Classical Risk-Averse Control for a Finite-Horizon Borel Model

Margaret P. Chapman, *Member, IEEE*, and Kevin M. Smith

***Abstract*—We study a risk-averse optimal control problem for a finite-horizon Borel model, where a cumulative cost is assessed via exponential utility. The setting permits non-linear dynamics, non-quadratic costs, and continuous state and control spaces but is less general than the problem of optimizing an expected utility. Our contribution is to show the existence of an optimal risk-averse controller without using state space augmentation and therefore offer a simpler solution method from first principles compared to what is currently available in the literature.**

***Index Terms*—Stochastic optimal control, exponential utility, Markov processes.**

I. Introduction

RECENTLY, there has been a renewed interest in risk-sensitive control for various applications, including robotics [1], [2], remote state estimation [3], and building evacuation [4]. A classical risk-sensitive control approach is to assess a random cost using the *exponential utility* functional. This functional takes the form $\rho_\theta(G) := \frac{-2}{\theta} \log E(e^{\frac{-\theta}{2}G})$, where θ is a parameter and G is a non-negative random variable. If $\theta < 0$, then large values of G are exaggerated through the exponential transformation, which represents a risk-averse perspective. In contrast, the case of $\theta > 0$ represents a risk-seeking perspective. It can be shown that $\rho_\theta(G) \approx E(G) - \frac{\theta}{4}\text{var}(G)$ approximates a weighted sum of the mean and variance of G under appropriate conditions, including $|\theta|$ being small [5]. Exponential-utility optimal control has been studied since the 1970s, and we first summarize early work.

In 1972, Howard and Matheson studied the optimization of an exponential utility criterion for a discrete-time Markov decision process (MDP) with finitely many states [6]. In 1973, Jacobson considered the problem of optimizing the exponential utility of a quadratic cost for a discrete-time linear system with Euclidean state and control spaces subject to additive Gaussian noise [7]. This problem is often called linear-exponential-quadratic-Gaussian (LEQG) or linear-exponential-quadratic-regulator (LEQR) control. LEQG theory was generalized by Whittle, in particular to the case of partially observed states, in the 1980s and 1990s; see [8] and the references therein. Relations between controllers satisfying an $\mathcal{H}_\infty$-norm bound and the infinite-time LEQG controller were analyzed in the late 1980s [9], [10]. In 1999, di Masi and Stettner studied MDPs on continuous state spaces and discrete infinite-time horizons with exponential utility criteria [11]. In the early 2000s, relations between robust model predictive control (MPC) and MPC with an exponential utility criterion were investigated in the linear-quadratic setting [12, Ch. 8.3]. The exponential utility criterion belongs to the broader family of *expected utility* criteria, which measure risk by transforming a random cost based on a user's subjective preferences.

Additional methods for quantifying and optimizing risk are presented by [13, Ch. 6], [14], for example. Specifically, Ruszczyński considered the optimization of a *nested risk functional* for a discrete-time Borel-space MDP [14], and related formulations have been studied, e.g., see [15], [16]. A nested risk functional takes the form $\rho_1(Z_1 + \rho_2(Z_2 + \cdots + \rho_{N-1}(Z_{N-1} + \rho_N(Z_N)) \cdots))$, where Z_i is a random variable and ρ_i is a mapping between spaces of random variables [14]. This functional is not straightforward to interpret, but it can be optimized using dynamic programming (DP) on the state space. In 2021, connections between nested risk functionals and distributionally robust MDPs were drawn [16], and similar ideas were suggested earlier, e.g., see [17, eq. (4.11)].

Another approach to risk-sensitive control is to optimize an expected cumulative cost subject to a risk constraint, for instance, see [18]–[22]. In particular, Tsiamis *et al.* considered a variance-like constraint in a linear-quadratic setting [22], whereas References [18], [20], [21] considered *Conditional Value-at-Risk* (CVaR) constraints. In contrast to expected utility criteria, CVaR is a quantile-based measure that quantifies an average cost in a fraction of worst cases. In prior work, we developed a controller using a CVaR objective [23] and a safety analysis framework using CVaR [24]. Risk-constrained MDPs and MDPs with expected utility or CVaR criteria were studied by [19] in an infinite-time setting using occupation measures and state space augmentation. State space augmentation is a technique for tracking history-dependent information that may be needed to characterize the stage-wise sub-problems of a multi-stage optimization problem. This technique is not needed when a sub-problem can be written in terms of the current state rather than the current and prior

Manuscript received July 5, 2021; revised August 26, 2021; accepted September 13, 2021. Date of publication September 21, 2021; date of current version October 21, 2021. The work of Margaret P. Chapman was supported by the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto. The work of Kevin M. Smith was supported in part by the U.S. National Science Foundation under Grant NSF-NRT 2021874. Recommended by Senior Editor V. Ugrinovskii. *(Corresponding author: Margaret P. Chapman.)*

Margaret P. Chapman is with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G8, Canada (e-mail: mchapman@ece.utoronto.ca).

Kevin M. Smith is with the Department of Civil and Environmental Engineering, Tufts University, Medford, MA 02155 USA (e-mail: kevin.smith@tufts.edu).

Digital Object Identifier 10.1109/LCSYS.2021.3114126

states. More details about risk-sensitive MDPs can be found in [16], for example.

While these maturing approaches to risk-sensitive control are intriguing, this letter concerns exponential utility, which remains popular in the literature. Specifically, the focus here is optimizing exponential utility for an MDP on a discrete finite-time horizon with Borel state and control spaces, which we call Problem A for convenience. Problem A permits non-Euclidean spaces, non-linear dynamics, non-quadratic costs, and non-Gaussian noise, and therefore generalizes the LEQG setting. In 2014, Bäuerle and Rieder developed the state-of-the-art approach for solving Problem A and a broader class of MDP problems with expected utility criteria [25]. However, the methodology employs an augmented state space, in which an extra state records the cumulative cost thus far [25, Th. 1, Corollary 1]. Our contribution is to demonstrate that the additional complexity of state space augmentation is not required to solve Problem A, as we provide an alternate solution pathway using first principles from measure theory and real analysis.

There are key distinctions between our paper and the existing literature on MDPs with exponential utility criteria, which we outline below. Many works concern MDPs on countable state spaces, see [26]–[30] for some examples, whereas we consider the more general setting of Borel spaces. As mentioned previously, the proof of [25, Th. 1, Corollary 1] uses a Bellman equation that is defined on an augmented state space, whereas our approach does not use an augmented state space. Numerous papers about discrete-time Borel-space MDPs examine infinite-time settings, e.g., [11], [31]–[36], which naturally require different techniques compared to our work. di Masi and Stettner studied infinite-time problems in which the stage cost is continuous using a span contraction approach and two discounting approaches [11], [31]. Later, di Masi and Stettner generalized their work by approximating an MDP with one that is uniformly ergodic [32]. In contrast, our paper concerns a finite-time horizon and lower semi-continuous costs, and thus requires measure-theoretic arguments that are properly adapted to this setting. Jaśkiewicz and colleagues examined infinite-time problems using a vanishing discount factor approach [33], MDPs in which the transition kernel only depends on the current control [34], and later using the Banach fixed point theorem [35]. Anantharam and Borkar studied an infinite-time reward maximization problem using occupation measures [36]. While we focus on the discrete-time case, we note that continuous-time MDPs with exponential utility criteria have been examined by [37]–[40], for example.

