

(2) State spaces $\mathcal{X}_t$, $t \in \{1, \ldots, T\}$: $\mathcal{X}_t$ is a nonempty Borel space.

(3) Control spaces $\mathcal{A}_t$, $t \in \{1, \ldots, T-1\}$: $\mathcal{A}_t$ is a nonempty Borel space. $u_t \in \mathcal{A}_t$ is a control chosen at time t.

(4) Control constraints U_t, $t \in \{1, \ldots, T-1\}$: U_t is a function from $\mathcal{X}_t$ to the set of nonempty subsets of $\mathcal{A}_t$. Suppose the set

$$\Gamma_t := \{(x_t, u_t) | x_t \in \mathcal{X}_t, u_t \in U_t(x_t)\} \tag{2.1}$$

is a Borel subset of $\mathcal{X}_t \times \mathcal{A}_t$.

(5) Transition kernels $P_t(dx_{t+1}|x_t, u_t)$, $t \in \{1, \ldots, T-1\}$: Suppose $P_t(dx_{t+1}|x_t, u_t)$ is a Borel-measurable stochastic kernel on $\mathcal{X}_{t+1}$ given Γ_t.

The system begins at some state $x_1 \in \mathcal{X}_1$ and proceeds successively to state spaces $\mathcal{X}_2, \ldots, \mathcal{X}_{T-1}$ and finally terminates in $\mathcal{X}_T$. A policy controlling the system is a sequence $\pi_{t:T-1} := (\pi_t, \ldots, \pi_{T-1})$, where $\pi_t(\cdot|x_{1:t}, u_{1:t-1})$ is a Borel measurable stochastic kernel on $\mathcal{A}_t$ given the history and it satisfies $\pi_t(U_t(x_t)|x_{1:t}, u_{1:t-1}) = 1$ for every $(x_{1:t}, u_{1:t-1})$. In this work, we consider $\pi_t(\cdot|x_{1:t}, u_{1:t-1})$ that depends only on the current state and time and concentrates at one point for each state x_t, which is called nonrandomized Markov policies. Thus, $\pi_t(\cdot|x_{1:t}, u_{1:t-1}) = u_t(x_t)$ for some Borel measurable functions $u_t(\cdot)$. We still use the notation $\pi^t = (u_t, \ldots, u_{T-1})$.

At time t, the agent wants to minimize a state-dependent and nonlinear cost objective (Björk and Murgoci, 2014, Equation 3.8):

$$J_t(x; \pi^t) := \int \left[\sum_{k=t}^{T-1} C_k(t, x, x_k, u_k) + F(t, x, x_T) \right] P(dx_{t+1:T}|x, \pi^t)$$
$$+ G\left(t, x, \int H(x_T) P(dx_T|x, \pi^t) \right), \tag{2.2}$$

where $P(dx_{t+1:T}|x, \pi^t) := P_{T-1}(dx_T|x_{T-1}, u_{T-1}(x_{T-1})) \ldots P_t(dx_{t+1}|x_t, u_t(x_t))$ under the policy π^t and $x_t = x$. The costs $C_k(s, y, x_k, u_k)$, $F(s, y, x_T)$, $H(x_T)$, and $G(s, y, h)$ are Borel measurable. Given $t \in \{1, \ldots, T-1\}$, the term (s, y) with $s \in \{1, \ldots, t\}$ and $y \in \mathcal{X}_s$ represents the state-dependence. For example, consider $F(t, x, x_T)$. At a different time $s \neq t$, it becomes $F(s, x_s, x_T)$ and thus the preference of the agent changes. $G(s, y, h)$ is nonlinear and does not satisfy the tower property in general. Therefore, Bellman optimality principle fails due to the state-dependence and nonlinearity. We refer readers to Björk and Murgoci (2014); Björk et al. (2021) for specific applications with time inconsistency.

Inspired by the consistent planning in Strotz (1955), Björk and Murgoci (2014, Definition 2.5) introduced a subgame perfect Nash equilibrium strategy as follows. Given time $t < T$ and state $x_t = x$, define a control policy $\pi^{t,u} = (u, u_{t+1}^*, \ldots, u_{T-1}^*)$ on $t, \ldots, T-1$, where $u \in U_t(x_t)$ is an arbitrary control value and $u_s^* = u_s^*(x_s), \forall x_s \in \mathcal{X}_s$, $s \in \{t+1, \ldots, T-1\}$, are Borel measurable functions. A Borel measurable function $u_t^*(x_t)$ is a nonrandomized Markov equilibrium control at time t if for every x, we have

$$\inf_{u \in U_t(x_t)} J_t(x; \pi^{t,u}) = J_t(x; \pi^{t,*}),$$

where $\pi^{t,*} = (u_t^*, u_{t+1}^*, \ldots, u_{T-1}^*)$. Denote the corresponding equilibrium value function as $V_t(x) = J_t(x; \pi^{t,*})$.