Notation: If $\mathcal{M}$ is a metrizable space, $\mathcal{B}_\mathcal{M}$ is the Borel sigma algebra on $\mathcal{M}$. $\mathcal{P}(D)$ is the set of probability measures on $(D, \mathcal{B}_D)$ with the weak topology, where D is a Borel space. Capital letters denote random objects, while lower-case letters denote the associated values; e.g., x_t is a value of X_t. We define $\mathbb{T} := \{0, 1, \ldots, N-1\}$ and $\mathbb{T}_N := \{0, 1, \ldots, N\}$, where $N \in \mathbb{N}$ is given. $\mathbb{R}^* := \mathbb{R} \cup \{-\infty, +\infty\}$ is the extended real line. We abbreviate lower semi-continuous as lsc.

II. Problem Statement

Let S, A, and D be Borel spaces of states, controls, and disturbances, respectively. Consider a system on a discrete finite-time horizon of length $N \in \mathbb{N}$ of the form

$$x_{t+1} = f_t(x_t, u_t, w_t) \quad \forall t \in \mathbb{T}, \tag{1}$$

where $x_t \in S$, $u_t \in A$, and $w_t \in D$ are values of the random state X_t, the random control U_t, and the random disturbance W_t, respectively. The initial state X_0 is fixed at an arbitrary initial condition $x \in S$. The dynamics function $f_t : S \times A \times D \to S$ is Borel measurable. Given (X_t, U_t), the disturbance W_t is conditionally independent of W_s for all $s \neq t$. The distribution of W_t, $p_t(\mathrm{d}w_t|x_t, u_t)$, is a Borel-measurable stochastic kernel on D given $S \times A$. That is, the function $\gamma_t : S \times A \to \mathcal{P}(D)$ defined by $\gamma_t(x_t, u_t) := p_t(\mathrm{d}w_t|x_t, u_t)$ is Borel measurable. If $(x_t, u_t) \in S \times A$ is the value of (X_t, U_t), then the distribution of X_{t+1} is given by

$$q_t(B|x_t, u_t) := p_t\big(\{w_t \in D : f_t(x_t, u_t, w_t) \in B\}\big|x_t, u_t\big) \tag{2}$$

for all $B \in \mathcal{B}_S$ and $t \in \mathbb{T}$. We consider the class of deterministic Markov policies Π. Each $\pi \in \Pi$ takes the form $\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1})$ such that $\mu_t : S \to A$ is Borel measurable for each $t \in \mathbb{T}$.

We aim to define and optimize a random cost (where we specify the precise notion of optimality later). For this task, we are required to define a probability space $(\Omega, \mathcal{B}_\Omega, P_x^\pi)$, which is parametrized by an initial condition $x \in S$ and a policy $\pi \in \Pi$. The sample space Ω is defined by $\Omega := (S \times A)^N \times S$. That is, an element $\omega = (x_0, u_0, \ldots, x_{N-1}, u_{N-1}, x_N) \in \Omega$ is a value of the random trajectory $(X_0, U_0, \ldots, X_{N-1}, U_{N-1}, X_N)$. The coordinates of ω have causal dependencies due to the form of the dynamics (1) and the class of policies Π. The distribution of the random trajectory is given by a probability measure P_x^π on $(\Omega, \mathcal{B}_\Omega)$. The form of P_x^π allows us to define a DP recursion for computing an optimal policy under certain conditions (to be specified).

Let δ_x denote the Dirac measure on $(S, \mathcal{B}_S)$ concentrated at x. With slight abuse of notation, let $\delta_{\mu_t(x_t)}$ denote the Dirac measure on $(A, \mathcal{B}_A)$ concentrated at $\mu_t(x_t)$, where $x_t \in S$ is a value of X_t. Let $B \in \mathcal{B}_\Omega$ be a measurable rectangle, i.e., $B = B_{X_0} \times B_{U_0} \times B_{X_1} \times B_{U_1} \times \cdots \times B_{X_N}$, where $B_{X_i} \in \mathcal{B}_S$ for all $i \in \mathbb{T}_N$ and $B_{U_j} \in \mathcal{B}_A$ for all $j \in \mathbb{T}$. Then, we have

$$\begin{aligned} P_x^\pi(B) = \int_{B_{X_0}} \int_{B_{U_0}} \int_{B_{X_1}} \int_{B_{U_1}} \cdots \int_{B_{X_N}} q_{N-1}(\mathrm{d}x_N|x_{N-1}, u_{N-1}) \\ \cdots \delta_{\mu_1(x_1)}(\mathrm{d}u_1)\, q_0(\mathrm{d}x_1|x_0, u_0)\, \delta_{\mu_0(x_0)}(\mathrm{d}u_0)\, \delta_x(\mathrm{d}x_0). \end{aligned} \tag{3}$$

The nested integrals should be taken from "the inside to the outside." The integral with respect to $x_N \in S$ is taken over the set B_{X_N}; the integral with respect to $u_1 \in A$ is taken over B_{U_1}, etc. The reader may refer to [41, Proposition C.10, Remark C.11, p. 178] or [42, Proposition 7.28, pp. 140–141] for details.

If $G : \Omega \to \mathbb{R}^*$ is Borel measurable, then the expectation of G with respect to P_x^π, $E_x^\pi(G)$, is defined by

$$\begin{aligned} &\textstyle\int_\Omega G(\omega)\, \mathrm{d}P_x^\pi(\omega) \\ &:= \int_S \int_A \int_S \int_A \cdots \int_S G(x_0, u_0, \ldots, x_{N-1}, u_{N-1}, x_N) \\ &\quad q_{N-1}(\mathrm{d}x_N|x_{N-1}, u_{N-1}) \cdots \delta_{\mu_1(x_1)}(\mathrm{d}u_1) \\ &\quad q_0(\mathrm{d}x_1|x_0, u_0)\, \delta_{\mu_0(x_0)}(\mathrm{d}u_0)\, \delta_x(\mathrm{d}x_0). \end{aligned} \tag{4}$$

The expectation $E_x^\pi(G)$ exists (i.e., does not take the form $+\infty - \infty$), if G is bounded or non-negative, for instance. G is an extended random variable on $(\Omega, \mathcal{B}_\Omega, P_x^\pi)$ for each $x \in S$ and $\pi \in \Pi$.

We consider a particular random variable Z, representing a cost, that is incurred as the system operates over time. For any value $\omega = (x_0, u_0, \ldots, x_{N-1}, u_{N-1}, x_N) \in \Omega$ of the random trajectory, we define

$$Z(\omega) := \sum_{t=0}^{N-1} c_t(x_t, u_t) + c_N(x_N), \tag{5}$$

where $c_t : S \times A \to \mathbb{R}$ and $c_N : S \to \mathbb{R}$ are Borel measurable and bounded. In (5), x_t is the value of X_t and u_t is the value of U_t associated with the trajectory value ω.