If $\pi^* = (u_1^*, \ldots, u_{T-1}^*)$ exists, Björk and Murgoci (2014, Theorem 3.14) proved that the value function V_t satisfies the extended Bellman equation given by

$$V_t(x_t) = \inf_{u \in U_t(x_t)} \left\{ C_t(t, x_t, x_t, u) + \int V_{t+1}(x_{t+1}) P_t(dx_{t+1}|x_t, u) \right.$$
$$+ \sum_{k=t+1}^{T-1} \int b_k(t, x_t, x_{t+1}) P_t(dx_{t+1}|x_t, u) \tag{2.3}$$
$$- \sum_{k=t+1}^{T-1} \int b_k(t+1, x_{t+1}, x_{t+1}) P_t(dx_{t+1}|x_t, u)$$

3

$$+ \int f(t, x_t, x_{t+1}) P_t(dx_{t+1}|x_t, u) - \int f(t+1, x_{t+1}, x_{t+1}) P_t(dx_{t+1}|x_t, u)$$

$$+ G\left(t, x_t, \int h(x_{t+1}) P_t(dx_{t+1}|x_t, u)\right)$$

$$- \int G\left(t+1, x_{t+1}, h(x_{t+1})\right) P_t(dx_{t+1}|x_t, u)\bigg\}.$$

The functions b_k, f, and h are defined backwardly as

$$b_k(s, y, x_{t+1}) := \int C_k(s, y, x_k, u_k^*(x_k)) P(dx_k|x_{t+1}, \pi^{t+1,*}), \quad k \in \{t+1, \ldots, T-1\},$$

$$f(s, y, x_{t+1}) := \int F(s, y, x_T) P(dx_T|x_{t+1}, \pi^{t+1,*}), \tag{2.4}$$

$$h(x_{t+1}) := \int H(x_T) P(dx_T|x_{t+1}, \pi^{t+1,*}),$$

with a nonrandomized Markov equilibrium policy $\pi^{t+1,*} = (u_{t+1}^*, \ldots, u_{T-1}^*)$ found for time $t+1, \ldots, T-1$.

In fact, the value function V_{t+1} satisfies

$$V_{t+1}(x_{t+1}) = \sum_{k=t+1}^{T-1} b_k(t+1, x_{t+1}, x_{t+1}) + f(t+1, x_{t+1}, x_{t+1}) + G(t+1, x_{t+1}, h(x_{t+1})). \tag{2.5}$$

Thus, the extended Bellman equation (2.3) reduces to

$$V_t(x_t) = \inf_{u \in U_t(x_t)} \mathcal{L}_{t+1}(t, x_t, x_t, u), \tag{2.6}$$

where

$$\mathcal{L}_{t+1}(s, y, x_t, u) := C_t(s, y, x_t, u) + \sum_{k=t+1}^{T-1} \int b_k(s, y, x_{t+1}) P_t(dx_{t+1}|x_t, u)$$

$$+ \int f(s, y, x_{t+1}) P_t(dx_{t+1}|x_t, u) + G\left(s, y, \int h(x_{t+1}) P_t(dx_{t+1}|x_t, u)\right).$$

We emphasize that the extended Bellman equation (2.6) relies on the auxiliary functions b_k, f, and h in (2.4) backwardly. When the existence of a nonrandomized Markov equilibrium policy $\pi^{t+1,*}$ is proved for time $t+1, \ldots, T-1$, then the characterization of u_t^* is given by (2.6) at time t.

3 Existence

In Lemma 3.1, we quote a Borel measurable selection theorem in Rieder (1978, Theorem 4.1 and Corollary 4.3), which is relevant to Brown and Purves (1973, Corollary 1). Let D be a subset of a product space $\mathcal{X} \times \mathcal{Y}$. Denote proj$(D)$ as the projection of D onto $\mathcal{X}$. For $x \in \mathcal{X}$, the x-section of D is denoted by D_x. Crucially, Lemma 3.1 does not require the compactness of D_x for the existence of a minimizer. Instead, it supposes the level sets in (3.1) are compact. Functions satisfying this property are called inf-compact, see Hernández-Lerma and Lasserre (1996, p.28).

Lemma 3.1. *(Rieder, 1978, Theorem 4.1 and Corollary 4.3) Let $\mathcal{X}$ and $\mathcal{Y}$ be two Borel spaces. Consider a Borel subset D of $\mathcal{X} \times \mathcal{Y}$ and a Borel measurable function $f(x,y) : D \to \mathbb{R}$. If the set*

$$\{y \in D_x | f(x,y) \leq r\} \tag{3.1}$$

is compact for every $r \in \mathbb{R}$ and $x \in \mathrm{proj}(D)$, then there exists a Borel measurable minimizer $g : \mathrm{proj}(D) \to \mathcal{Y}$, such that

$$g(x) \in D_x \text{ and } f(x,g(x)) = \inf_{y \in D_x} f(x,y), \quad x \in \mathrm{proj}(D). \tag{3.2}$$