A standard (risk-neutral) approach to manage Z is to minimize its expectation $E_x^\pi(Z)$ over the class of policies Π. An alternative approach is to use the (risk-averse) *exponential utility* functional. Let $\theta \in \Theta \subseteq (-\infty, 0)$ be given, and define the optimal value function $V_\theta^* : S \to \mathbb{R}^*$ as follows:

$$V_\theta^*(x) := \inf_{\pi \in \Pi} \frac{-2}{\theta} \log E_x^\pi \left(e^{\frac{-\theta}{2} Z} \right). \tag{6}$$

If there is a policy $\pi_\theta^* \in \Pi$ such that $V_\theta^*(x) = \frac{-2}{\theta} \log E_x^{\pi_\theta^*}(e^{\frac{-\theta}{2} Z})$ for all $x \in S$, we say that π_θ^* is *optimal for* V_θ^*. The next assumption ensures that such a policy exists.

Assumption 1 (Measurable Selection): We assume that

1) $p_t(\mathrm{d}w_t | x_t, u_t)$ is a continuous stochastic kernel for all $t \in \mathbb{T}$. That is, the function $\gamma_t : S \times A \to \mathcal{P}(D)$ defined by $\gamma_t(x_t, u_t) := p_t(\mathrm{d}w_t | x_t, u_t)$ is continuous.
2) f_t is continuous for all $t \in \mathbb{T}$. c_t is lower semi-continuous and bounded for all $t \in \mathbb{T}_N$.
3) The set of controls A is compact.

Remark 1 (Justification of Assumption 1): Assumption 1 is an example of a measurable selection condition. Such conditions are standard in stochastic control problems on Borel spaces, e.g., see [42, Definition 8.7], [41, Sec. 3.3]. In a risk-neutral problem, which optimizes $E_x^\pi(Z)$, it is common to assume that c_t is only bounded below. However, assuming that c_t is bounded simplifies arguments for risk-sensitive MDPs, e.g., see [19], [25], [35]. c_t being bounded ensures that Z is bounded, which implies that V_θ^* is finite for all $\theta \in \Theta$.

III. Dynamic Programming Algorithm

Next, we provide a DP algorithm for V_θ^*.

Algorithm 1 (DP for V_θ^*): For any $\theta \in \Theta$, define $V_t^\theta : S \to \mathbb{R}^*$ recursively, $V_N^\theta(x) := c_N(x)$ and for $t = N-1, \ldots, 1, 0$,

$$V_t^\theta(x) := \inf_{u \in A} v_{t+1}^\theta(x, u), \tag{7a}$$

where $v_{t+1}^\theta : S \times A \to \mathbb{R}^*$ is defined by

$$v_{t+1}^\theta(x, u) := c_t(x, u) + \psi_t^\theta(x, u),$$
$$\psi_t^\theta(x, u) := \tfrac{-2}{\theta} \log\left(\textstyle\int_D e^{\frac{-\theta}{2} V_{t+1}^\theta(f_t(x,u,w))}\, p_t(\mathrm{d}w | x, u)\right). \tag{7b}$$

Algorithm 1 is a backwards recursion that applies an exponential transformation to the cost-to-go V_{t+1}^θ and resembles existing formulations, e.g., see [11, eq. (2.3)], [25, Remark 1], and [35, eq. (1.1)]. Our contribution is not the algorithm itself but instead the direct pathway that we follow to solve the risk-averse control problem of interest. Namely, we use first principles from real analysis and measure theory and build on arguments from [42], [43] to prove two theorems. Theorem 1 shows that V_t^θ is lower semi-continuous (lsc) and bounded, and there is a Borel-measurable function $\mu_t^\theta : S \to A$ such that

$$V_t^\theta(x) = v_{t+1}^\theta(x, \mu_t^\theta(x)) \quad \forall x \in S. \tag{8}$$

In Theorem 2, we show that the (non-unique) policy $\pi_\theta^* := (\mu_0^\theta, \mu_1^\theta, \ldots, \mu_{N-1}^\theta)$ is optimal for V_θ^* and $V_0^\theta = V_\theta^*$.

IV. Analysis of Dynamic Programming Iterates

In this section, we first prove Proposition 1, which shows that the composition of a continuous increasing function and a bounded lsc function is lsc. Then, we use Proposition 1 and another preliminary result (Lemma 2, Appendix) to analyze the DP iterates $V_0^\theta, V_1^\theta, \ldots, V_N^\theta$ in Theorem 1.

Proposition 1 (Comp. of con't, inc. and lsc): Let $\mathcal{M}$ be a metrizable space. Assume that $\mathcal{Y}_i := (a_i, b_i) \subseteq \mathbb{R}$ is non-empty for $i = 1, 2$. Suppose that $\kappa_1 : \mathcal{Y}_1 \to \mathcal{Y}_2$ is continuous and increasing, and $\kappa_2 : \mathcal{M} \to \mathcal{Y}_1$ is lsc and bounded. (There are scalars $\underline{c}$ and $\overline{c}$ such that $[\underline{c}, \overline{c}] \subset \mathcal{Y}_1$ and $\underline{c} \leq \kappa_2(y) \leq \overline{c}$ for all $y \in \mathcal{M}$.) Then, $\kappa_1 \circ \kappa_2 : \mathcal{M} \to \mathcal{Y}_2$ is lsc.

Proof: To show that $\kappa_1 \circ \kappa_2$ is lsc, we must show that $\liminf_{i \to \infty} \kappa_1(\kappa_2(x^i)) \geq \kappa_1(\kappa_2(x))$, where $\{x^i\}_{i=1}^\infty$ is a sequence in $\mathcal{M}$ converging to $x \in \mathcal{M}$.[1] Since $\{x^i\}_{i=1}^\infty$ converges to x and κ_2 is lsc, it holds that

$$\lim_{i \to \infty} \inf_{k \geq i} \kappa_2(x^k) := \liminf_{i \to \infty} \kappa_2(x^i) \geq \kappa_2(x). \tag{9}$$

Since κ_2 is bounded below by $\underline{c}$ and above by $\overline{c}$, we have

$$\underline{c} \leq \inf_{k \geq i} \kappa_2(x^k) \leq \inf_{k \geq i+1} \kappa_2(x^k) \leq \overline{c} \quad \forall i \in \mathbb{N}, \tag{10}$$

which implies that $\underline{c} \leq \lim_{i \to \infty} \inf_{k \geq i} \kappa_2(x^k) \leq \overline{c}$. Since $\{\inf_{k \geq i} \kappa_2(x^k)\}_{i=1}^\infty$ is a sequence in $[\underline{c}, \overline{c}]$ which converges to a point in $[\underline{c}, \overline{c}]$, $[\underline{c}, \overline{c}]$ is a non-empty subset of $\mathcal{Y}_1$, and κ_1 is continuous on $\mathcal{Y}_1$, we find that

$$\kappa_1\left(\lim_{i \to \infty} \inf_{k \geq i} \kappa_2(x^k)\right) = \lim_{i \to \infty} \kappa_1\left(\inf_{k \geq i} \kappa_2(x^k)\right). \tag{11}$$