Remark 3.2. *Since compact subsets of metric spaces are closed, the inf-compact assumption in Lemma 3.1 implies that $f(x,\cdot)$ is l.s.c. with each $x \in \mathrm{proj}(D)$. Moreover, since $D_x = \cup_{r \in \mathbb{Z}}\{y \in D_x | f(x,y) \leq r\}$, D_x is σ-compact with each $x \in \mathrm{proj}(D)$. Note that Brown and Purves (1973, Corollary 1) also holds for Borel spaces (Brown and Purves, 1973, Remark 2). Hence, the assumptions in Brown and Purves (1973, Corollary 1) are satisfied under the assumptions in Lemma 3.1.*

If we can verify that

$$\{u \in U_t(x_t) | \mathcal{L}_{t+1}(t,x_t,x_t,u) \leq r\} \tag{3.3}$$

is compact for every $r \in \mathbb{R}$ and $x_t \in \mathcal{X}_t$, then a Borel measurable minimizer $u_t^*(x_t)$ exists for (2.6) at time t. If the inf-compact condition (3.3) can be verified for every time t after the existence of an equilibrium is proved for time $t+1, \ldots, T-1$, then a nonrandomized Markov equilibrium policy exists. Assumption 3.3 provides a sufficient condition.

Assumption 3.3. *(1) The functions F, G, H, and C_t, $t \in \{1,\ldots,T-1\}$ in the objective (2.2) are Borel measurable, real-valued, and nonnegative.*

(2) For every $t \in \{1,\ldots,T-1\}$, $C_t(s,y,x_t,u)$ is l.s.c. in u for each (s,y,x_t), where $s \in \{1,\ldots,t\}$, $y \in \mathcal{X}_s$, and $x_t \in \mathcal{X}_t$. Moreover, the set

$$U_t(r,x_t) := \{u \in U_t(x_t) | C_t(t,x_t,x_t,u) \leq r\} \tag{3.4}$$

is compact for every $r \in \mathbb{R}$, $x_t \in \mathcal{X}_t$, and $t \in \{1,\ldots,T-1\}$.

(3) $G(s,y,h)$ is l.s.c. in h and nondecreasing in h for each (s,y), where $s \in \{1,\ldots,T-1\}$ and $y \in \mathcal{X}_s$.

(4) For each $t \in \{1,\ldots,T-1\}$, $\int V(x_{t+1})P_t(dx_{t+1}|x_t,u)$ is continuous in u for each $x_t \in \mathcal{X}_t$ and bounded Borel measurable function V.

Remark 3.4. *Assumption 3.3 (4) is equivalent to imposing that $\int V(x_{t+1})P_t(dx_{t+1}|x_t,u)$ is l.s.c. in u for each $x_t \in \mathcal{X}_t$ and bounded Borel measurable function V. We could also consider nonnegative V only; see Hernández-Lerma and Lasserre (1999, p. 44). Assumption 3.3 (3) is a sufficient condition that the composition $G\left(s,y,\int h(x_{t+1})P_t(dx_{t+1}|x_t,u)\right)$ is l.s.c. in u when (s,y,x_t) is fixed. However, for a special case in Section 4.2, the nondecreasing property is not needed and the composition is even continuous.*

An immediate consequence of Assumption 3.3 (4) is given by Lemma 3.5.

Lemma 3.5. *Suppose Assumption 3.3 (4) holds, then $\int V(x_{t+1})P_t(dx_{t+1}|x_t,u)$ is l.s.c. in u for each $x_t \in \mathcal{X}_t$ and Borel measurable function $V \geq 0$.*

Proof. For a nonnegative Borel measurable function V, there exists a sequence V_k of Borel measurable and bounded functions converging increasingly to V. By monotone convergence theorem,

$$\int V(x_{t+1})P_t(dx_{t+1}|x_t,u) = \int \sup_k V_k(x_{t+1})P_t(dx_{t+1}|x_t,u) = \sup_k \int V_k(x_{t+1})P_t(dx_{t+1}|x_t,u),$$

with a fixed x_t. Under Assumption 3.3 (4), it is a supremum of continuous functions in u. Thus, it is l.s.c. in u by Santambrogio (2015, Box 1.5, p.6). $\square$

Assumption 3.3 (4) is crucial and specifies the topology needed for the proof of the existence Theorem 3.6. The usual weak topology considers bounded and continuous functions V, which is not enough to apply the Borel measurable selection theorem 3.1 repeatedly. Assumption 3.3 (4) implies the setwise convergence. It is widely used in the literature of Markov decision processes with average costs; see Schäl (1993, Condition (S)(2)), Hernández-Lerma (1991, Assumption 2.1 (b)), Hernández-Lerma and Lasserre (1999, Assumption 8.3.1 (c)), and Saldi et al. (2018, Assumption 3.1 (b)).

Theorem 3.6. *Suppose Assumption 3.3 holds, then there exists a nonrandomized Markov equilibrium policy $\pi^* = (u_1^*, \ldots, u_{T-1}^*)$.*

Proof. We prove the claim by backward induction. At time $T - 1$, the objective function is

$$\begin{aligned}
\mathcal{L}_T(T-1, x_{T-1}, x_{T-1}, u) =& C_{T-1}(T-1, x_{T-1}, x_{T-1}, u) \\
&+ \int F(T-1, x_{T-1}, x_T)P_{T-1}(dx_T|x_{T-1}, u) \\
&+ G\left(T-1, x_{T-1}, \int H(x_T)P_{T-1}(dx_T|x_{T-1}, u)\right).
\end{aligned}$$

By Lemma 3.5, $\int F(T-1, x_{T-1}, x_T)P_{T-1}(dx_T|x_{T-1}, u)$ and $\int H(x_T)P_{T-1}(dx_T|x_{T-1}, u)$ are l.s.c. in u for each x_{T-1}. Since $G(T-1, x_{T-1}, \cdot)$ is l.s.c. and nondecreasing, the composition $G(T-1, x_{T-1}, \int H(x_T)P_{T-1}(dx_T|x_{T-1}, u))$ is l.s.c. in u when x_{T-1} is fixed. Hence, the objective $\mathcal{L}_T(T-1, x_{T-1}, x_{T-1}, u)$ is l.s.c. in u for each x_{T-1}.

Since $F, G \geq 0$, $\mathcal{L}_T(T-1, x_{T-1}, x_{T-1}, u) \leq r$ implies that $C_{T-1}(T-1, x_{T-1}, x_{T-1}, u) \leq r$. Therefore,

$$\{u \in U_{T-1}(x_{T-1})|\mathcal{L}_T(T-1, x_{T-1}, x_{T-1}, u) \leq r\} \tag{3.5}$$

is a subset of the compact set $U_{T-1}(r, x_{T-1})$ defined in (3.4). Moreover, the subset in (3.5) is closed by the definition of lower semicontinuity. Since closed subsets of compact sets are compact, we have verified the inf-compact condition (3.1). By Lemma 3.1, there exists a Borel measurable minimizer $u_{T-1}^*(x_{T-1})$ for

$$V_{T-1}(x_{T-1}) = \inf_{u \in U_{T-1}(x_{T-1})} \mathcal{L}_T(T-1, x_{T-1}, x_{T-1}, u).$$

Moreover,

$$b_{T-1}(s, y, x_{T-1}) := C_{T-1}(s, y, x_{T-1}, u_{T-1}^*(x_{T-1})),$$

$$f(s, y, x_{T-1}) := \int F(s, y, x_T)P_{T-1}(dx_T|x_{T-1}, u_{T-1}^*(x_{T-1})),$$

$$h(x_{T-1}) := \int H(x_T)P_{T-1}(dx_T|x_{T-1}, u_{T-1}^*(x_{T-1})),$$

6

and $V_{T-1}(x_{T-1})$ are nonnegative Borel measurable functions.

At time $T-2$, the objective is

$$\mathcal{L}_{T-1}(T-2, x_{T-2}, x_{T-2}, u) = C_{T-2}(T-2, x_{T-2}, x_{T-2}, u)$$
$$+ \int b_{T-1}(T-2, x_{T-2}, x_{T-1}) P_{T-2}(dx_{T-1}|x_{T-2}, u)$$
$$+ \int f(T-2, x_{T-2}, x_{T-1}) P_{T-2}(dx_{T-1}|x_{T-2}, u)$$
$$+ G\left(T-2, x_{T-2}, \int h(x_{T-1}) P_{T-2}(dx_{T-1}|x_{T-2}, u)\right).$$

Similarly, $\mathcal{L}_{T-1}(T-2, x_{T-2}, x_{T-2}, u)$ is l.s.c. in u for each x_{T-2}. Hence, we can prove the stated claim backwardly with Lemma 3.1. $\qquad\square$

Lemma 3.7 gives a sufficient condition for Assumption 3.3 (4). The proof is similar to Feinberg and Lewis (2007, Lemma 4.1), while some calculation details are different. We include the proof in the Appendix for the completeness of the paper.

Lemma 3.7. *Let the state spaces $\mathcal{X}_t = \mathbb{R}$. Consider the transition kernels $P_t(dx_{t+1}|x_t, u)$ specified by the following system:*

$$x_{t+1} = \mu(x_t, u_t) + \sigma(x_t, u_t)W_t, \quad t \in \{1, \ldots, T-1\}. \tag{3.6}$$

Suppose

(1) $\{W_t\}_{t=1}^{T-1}$ are i.i.d. random noises with a Borel density function $p(w)$, $w \in (-\infty, \infty)$, with respect to the Lebesgue measure.

(2) The Borel measurable functions $\mu(x_t, u)$ and $\sigma(x_t, u)$ are continuous in u for each $x_t \in \mathbb{R}$.