Since κ_1 is increasing and by (9), we have

$$\kappa_1\left(\lim_{i \to \infty} \inf_{k \geq i} \kappa_2(x^k)\right) \geq \kappa_1(\kappa_2(x)). \tag{12}$$

Moreover, since κ_1 is increasing, for any $i \in \mathbb{N}$, it holds that

$$\forall k \geq i, \quad \kappa_1(\kappa_2(x^k)) \geq \kappa_1\left(\inf_{k \geq i} \kappa_2(x^k)\right). \tag{13}$$

Thus, $\kappa_1(\inf_{k \geq i} \kappa_2(x^k)) \in \mathbb{R}$ is a lower bound for the set $\{\kappa_1(\kappa_2(x^k)) : k \geq i\}$, which implies that

$$\inf_{k \geq i} \kappa_1(\kappa_2(x^k)) \geq \kappa_1\left(\inf_{k \geq i} \kappa_2(x^k)\right). \tag{14}$$

By letting i tend to infinity, it holds that

$$\lim_{i \to \infty} \inf_{k \geq i} \kappa_1(\kappa_2(x^k)) \geq \lim_{i \to \infty} \kappa_1\left(\inf_{k \geq i} \kappa_2(x^k)\right), \tag{15}$$

which is equivalent to

$$\liminf_{i \to \infty} \kappa_1(\kappa_2(x^i)) \geq \kappa_1\left(\lim_{i \to \infty} \inf_{k \geq i} \kappa_2(x^k)\right) \tag{16}$$

by (11). Finally, by (12), we derive the desired result. ■

Next, we use Proposition 1 to prove Theorem 1.

[1] A key aspect of the proof of Proposition 1 is the use of the bounds $\underline{c}$ and $\overline{c}$ to guarantee that a limit inferior is in the domain of κ_1. This and the continuity of κ_1 allow us to exchange the order of a limit and κ_1.

Theorem 1 (Properties of V_t^θ): Assume Assumption 1. For all $t \in \mathbb{T}_N$, V_t^θ is lsc and bounded. For all $t \in \mathbb{T}$, there is a Borel-measurable function $\mu_t^\theta : S \to A$ such that (8) holds.

Proof: By induction. For brevity, denote $\mathcal{S} := S \times A \times D$. $V_N^\theta = c_N$ is lsc and bounded by Assumption 1. Now, suppose (the induction hypothesis) that for some $t \in \mathbb{T}$, V_{t+1}^θ is lsc and bounded. The key step is to show that ψ_t^θ (7b) is lsc and bounded.[2] Boundedness of ψ_t^θ follows from boundedness of V_{t+1}^θ. For showing boundedness, note that the function $\phi_t : \mathcal{S} \to \mathbb{R}$ defined by

$$\phi_t(x, u, w) := e^{\frac{-\theta}{2} V_{t+1}^\theta(f_t(x,u,w))} \tag{17}$$

is non-negative and bounded. (We drop the superscript θ on the left-hand-side for brevity.) Also, for any $(x, u) \in S \times A$, the function $\phi_t(x, u, \cdot) : D \to \mathbb{R}$ is Borel measurable because it is a composition of Borel-measurable functions. In addition, since $p_t(\mathrm{d}w|x, u)$ is a probability measure on $(D, \mathcal{B}_D)$, it follows that the (Lebesgue) integral

$$\phi_t'(x, u) := \int_D \phi_t(x, u, w)\, p_t(\mathrm{d}w|x, u) \tag{18}$$

exists and is finite for all $(x, u) \in S \times A$.

To complete the proof, we show that ψ_t^θ (7b) is lsc. Since f_t is continuous by Assumption 1, V_{t+1}^θ is lsc by the induction hypothesis, and $\frac{-\theta}{2} > 0$, the function $g_t : \mathcal{S} \to \mathbb{R}$ defined by

$$g_t(x, u, w) := \frac{-\theta}{2} V_{t+1}^\theta(f_t(x, u, w)) \tag{19}$$

is lsc. Indeed, let $\{(x^i, u^i, w^i)\}_{i=1}^\infty$ be a sequence in $\mathcal{S}$ converging to $(x, u, w) \in \mathcal{S}$. To show that $V_{t+1}^\theta \circ f_t$ is lsc, we must prove that

$$\liminf_{i\to\infty} V_{t+1}^\theta(f_t(x^i, u^i, w^i)) \geq V_{t+1}^\theta(f_t(x, u, w)). \tag{20}$$

Since $\{(x^i, u^i, w^i)\}_{i=1}^\infty$ converges to (x, u, w) and f_t is continuous, $\{f_t(x^i, u^i, w^i)\}_{i=1}^\infty$ converges to $f_t(x, u, w)$. Since the latter is a converging sequence in S and $V_{t+1}^\theta : S \to \mathbb{R}$ is lsc, we have

$$\liminf_{i\to\infty} V_{t+1}^\theta(f_t(x^i, u^i, w^i)) \geq V_{t+1}^\theta(f_t(x, u, w)), \tag{21}$$

which proves that the composition $V_{t+1}^\theta \circ f_t$ is lsc. Since $V_{t+1}^\theta \circ f_t$ is lsc and $\frac{-\theta}{2} > 0$, the function $g_t = \frac{-\theta}{2} V_{t+1}^\theta \circ f_t$ is lsc because multiplying (21) by a positive constant preserves the direction of the inequality.

By Proposition 1, it holds that $\phi_t = \exp \circ g_t$ (17) is lsc because $\exp : \mathbb{R} \to (0, +\infty)$ is continuous and increasing and $g_t : \mathcal{S} \to \mathbb{R}$ (19) is lsc and bounded. In addition, ϕ_t is bounded as a consequence of g_t being bounded.

It follows that ϕ_t' (18) is bounded and lsc. The latter property holds in particular because ϕ_t (17) is lsc and $p_t(\mathrm{d}w|x, u)$ is a continuous stochastic kernel (see Lemma 2, Appendix).

[2]If ψ_t^θ is lsc and bounded, then $v_{t+1}^\theta = c_t + \psi_t^\theta$ is lsc and bounded because the sum of two lsc and bounded functions is lsc and bounded. Since v_{t+1}^θ is bounded, V_t^θ is bounded (7a). The remaining desired conclusions follow from a known result, which we describe next. Since $v_{t+1}^\theta : S \times A \to \mathbb{R}$ is lsc, A is compact, and $V_t^\theta(x) = \inf_{u\in A} v_{t+1}^\theta(x, u)$ $\forall x \in S$, it holds that V_t^θ is lsc, and there is a Borel-measurable function $\mu_t^\theta : S \to A$ such that $V_t^\theta(x) = v_{t+1}^\theta(x, \mu_t^\theta(x))$ $\forall x \in S$ by a special case of [42, Proposition 7.33, p. 153]. In summary, if ψ_t^θ is lsc and bounded, then a) V_t^θ is lsc and bounded and b) a Borel-measurable function μ_t^θ satisfying (8) exists. This logic repeats backwards in time to complete the proof.