(3) There exists a constant $\underline{\sigma} > 0$ such that $\sigma(x_t, u) \geq \underline{\sigma}$.

Then $\int V(x_{t+1}) P_t(dx_{t+1}|x_t, u)$ is continuous in u for each $x_t \in \mathbb{R}$ and bounded Borel measurable function V.

Remark 3.8. *(1) If we assume the density $p(w)$ of random noises is continuous, then Lemma 3.7 follows from Scheffé's Theorem (Durrett, 2019, p.117).*

(2) The assumption $\sigma(x_t, u) \geq \underline{\sigma} > 0$ is needed. For example, consider the state following $x_{t+1} = u_t W_t$, where $W_t \sim N(0, 1)$ is a standard normal random variable. The Borel measurable function V satisfies $V(0) = 1$ and $V(x) = 0$ if $x \neq 0$. Then $\int V(x_{t+1}) P_t(dx_{t+1}|x_t, u) = \int V(uw)p(w)dw = 1$ if $u = 0$ and 0 if $u \neq 0$. Hence, it is not l.s.c. at $u = 0$.

4 Examples

4.1 Time-inconsistent linear-quadratic regulator

Consider a time-inconsistent linear-quadratic regulator in Björk and Murgoci (2014, Section 9.4). The state-dependent cost objective is given by

$$J_t(x; \pi^t) = \int \left[\sum_{k=t}^{T-1} u_k^2\right] P(dx_{t+1:T}|x, \pi^t) + \int (x_T - x)^2 P(dx_T|x, \pi^t), \tag{4.1}$$

7

with a scalar state following

$$x_{t+1} = ax_t + bu_t + \sigma W_t,$$

where a and b are known constants. The random noises $\{W_t\}_{t=1}^{T-1}$ are i.i.d. and follow standard normal distribution $N(0,1)$. By Lemma 3.7, Assumption 3.3 (4) is valid. It is also direct to verify other conditions in Assumption 3.3. By Theorem 3.6, a nonrandomized Markov equilibrium policy exists. The explicit solution is given in Björk and Murgoci (2014, Section 9.4).

Besides the objective (4.1), we can also consider nonlinear cost functionals satisfying Assumption 3.3, such as

$$J_t(x; \pi^t) = \int \Big[\sum_{k=t}^{T-1} u_k^2 \Big] P(dx_{t+1:T}|x, \pi^t) + \Big(\int \max\{x_T, 0\} P(dx_T|x, \pi^t) \Big)^2,$$

since h^2 is a continuous and nondecreasing function on $[0, \infty)$.

4.2 Mean-variance portfolio selection

Assumption 3.3 is only a sufficient condition for the existence. In fact, if we can verify that the objective in (2.6) is inf-compact backwardly, then the existence holds.

Consider the mean-variance portfolio selection with the objective

$$J_t(x; \pi^t) = \mathrm{Var}_t[x_T] - \gamma \mathbb{E}_t[x_T] = \int (x_T^2 - \gamma x_T) P(dx_T|x, \pi^t) - \Big(\int x_T P(dx_T|x, \pi^t) \Big)^2 \qquad (4.2)$$

and the wealth (state) process

$$x_{t+1} = Rx_t + u_t Z_t. \qquad (4.3)$$

$R \geq 1$ is a known constant. $\{Z_t\}_{t=1}^{T-1}$ are i.i.d. random variables with mean μ and variance $\sigma^2 > 0$.

With our notations, the objective (4.2) corresponds to

$$F(x_T) = x_T^2 - \gamma x_T, \quad H(x_T) = x_T, \quad G(h) = -h^2.$$

While Assumption 3.3 is not satisfied, we have the following proposition for the inf-compactness and the existence.

Proposition 4.1. *For the objective (4.2) with the state (4.3), the level set (3.3) is inf-compact and thus a nonrandomized Markov equilibrium policy exists.*

Proof. At time $T-1$, it is direct to show the objective is quadratic in u and an equilibrium control u_{T-1}^* is a constant independent of x_{T-1}. To apply backward induction, we assume the equilibrium controls $\{u_{t+1}^*, \ldots, u_{T-1}^*\}$ exist and are constant. Then we calculate the objective at time t. The wealth process yields

$$x_T = R^{T-t-1} x_{t+1} + \sum_{k=t+1}^{T-1} R^{T-1-k} u_k^* Z_k.$$

Therefore,

$$h(x_{t+1}) = \int x_T P(dx_T|x_{t+1}, \pi^{t+1,*}) = R^{T-t-1} x_{t+1} + \sum_{k=t+1}^{T-1} R^{T-1-k} \mu u_k^* =: R^{T-t-1} x_{t+1} + h_{t+1}^0.$$

$$\int x_T^2 P(dx_T | x_{t+1}, \pi^{t+1,*}) = R^{2(T-t-1)} x_{t+1}^2 + 2R^{T-t-1} x_{t+1} \sum_{k=t+1}^{T-1} R^{T-1-k} \mu u_k^*$$

$$+ \sum_{\substack{k \neq l, \\ k,l \in \{t+1,\ldots,T-1\}}} R^{T-1-k} R^{T-1-l} \mu^2 u_k^* u_l^* + \sum_{k=t+1}^{T-1} R^{2(T-1-k)} (u_k^*)^2 (\mu^2 + \sigma^2)$$

$$=: R^{2(T-t-1)} x_{t+1}^2 + a_{t+1}^1 x_{t+1} + a_{t+1}^0.$$

At time t, we have $x_{t+1} = Rx_t + uZ_t$, with the candidate equilibrium control u to be solved. We note that

$$\int f(x_{t+1}) P_t(dx_{t+1} | x_t, u)$$

$$= \int \left[R^{2(T-t-1)} x_{t+1}^2 + a_{t+1}^1 x_{t+1} + a_{t+1}^0 - \gamma (R^{T-t-1} x_{t+1} + h_{t+1}^0) \right] P_t(dx_{t+1} | x_t, u)$$

$$= R^{2(T-t-1)} \left[R^2 x_t^2 + 2Rx_t \mu u + (\mu^2 + \sigma^2) u^2 \right] + (a_{t+1}^1 - \gamma R^{T-t-1})(Rx_t + \mu u)$$

$$+ a_{t+1}^0 - \gamma h_{t+1}^0.$$

$$G\left(\int h(x_{t+1}) P_t(dx_{t+1} | x_t, u) \right) = -(R^{T-t} x_t + R^{T-t-1} \mu u + h_{t+1}^0)^2.$$

Clearly, the objective $\mathcal{L}_{t+1}(x_t, u) = \int f(x_{t+1}) P_t(dx_{t+1} | x_t, u) + G\left(\int h(x_{t+1}) P_t(dx_{t+1} | x_t, u) \right)$ at time t is a quadratic function of u. Since the coefficient of u^2 is $R^{2(T-t-1)} \sigma^2 > 0$, the level set (3.3) is compact. Thus, an equilibrium control u_t^* at time t exists. Indeed, u_t^* is also constant, since coefficients of u^2 and u do not rely on x_t. Then we can prove the existence by backward induction. $\square$

The existence of an equilibrium strategy with the so-called state-dependent risk aversion (Björk and Murgoci, 2014, Section 9.2) can be proved similarly. Indeed, these two cases have explicit equilibrium policies.

4.3 Non-exponential discounting

Consider the portfolio selection with non-exponential discounting and the exponential utility:

$$J_t(x; \pi^t) = \int \varphi(T - t) \frac{e^{-\gamma x_T}}{\gamma} P(dx_T | x, \pi^t), \tag{4.4}$$

where $\gamma > 0$ is the risk aversion parameter and $\varphi(T - t) > 0$ is the non-exponential discounting function. The wealth process is still given by (4.3). The following existence result is directly from Theorem 3.6 and Lemma 3.7.

Proposition 4.2. *Suppose the random variables Z_t in (4.3) satisfy Condition (1) in Lemma 3.7. Moreover, assume the control $0 < \underline{u} \leq u_t \leq \bar{u} < \infty$ is in a compact interval. Then there exists a nonrandomized Markov equilibrium policy for (4.4).*

Acknowledgment

References

Basak, S. and Chabakauri, G. (2010). Dynamic mean-variance asset allocation. *The Review of Financial Studies*, 23(8):2970–3016.

Bayraktar, E. and Han, B. (2023). Equilibrium transport with time-inconsistent costs: An application to matching problems in the job market. *arXiv preprint arXiv:2302.01498*.

Bayraktar, E., Huang, Y.-J., Wang, Z., and Zhou, Z. (2023). Relaxed equilibria for time-inconsistent markov decision processes. *arXiv preprint arXiv:2307.04227*.

Bayraktar, E., Zhang, J., and Zhou, Z. (2021). Equilibrium concepts for time-inconsistent stopping problems in continuous time. *Mathematical Finance*, 31(1):508–530.

Bertsekas, D. and Shreve, S. E. (1978). *Stochastic optimal control: The discrete-time case*. Academic Press.

Björk, T., Khapko, M., and Murgoci, A. (2021). *Time-inconsistent control theory with finance applications*, volume 732. Springer.

Björk, T. and Murgoci, A. (2014). A theory of Markovian time-inconsistent stochastic control in discrete time. *Finance and Stochastics*, 18(3):545–592.

Brown, L. D. and Purves, R. (1973). Measurable selections of extrema. *The Annals of Statistics*, pages 902–912.

Durrett, R. (2019). *Probability: Theory and examples*, volume 49. Cambridge University Press.

Feinberg, E. A. and Lewis, M. E. (2007). Optimality inequalities for average cost Markov decision processes and the stochastic cash balance problem. *Mathematics of Operations Research*, 32(4):769–783.