We use Proposition 1 again to conclude that $\log \circ \phi_t' : S \times A \to \mathbb{R}$ is lsc. To apply Proposition 1, we choose $\mathcal{M} = S \times A$, $\mathcal{Y}_1 = (0, +\infty)$, $\mathcal{Y}_2 = \mathbb{R}$, $\kappa_1 = \log$, $\kappa_2 = \phi_t'$, and $[\underline{c}, \overline{c}] = [e^{\frac{-\theta}{2}\underline{b}}, e^{\frac{-\theta}{2}\overline{b}}] \subset \mathcal{Y}_1$, where $\underline{b}$ is a lower bound and $\overline{b}$ is an upper bound for V_{t+1}^θ. Finally, since $\log \circ \phi_t'$ is lsc and $\frac{-2}{\theta} > 0$, we conclude that $\psi_t^\theta = \frac{-2}{\theta} \log \circ \phi_t'$ is lsc. ∎

By proving Theorem 1, we guarantee that the functions $V_0^\theta, V_1^\theta, \dots, V_N^\theta$ satisfy properties which facilitate the optimality result (Theorem 2) in the following section.

V. Existence of an Optimal Risk-Averse Policy

In this section, first we prove a DP recursion for the risk-averse control problem; the sum-to-product property of the exponential function is particularly useful for this proof (Lemma 1). Then, we use Lemma 1 and Theorem 1 to show the equality $V_0^\theta = V_\theta^*$ and to construct an optimal risk-averse policy (Theorem 2).

Define the random cost-to-go Z_t for time $t \in \mathbb{T}_N$ as follows: for all $\omega = (x_0, u_0, \dots, x_{N-1}, u_{N-1}, x_N) \in \Omega$,

$$Z_t(\omega) := \begin{cases} c_N(x_N) + \sum_{i=t}^{N-1} c_i(x_i, u_i) & \text{if } t \in \mathbb{T} \\ c_N(x_N) & \text{if } t = N \end{cases}. \tag{22}$$

Note that $Z_t(\omega) = c_t(x_t, u_t) + Z_{t+1}(\omega)$ for any $t \in \mathbb{T}$ and $\omega \in \Omega$ of the form specified above, and $Z_0 = Z$ (5). While Z_t is a random variable whose domain is Ω, Z_t does not depend on the trajectory prior to time t, which is required to derive a recursion that is history-dependent only through the current state. For any $t \in \mathbb{T}_N$, $x \in S$, $\theta \in \Theta$, and $\pi \in \Pi$, we denote a conditional expectation of $e^{\frac{-\theta}{2} Z_t}$ by $W_t^{\pi,\theta}(x) := E^\pi(e^{\frac{-\theta}{2} Z_t}|X_t = x)$.

Lemma 1 (A DP Recursion): Let $\theta \in \Theta$ and $\pi = (\mu_0, \mu_1, \dots, \mu_{N-1}) \in \Pi$ be given. Under Assumption 1, it holds that $W_t^{\pi,\theta}(x) \in (0, +\infty)$ for all $x \in S$, and

$$W_t^{\pi,\theta}(x) = e^{\frac{-\theta}{2} c_t(x,\mu_t(x))} \int_D W_{t+1}^{\pi,\theta}(f_t(x, \mu_t(x), w))\, p_t(\mathrm{d}w|x, \mu_t(x))$$

for all $t \in \mathbb{T}$ and $x \in S$.

Proof: To derive the form of $W_t^{\pi,\theta}$, the first step is to use P_x^π (3) to derive the induced probability measure $P_{X_t}^\pi(B) := P_x^\pi(\{X_t \in B\})$, where $B \in \mathcal{B}_S$. The second step is to apply the definition of conditional expectation [43, Th. 6.3.3, p. 245]. It follows that the function $W_t^{\pi,\theta} : S \to \mathbb{R}^*$ is Borel measurable and $W_t^{\pi,\theta}(x_t)$ is given by

$$\begin{aligned} W_t^{\pi,\theta}(x_t) = \int_A \int_S \int_A \cdots \int_A \int_S e^{\frac{-\theta}{2}(c_N(x_N) + \sum_{i=t}^{N-1} c_i(x_i,u_i))} \\ q_{N-1}(\mathrm{d}x_N|x_{N-1}, u_{N-1})\, \delta_{\mu_{N-1}(x_{N-1})}(\mathrm{d}u_{N-1}) \cdots \\ \delta_{\mu_{t+1}(x_{t+1})}(\mathrm{d}u_{t+1})\, q_t(\mathrm{d}x_{t+1}|x_t, u_t)\, \delta_{\mu_t(x_t)}(\mathrm{d}u_t) \end{aligned} \tag{23}$$

for all $x_t \in S$ and $t \in \mathbb{T}$. Similarly, $W_N^{\pi,\theta} : S \to \mathbb{R}^*$ is Borel measurable and satisfies $W_N^{\pi,\theta}(x_N) = e^{\frac{-\theta}{2} c_N(x_N)}$ for all $x_N \in S$. Since c_t is bounded, we have that $W_t^{\pi,\theta}(x_t) \in (0, +\infty)$ for all $x_t \in S$ and $t \in \mathbb{T}_N$. Details about applying [43, Th. 6.3.3] are provided in a footnote.[3]

Let $t \in \{0, 1, \dots, N-2\}$. By the definition of $Z_t(\omega)$ (22), it holds that $e^{\frac{-\theta}{2} Z_t(\omega)} = e^{\frac{-\theta}{2} c_t(x_t,u_t)} e^{\frac{-\theta}{2} Z_{t+1}(\omega)}$, which equals

[3]To apply [43, Th. 6.3.3], note that $e^{\frac{-\theta}{2} Z_t}$ is a random variable on $(\Omega, \mathcal{B}_\Omega, P_x^\pi)$ for any $x \in S$ and $\pi \in \Pi$. X_t is a random object. The expectation $E_x^\pi(e^{\frac{-\theta}{2} Z_t}) := \int_\Omega e^{\frac{-\theta}{2} Z_t(\omega)} \mathrm{d}P_x^\pi(\omega)$ exists because $e^{\frac{-\theta}{2} Z_t(\omega)} \geq 0$ for all $\omega \in \Omega$.