Hernández-Lerma, O. (1991). Average optimality in dynamic programming on Borel spaces—unbounded costs and controls. *Systems & Control Letters*, 17(3):237–242.

Hernández-Lerma, O. and Lasserre, J. B. (1996). *Discrete-time Markov control processes: Basic optimality criteria*, volume 30. Springer Science & Business Media.

Hernández-Lerma, O. and Lasserre, J. B. (1999). *Further topics on discrete-time Markov control processes*, volume 42. Springer Science & Business Media.

Huang, Y.-J. and Zhou, Z. (2021). Strong and weak equilibria for time-inconsistent stochastic control in continuous time. *Mathematics of Operations Research*, 46(2):428–451.

Jaśkiewicz, A. and Nowak, A. S. (2021). Markov decision processes with quasi-hyperbolic discounting. *Finance and Stochastics*, 25(2):189–229.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Rieder, U. (1978). Measurable selection theorems for optimization problems. *Manuscripta Mathematica*, 24(1):115–131.

Saldi, N., Linder, T., and Yüksel, S. (2018). *Finite approximations in discrete-time stochastic control.* Springer.

Santambrogio, F. (2015). Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63).

Schäl, M. (1993). Average optimality in dynamic programming with general state space. *Mathematics of Operations Research*, 18(1):163–172.

Strotz, R. (1955). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23(3):165–180.

A   Proof of Lemma 3.7

Fix $x_t \in \mathbb{R}$. Consider a sequence u^k converging to u. For notational simplicity, denote $\mu^k := \mu(x_t, u^k)$, $\mu := \mu(x_t, u)$, $\sigma^k := \sigma(x_t, u^k)$, and $\sigma := \sigma(x_t, u)$. Under condition (2), we have $\mu^k \to \mu$ and $\sigma^k \to \sigma$, when $k \to \infty$.

A change of variables gives

$$\int V(x_{t+1})P_t(dx_{t+1}|x_t, u^k) = \int_{-\infty}^{\infty} V(\mu^k + \sigma^k w)p(w)dw = \int_{-\infty}^{\infty} V(z)p\Big(\frac{z - \mu^k}{\sigma^k}\Big)\frac{dz}{\sigma^k}.$$

Then

$$\left| \int V(x_{t+1})P_t(dx_{t+1}|x_t, u^k) - \int V(x_{t+1})P_t(dx_{t+1}|x_t, u) \right|$$

$$\leq \left| \int_{-\infty}^{\infty} V(z)p\Big(\frac{z - \mu^k}{\sigma^k}\Big)\frac{dz}{\sigma^k} - \int_{-\infty}^{\infty} V(z)p\Big(\frac{z - \mu}{\sigma}\Big)\frac{dz}{\sigma} \right|$$

$$\leq M \int_{-\infty}^{\infty} \left| p\Big(\frac{z - \mu^k}{\sigma^k}\Big)\frac{1}{\sigma^k} - p\Big(\frac{z - \mu}{\sigma}\Big)\frac{1}{\sigma} \right| dz, \tag{A.1}$$

where the second inequality uses the boundedness of $|V| \leq M$. Denote $y = (z - \mu)/\sigma$, then

$$\int_{-\infty}^{\infty} \left| p\Big(\frac{z - \mu^k}{\sigma^k}\Big)\frac{1}{\sigma^k} - p\Big(\frac{z - \mu}{\sigma}\Big)\frac{1}{\sigma} \right| dz = \int_{-\infty}^{\infty} \left| p\Big(\frac{\sigma y + \mu - \mu^k}{\sigma^k}\Big)\frac{1}{\sigma^k} - p(y)\frac{1}{\sigma} \right| \sigma dy.$$

Note that $\sigma > 0$ and can be moved into the absolute value.

For simplicity, denote $\tilde{\sigma}^k := \sigma/\sigma^k$ and $\tilde{\mu}^k := (\mu - \mu^k)/\sigma^k$. For given $0 < \delta < 1$ and $\delta' > 0$, choose an integer k large enough such that

$$|\tilde{\sigma}^k - 1| \leq \delta, \quad |\tilde{\mu}^k| \leq \delta'.$$

For a fixed $\varepsilon > 0$, consider $K > 0$ such that $\int_{-K}^{K} p(y)dy \geq 1 - \varepsilon/8$. Denote a constant $K^* := (K + \delta')/(1 - \delta)$. We have

$$\int_{K^*}^{\infty} \left| p\Big(\frac{\sigma y + \mu - \mu^k}{\sigma^k}\Big)\frac{1}{\sigma^k} - p(y)\frac{1}{\sigma} \right| \sigma dy \leq \int_{K^*}^{\infty} p(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k dy + \int_{K^*}^{\infty} p(y)dy$$

$$\leq \int_{\tilde{\sigma}^k K^* + \tilde{\mu}^k}^{\infty} p(y')dy' + \int_{K^*}^{\infty} p(y)dy$$

$$\leq 2\int_{K}^{\infty} p(y)dy,$$

11

where the last inequality used the fact that $\max\{\tilde{\sigma}^k K^* + \tilde{\mu}^k, K^*\} \geq (1-\delta)K^* - \delta' = K$, by the definition of K^*. Similarly, we have

$$\int_{-\infty}^{-K^*} \left| p\Big(\frac{\sigma y + \mu - \mu^k}{\sigma^k}\Big)\frac{1}{\sigma^k} - p(y)\frac{1}{\sigma}\right| \sigma dy \leq 2 \int_{-\infty}^{-K} p(y)dy.$$