$e^{\frac{-\theta}{2}(c_N(x_N)+\sum_{i=t}^{N-1} c_i(x_i,u_i))}$ in (23). Since $e^{\frac{-\theta}{2}c_t(x_t,u_t)}$ does not depend on the trajectory after time t, it can be placed "outside" several integrals so that (23) becomes

$$W_t^{\pi,\theta}(x_t) = \int_A e^{\frac{-\theta}{2}c_t(x_t,u_t)} \int_S \int_A \cdots \int_A \int_S e^{\frac{-\theta}{2}Z_{t+1}(\omega)} q_{N-1}(\mathrm{d}x_N|x_{N-1},u_{N-1})\, \delta_{\mu_{N-1}(x_{N-1})}(\mathrm{d}u_{N-1}) \cdots \delta_{\mu_{t+1}(x_{t+1})}(\mathrm{d}u_{t+1})\, q_t(\mathrm{d}x_{t+1}|x_t,u_t)\, \delta_{\mu_t(x_t)}(\mathrm{d}u_t). \quad (24)$$

Since $t+1 \in \mathbb{T}$, by (23), it holds that

$$W_{t+1}^{\pi,\theta}(x_{t+1}) = \int_A \cdots \int_A \int_S e^{\frac{-\theta}{2}(c_N(x_N)+\sum_{i=t+1}^{N-1} c_i(x_i,u_i))} q_{N-1}(\mathrm{d}x_N|x_{N-1},u_{N-1})\delta_{\mu_{N-1}(x_{N-1})}(\mathrm{d}u_{N-1}) \cdots \delta_{\mu_{t+1}(x_{t+1})}(\mathrm{d}u_{t+1}) \quad (25)$$

for all $x_{t+1} \in S$, where $e^{\frac{-\theta}{2}(c_N(x_N)+\sum_{i=t+1}^{N-1} c_i(x_i,u_i))}$ in (25) equals $e^{\frac{-\theta}{2}Z_{t+1}(\omega)}$ in (24) by applying the definition (22). Moreover, the expression for $W_{t+1}^{\pi,\theta}(x_{t+1})$ (25) appears in (24), which permits the following conclusion:

$$W_t^{\pi,\theta}(x_t) = \int_A e^{\frac{-\theta}{2}c_t(x_t,u_t)} \int_S W_{t+1}^{\pi,\theta}(x_{t+1}) q_t(\mathrm{d}x_{t+1}|x_t,u_t)\delta_{\mu_t(x_t)}(\mathrm{d}u_t). \quad (26)$$

By using the definition of the Dirac measure $\delta_{\mu_t(x_t)}$ and the definition of q_t (2), we complete the derivation of the recursion for $t \in \{0,1,\ldots,N-2\}$. The derivation for $t = N-1$ is analogous. ■

The last result proves optimality.

Theorem 2 (Optimality of V_0^θ and π_θ^*): Under Assumption 1, it holds that $V_0^\theta(x) = V_\theta^*(x) = \frac{-2}{\theta}\log E_x^{\pi_\theta^*}(e^{\frac{-\theta}{2}Z})$ for all $x \in S$, where $\pi_\theta^* := (\mu_0^\theta, \mu_1^\theta, \ldots, \mu_{N-1}^\theta)$ is non-unique and is given by Theorem 1.

Proof: First, note that $W_0^{\pi,\theta}(x) = E_x^\pi(e^{\frac{-\theta}{2}Z})$ for all $x \in S$ and $\pi \in \Pi$, and recall that $Z = Z_0$. To show the desired statement, it suffices to show that

$$\frac{-2}{\theta}\log W_t^{\pi,\theta}(x) \geq V_t^\theta(x) = \frac{-2}{\theta}\log W_t^{\pi_\theta^*,\theta}(x) \quad (27)$$

for all $t \in \mathbb{T}_N$, $x \in S$, and $\pi \in \Pi$. (Let $t = 0$, note that $\pi_\theta^* \in \Pi$, and use the definition of the infimum.) Proceed by induction. For the base case, we have $\frac{-2}{\theta}\log W_N^{\pi,\theta}(x) = \frac{-2}{\theta}\log(e^{\frac{-\theta}{2}c_N(x)}) = V_N^\theta(x)$ for all $x \in S$ and $\pi \in \Pi$ because $W_N^{\pi,\theta}(x) = e^{\frac{-\theta}{2}c_N(x)}$ and by the definition of V_N^θ. Now, assume (the induction hypothesis) that for some $t \in \mathbb{T}$, it holds that $\frac{-2}{\theta}\log W_{t+1}^{\pi,\theta}(x) \geq V_{t+1}^\theta(x) = \frac{-2}{\theta}\log W_{t+1}^{\pi_\theta^*,\theta}(x)$ for all $x \in S$ and $\pi \in \Pi$. Since $\frac{-\theta}{2} > 0$, the exponential is increasing, and $e^{\log a} = a$ for all $a \in (0,+\infty)$, the induction hypothesis is equivalent to

$$W_{t+1}^{\pi,\theta}(x) \geq e^{\frac{-\theta}{2}V_{t+1}^\theta(x)} = W_{t+1}^{\pi_\theta^*,\theta}(x) \quad \forall x \in S. \quad (28)$$

Now, let $x \in S$ and $\pi = (\mu_0,\mu_1,\ldots,\mu_{N-1}) \in \Pi$ be given. We use the recursion provided by Lemma 1, the inequality in (28), V_{t+1}^θ being lsc and bounded below (Theorem 1), and $W_{t+1}^{\pi,\theta}$ being Borel measurable to derive the inequality

$$W_t^{\pi,\theta}(x) \geq e^{\frac{-\theta}{2}c_t(x,\mu_t(x))} \int_D e^{\frac{-\theta}{2}V_{t+1}^\theta(f_t(x,\mu_t(x),w))} p_t(\mathrm{d}w|x,\mu_t(x)). \quad (29)$$

The right-hand-side of the inequality in (29) is a product of elements of $(0,+\infty)$ in particular since V_{t+1}^θ is bounded (Theorem 1). Since $\log(ab) = \log a + \log b$ for any $a \in (0,+\infty)$ and $b \in (0,+\infty)$, the natural logarithm is increasing, and $W_t^{\pi,\theta}(x) \in (0,+\infty)$, it holds that

$$\log W_t^{\pi,\theta}(x) \geq \log e^{\frac{-\theta}{2}c_t(x,\mu_t(x))} + \log\left(\int_D e^{\frac{-\theta}{2}V_{t+1}^\theta(f_t(x,\mu_t(x),w))} p_t(\mathrm{d}w|x,\mu_t(x))\right). \quad (30)$$

By simplifying the first term in the sum and multiplying by $\frac{-2}{\theta} > 0$, it follows that $\frac{-2}{\theta}\log W_t^{\pi,\theta}(x) \geq v_{t+1}^\theta(x,\mu_t(x))$, where v_{t+1}^θ is given by (7b). Since $v_{t+1}^\theta(x,\mu_t(x)) \geq \inf_{u\in A} v_{t+1}^\theta(x,u) = V_t^\theta(x)$ (7a), we conclude that $\frac{-2}{\theta}\log W_t^{\pi,\theta}(x) \geq V_t^\theta(x)$.

A similar procedure shows that $V_t^\theta(x) = \frac{-2}{\theta}\log W_t^{\pi_\theta^*,\theta}(x)$ to complete the induction. In particular, one uses $\pi_\theta^* = (\mu_0^\theta, \mu_1^\theta, \ldots, \mu_{N-1}^\theta) \in \Pi$ as the policy in the recursion provided by Lemma 1, where each $\mu_t^\theta : S \to A$ is Borel measurable, satisfies (8), and exists by Theorem 1. ■

VI. Concluding Remarks

Here, we have studied a classical risk-averse control problem for an MDP with Borel state and control spaces on a discrete finite-time horizon, where risk is characterized using the exponential utility functional. While exponential-utility optimal control is well-understood in settings with linear-quadratic assumptions, countable state spaces, or infinite-time horizons, it is not well-understood outside of these settings from a basic analytical perspective. Using first principles from measure theory and real analysis, we have presented a more basic path to the solution in comparison to the existing literature. Topics for future work include investigating an alternative path using the Interchangeability Principle [13] and extensions to partial state information and universally measurable policies, e.g., by building on techniques from [44].

Appendix

We state and prove Lemma 2 below.

Lemma 2 (ϕ_t' is lsc and finite): Recall that $\phi_t : S\times A\times D \to (0,+\infty)$ is lsc and bounded (17), and $p_t(\mathrm{d}w_t|x_t,u_t)$ is a continuous stochastic kernel on D given $S \times A$. It holds that $\phi_t' : S\times A \to (0,+\infty)$ (18) is lsc.

Proof: Showing that ϕ_t' is lsc is a special case of [42, Proposition 7.31], which we call Corollary 1: *Let $\mathcal{X}$ and $\mathcal{Y}$ be separable metrizable spaces, and let $q(\mathrm{d}y|x)$ be a continuous stochastic kernel on $\mathcal{Y}$ given $\mathcal{X}$. Suppose that $g : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ is lsc, and there are scalars $\underline{c}$ and $\overline{c}$ such that $\underline{c} \leq g(x,y) \leq \overline{c}$ for all $(x,y) \in \mathcal{X}\times\mathcal{Y}$. Then, the function $\lambda : \mathcal{X} \to \mathbb{R}$ defined by $\lambda(x) := \int_\mathcal{Y} g(x,y)\, q(\mathrm{d}y|x)$ is lsc.*

To apply Corollary 1, choose $\mathcal{X} = S\times A$, $\mathcal{Y} = D$, $g = \phi_t$, $\lambda = \phi_t'$, $\underline{c} = e^{\frac{-\theta}{2}\underline{b}}$, $\overline{c} = e^{\frac{-\theta}{2}\overline{b}}$, where $\underline{b}$ is a lower bound and $\overline{b}$ is an upper bound for V_{t+1}^θ, and $q(\mathrm{d}y|x) = p_t(\mathrm{d}w|x,u)$.

To prove Corollary 1, one uses the fact that $g : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ being lsc and bounded ($\underline{c} \leq g \leq \overline{c}$), where $\mathcal{X}\times\mathcal{Y}$ is a metrizable space, implies that there is a sequence of continuous functions $g_m : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ such that $\underline{c} \leq g_m \leq g_{m+1} \leq g \leq \overline{c}$ for all $m \in \mathbb{N}$, and $\{g_m\}_{m=1}^\infty$ converges to g pointwise. The proof of this fact uses techniques from [42, Lemmas 7.7 & 7.14] and [43, Th. A6.6]. We outline some key steps below for clarity.

One may choose $g_m(z) := \inf\{g(s) + m\rho(z,s) : s = (s_1,s_2), s_1 \in \mathcal{X}, s_2 \in \mathcal{Y}\}$, where $z = (z_1,z_2)$, $z_1 \in \mathcal{X}$, $z_2 \in \mathcal{Y}$, $m \in \mathbb{N}$, and ρ is an appropriate metric on $\mathcal{X}\times\mathcal{Y}$.

Now, let $z = (z_1, z_2) \in \mathcal{X} \times \mathcal{Y}$ and $\epsilon > 0$ be given. For each $m \in \mathbb{N}$, $g_m(z)$ is finite, and hence, there is a point $z_m = (z_{1m}, z_{2m}) \in \mathcal{X} \times \mathcal{Y}$ such that

$$g(z_m) + m\rho(z, z_m) \leq g_m(z) + \epsilon. \tag{31}$$

Since $\underline{c} \leq g(z_m)$ and $g_m(z) \leq g(z)$ for all $m \in \mathbb{N}$, we have

$$\underline{c} + m\rho(z, z_m) \leq g_m(z) + \epsilon \leq g(z) + \epsilon \quad \forall m \in \mathbb{N}. \tag{32}$$

Since $\underline{c}$ is finite, m is finite and positive, ρ is bounded below by zero, and from (32), it follows that

$$0 \leq \rho(z, z_m) \leq \frac{g(z) + \epsilon - \underline{c}}{m} \quad \forall m \in \mathbb{N}. \tag{33}$$

The inequality (33) and $g(z)$ being finite imply that

$$\liminf_{m\to\infty} \rho(z, z_m) = \limsup_{m\to\infty} \rho(z, z_m) = 0, \tag{34}$$

which shows that the limit of $\{\rho(z, z_m)\}_{m=1}^{\infty}$ exists and equals zero. Then, g being lsc and $\rho(z, z_m) \to 0$ implies that $g(z) \leq \liminf_{m\to\infty} g(z_m)$.

Moreover, the proof of Corollary 1 uses the Extended Monotone Convergence Theorem [43, p. 47], which applies in particular because $g_m \geq \underline{c}$ for all $m \in \mathbb{N}$. ■

Acknowledgment

The authors thank Riccardo Bonalli, Marco Pavone, and Dimitri Bertsekas for discussions. The authors are also grateful for advice provided by the Associate Editor and three reviewers, whose comments improved the presentation of this work.

References

[1] A. Majumdar and M. Pavone, "How should a robot assess risk? Towards an axiomatic theory of risk in robotics," in *Robotics Research*. Cham, Switzerland: Springer, 2020, pp. 75–84.

[2] B. Hammoud, M. Khadiv, and L. Righetti, "Impedance optimization for uncertain contact interactions through risk sensitive optimal control," *IEEE Robot. Autom. Lett.*, vol. 6, no. 3, pp. 4766–4773, Jul. 2021.

[3] Y.-C. Sun and G.-H. Yang, "Remote state estimation for nonlinear systems via a fading channel: A risk-sensitive approach," *IEEE Trans. Cybern.*, early access, Mar. 4, 2021, doi: 10.1109/TCYB.2021.3052563.

[4] J. Barreiro-Gomez, S. E. Choutri, and H. Tembine, "Risk-awareness in multi-level building evacuation with smoke: Burj Khalifa case study," *Automatica*, vol. 129, Jul. 2021, Art. no. 109625.

[5] P. Whittle, "Risk-sensitive linear/quadratic/Gaussian control," *Adv. Appl. Probab.*, vol. 13, no. 4, pp. 764–777, 1981.

[6] R. A. Howard and J. E. Matheson, "Risk-sensitive Markov decision processes," *Manag. Sci.*, vol. 18, no. 7, pp. 356–369, 1972.

[7] D. Jacobson, "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games," *IEEE Trans. Autom. Control*, vol. 18, no. 2, pp. 124–131, Apr. 1973.

[8] P. Whittle, "A risk-sensitive maximum principle: The case of imperfect state observation," *IEEE Trans. Autom. Control*, vol. 36, no. 7, pp. 793–801, Jul. 1991.