Hence, noting the property of K, we obtain

$$\int_{-\infty}^{\infty} \left| p\Big(\frac{z-\mu^k}{\sigma^k}\Big)\frac{1}{\sigma^k} - p\Big(\frac{z-\mu}{\sigma}\Big)\frac{1}{\sigma}\right| dz \leq \int_{-K^*}^{K^*} \left| p(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k - p(y)\right| dy + \frac{\varepsilon}{4}. \tag{A.2}$$

Next, we choose a constant $G > 0$, such that $\int_{-\infty}^{\infty} p(y)\mathbf{1}_{\{p(y)>G\}}dy \leq \varepsilon/8$. Define

$$p_G(y) = p(y)\mathbf{1}_{\{p(y)\leq G\}} + G\mathbf{1}_{\{p(y)>G\}},$$

which is bounded. It yields

$$\int_{-K^*}^{K^*} \left| p(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k - p(y)\right| dy - \int_{-K^*}^{K^*} \left| p_G(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k - p_G(y)\right| dy$$

$$\leq \int_{-K^*}^{K^*} \left| p(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k - p_G(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k\right| dy + \int_{-K^*}^{K^*} \left| p(y) - p_G(y)\right| dy$$

$$\leq 2 \int_{-\infty}^{\infty} \left| p(y) - p_G(y)\right| dy = 2 \int_{-\infty}^{\infty} (p(y) - G)\mathbf{1}_{\{p(y)>G\}}dy$$

$$\leq 2 \int_{-\infty}^{\infty} p(y)\mathbf{1}_{\{p(y)>G\}}dy \leq \varepsilon/4. \tag{A.3}$$

Finally, we apply Lusin's theorem to $p(y)$, which is Borel measurable on $\mathbb{R}$. There exists a continuous function $q(y)$, such that the Lebesgue measure of the set $\{p(y) \neq q(y)\}$ is not greater than $\varepsilon/(8G)$. Define

$$q_G(y) = q(y)\mathbf{1}_{\{0\leq q(y)\leq G\}} + G\mathbf{1}_{\{q(y)>G\}}.$$

Obviously, if $p(y) = q(y)$, then $p_G(y) = q_G(y)$. It implies that $\{p_G(y) \neq q_G(y)\} \subseteq \{p(y) \neq q(y)\}$. As a consequence,

$$\int_{-K^*}^{K^*} \left| p_G(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k - p_G(y)\right| dy$$

$$\leq \int_{-K^*}^{K^*} \left| q_G(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k - q_G(y)\right| dy + \int_{-K^*}^{K^*} \left| p_G(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k - q_G(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k\right| dy$$

$$+ \int_{-K^*}^{K^*} \left| p_G(y) - q_G(y)\right| dy$$

$$\leq \int_{-K^*}^{K^*} \left| q_G(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k - q_G(y)\right| dy + 2 \int_{-\infty}^{\infty} \left| p_G(y) - q_G(y)\right| dy$$

$$\leq \int_{-K^*}^{K^*} \left| q_G(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k - q_G(y)\right| dy + 2 \int_{-\infty}^{\infty} G\mathbf{1}_{\{p_G(y)\neq q_G(y)\}}dy$$

$$\leq \int_{-K^*}^{K^*} \left| q_G(\tilde{\sigma}^k y + \tilde{\mu}^k)\tilde{\sigma}^k - q_G(y)\right| dy + \varepsilon/4. \tag{A.4}$$

We can regard $q_G(ay + b)a$ as a function of three variables (a, b, y). Clearly, $q_G(ay + b)a$ is jointly continuous in (a, b, y), since $q_G(\cdot)$ is continuous. By the uniform continuity on the compact set $[1 - \delta, 1 + \delta] \times [-\delta', \delta'] \times [-K^*, K^*]$, there exists an integer $N > 0$ such that

$$\sup_{k \geq N, y \in [-K^*, K^*]} \left| q_G(\tilde{\sigma}^k y + \tilde{\mu}^k) \tilde{\sigma}^k - q_G(y) \right| \leq \varepsilon/(8K^*). \tag{A.5}$$

Combining (A.1), (A.2), (A.3), (A.4), and (A.5), we obtain

$$\left| \int V(x_{t+1}) P_t(dx_{t+1} | x_t, u^k) - \int V(x_{t+1}) P_t(dx_{t+1} | x_t, u) \right| \leq M\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the claim follows. $\qquad\square$