[9] K. Glover and J. C. Doyle, "State-space formulae for all stabilizing controllers that satisfy an $\mathcal{H}_\infty$-norm bound and relations to risk sensitivity," *Syst. Control Lett.*, vol. 11, no. 3, pp. 167–172, 1988.

[10] K. Glover, "Minimum entropy and risk-sensitive control: The continuous time case," in *Proc. IEEE Conf. Decis. Control*, 1989, pp. 388–391.

[11] G. B. di Masi and Ł. Stettner, "Risk-sensitive control of discrete-time Markov processes with infinite horizon," *SIAM J. Control Optim.*, vol. 38, no. 1, pp. 61–78, 1999.

[12] J. Löfberg, *Minimax Approaches to Robust Model Predictive Control*, vol. 812. Linköping, Sweden: Linköping Univ. Electron. Press, 2003.

[13] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia, PA, USA: SIAM, 2009.

[14] A. Ruszczyński, "Risk-averse dynamic programming for Markov decision processes," *Math. Program.*, vol. 125, no. 2, pp. 235–261, 2010.

[15] Y. Shen, W. Stannat, and K. Obermayer, "Risk-sensitive Markov control processes," *SIAM J. Control Optim.*, vol. 51, no. 5, pp. 3652–3672, 2013.

[16] N. Bäuerle and A. Glauner, "Markov decision processes with recursive risk measures," *Eur. J. Oper. Res.*, to be published, 2021.

[17] A. Shapiro, "Minimax and risk averse multistage stochastic programming," *Eur. J. Oper. Res.*, vol. 219, no. 3, pp. 719–726, 2012.

[18] V. Borkar and R. Jain, "Risk-constrained Markov decision processes," *IEEE Trans. Autom. Control*, vol. 59, no. 9, pp. 2574–2579, Sep. 2014.

[19] W. B. Haskell and R. Jain, "A convex analytic approach to risk-aware Markov decision processes," *SIAM J. Control Optim.*, vol. 53, no. 3, pp. 1569–1598, 2015.

[20] B. P. G. van Parys, D. Kuhn, P. J. Goulart, and M. Morari, "Distributionally robust control of constrained stochastic systems," *IEEE Trans. Autom. Control*, vol. 61, no. 2, pp. 430–442, Feb. 2016.

[21] S. Samuelson and I. Yang, "Safety-aware optimal control of stochastic systems using Conditional value-at-risk," in *Proc. Amer. Control Conf.*, 2018, pp. 6285–6290.

[22] A. Tsiamis, D. S. Kalogerias, L. F. Chamon, A. Ribeiro, and G. J. Pappas, "Risk-constrained linear-quadratic regulators," in *Proc. IEEE Conf. Decis. Control*, 2020, pp. 3040–3047.

[23] M. P. Chapman and L. Lessard, "Toward a scalable upper bound for a CVaR-LQ problem," *IEEE Control Syst. Lett.*, vol. 6, pp. 920–925, Jun. 2021.

[24] M. P. Chapman, R. Bonalli, K. M. Smith, I. Yang, M. Pavone, and C. J. Tomlin, "Risk-sensitive safety analysis using Conditional Value-at-Risk," 2021. [Online]. Available: arXiv:2101.12086.

[25] N. Bäuerle and U. Rieder, "More risk-sensitive Markov decision processes," *Math. Oper. Res.*, vol. 39, no. 1, pp. 105–120, 2014.

[26] T. Bielecki, D. Hernández-Hernández, and S. R. Pliska, "Risk sensitive control of finite state Markov chains in discrete time, with applications to portfolio management," *Math. Methods Oper. Res.*, vol. 50, no. 2, pp. 167–188, 1999.

[27] D. Hernández-Hernández and S. I. Marcus, "Existence of risk-sensitive optimal stationary policies for controlled Markov processes," *Appl. Math. Optim.*, vol. 40, no. 3, pp. 273–285, 1999.

[28] R. Cavazos-Cadena, "Optimality equations and inequalities in a class of risk-sensitive average cost Markov decision chains," *Math. Methods Oper. Res.*, vol. 71, no. 1, pp. 47–84, 2010.

[29] R. Cavazos-Cadena and D. Hernández-Hernández, "Discounted approximations for risk-sensitive average criteria in Markov decision chains with finite state space," *Math. Oper. Res.*, vol. 36, no. 1, pp. 133–146, 2011.

[30] R. Blancas-Rivera, R. Cavazos-Cadena, and H. Cruz-Suárez, "Discounted approximations in risk-sensitive average Markov cost chains with finite state space," *Math. Methods Oper. Res.*, vol. 91, no. 2, pp. 241–268, 2020.

[31] G. B. di Masi and Ł. Stettner, "Infinite horizon risk sensitive control of discrete time Markov processes with small risk," *Syst. Control Lett.*, vol. 40, no. 1, pp. 15–20, 2000.

[32] G. B. di Masi and Ł. Stettner, "Infinite horizon risk sensitive control of discrete time Markov processes under minorization property," *SIAM J. Control Optim.*, vol. 46, no. 1, pp. 231–252, 2007.

[33] A. Jaśkiewicz, "Average optimality for risk-sensitive control with general state space," *Ann. Appl. Probab.*, vol. 17, no. 2, pp. 654–675, 2007.

[34] A. Jaśkiewicz, "A note on risk-sensitive control of invariant models," *Syst. Control Lett.*, vol. 56, nos. 11–12, pp. 663–668, 2007.

[35] H. Asienkiewicz and A. Jaśkiewicz, "A note on a new class of recursive utilities in Markov decision processes," *Appl. Mathematicae*, vol. 44, no. 2, pp. 149–161, 2017.

[36] V. Anantharam and V. S. Borkar, "A variational formula for risk-sensitive reward," *SIAM J. Control Optim.*, vol. 55, no. 2, pp. 961–988, 2017.

[37] Q. Wei, "Continuous-time Markov decision processes with risk-sensitive finite-horizon cost criterion," *Math. Methods Oper. Res.*, vol. 84, no. 3, pp. 461–487, 2016.

[38] Y. Zhang, "Continuous-time Markov decision processes with exponential utility," *SIAM J. Control Optim.*, vol. 55, no. 4, pp. 2636–2660, 2017.

[39] C. Pal and S. Pradhan, "Risk sensitive control of pure jump processes on a general state space," *Stochastics*, vol. 91, no. 2, pp. 155–174, 2019.

[40] X. Guo and J. Zhang, "Risk-sensitive continuous-time Markov decision processes with unbounded rates and Borel spaces," *Discr. Event Dyn. Syst.*, vol. 29, no. 4, pp. 445–471, 2019.

[41] O. Hernández-Lerma and J. B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, vol. 30. New York, NY, USA: Springer, 1996.

[42] D. P. Bertsekas and S. Shreve, *Stochastic Optimal Control: The Discrete-Time Case*. Belmont, MA, USA: Athena Sci., 1996.

[43] R. Ash, *Probability and Real Analysis*. New York, NY, USA: Wiley, 1972.

[44] H. Yu, "Average-cost optimality results for Borel-space Markov decision processes with universally measurable policies," 2021. [Online]. Available: arXiv:2104.00181.