Robust Optimal Control Using Conditional Risk Mappings in Infinite Horizon

Kerem Uğurlu

Monday 9th April, 2018

Department of Applied Mathematics, University of Washington, Seattle, WA 98195
e-mail: keremu@uw.edu

Abstract

We use one-step conditional risk mappings to formulate a risk averse version of a total cost problem on a controlled Markov process in discrete time infinite horizon. The nonnegative one step costs are assumed to be lower semi-continuous but not necessarily bounded. We derive the conditions for the existence of the optimal strategies and solve the problem explicitly by giving the robust dynamic programming equations under very mild conditions. We further give an ϵ-optimal approximation to the solution and illustrate our algorithm in two examples of optimal investment and LQ regulator problems.

1 Introduction

Controlled Markov decision processes have been an active research area in sequential decision making problems in operations research and in mathematical finance. We refer the reader to [24, 2, 23] for an extensive treatment on theoretical background. Classically, the evaluation operator has been the expectation operator, and the optimal control problem is to be solved via Bellman's dynamic programming [5]. This approach and the corresponding problems continue to be an active research area in various scenarios (see e.g. the recent works [33, 34, 37] and the references therein)

On the other hand, expected values are not appropriate to measure the performance of the agent. Hence, expected criteria with utility functions have been extensively used in the literature (see e.g. [35, 36] and the references therein). Other than the evaluation of the performance via utility functions, to put risk aversion into an axiomatic framework, coherent risk measures has been introduced in the seminal paper [1]. [7] has removed the positive homogeneity assumption of a coherent risk measure and named it as a convex risk measure (see [8] for an extensive treatment on this subject).

However, this kind of operator has brought up another difficulty. Deriving dynamic programming equations with these operators in multistage optimization problems is challenging or impossible in many optimization problems. The reason for it is that the Bellman's optimality principle is not necessarily true using this type of operators. That is to say, the optimization problems are not *time-consistent*. Namely, a multistage stochastic decision problem is time-consistent, if resolving the problem at later stages (i.e., after observing some random outcomes), the original solutions remain optimal for the later stages. We refer the reader to [9, 10, 15, 38, 22] for further elaboration and examples on this type of inconsistency. Hence, optimal control problems on multi-period setting using risk measures on bounded and unbounded costs are not vast, but still, some works in this direction are [11, 12, 14, 13].

To overcome this deficit, dynamic extensions of convex/coherent risk measures so called conditional risk measures are introduced in [25] and studied extensively in [18]. In [16], so called Markov risk measures are introduced and an optimization problem is solved in a controlled Markov decision framework both in finite and discounted infinite horizon, where the cost functions are assumed to be bounded. This idea is extended to transient models in [26, 27] and to unbounded costs with w-weighted bounds in [28, 29, 30] and to so called *process-based* measures in [31] and to partially observable Markov chain frameworks in [32].

In this paper, we derive *robust* dynamic programming equations in discrete time on infinite horizon using one step conditional risk mappings that are dynamic analogues of coherent risk measures. We assume that our one step costs are nonnegative, but may well be unbounded from above. We show the existence of an optimal policy via dynamic programming under very mild assumptions. Since our methodology is based on dynamic programming, our optimal policy is by construction time consistent. We further give a recipe to construct an ϵ-optimal policy for the infinite horizon problem and illustrate our theory in two examples of optimal investment and LQ regulator control problem, respectively. To the best of our knowledge, this is the first work solving the optimal control problem in infinite horizon with the minimal assumptions stated in our model.

The rest of the paper is as follows. In Section 2, we briefly review the theoretical background on coherent risk measures and their dynamic analogues in multistage setting, and further describe the framework for the controlled Markov chain that we will work on. In Section 3, we state our main result on the existence of the optimal policy and the existence of optimality equations. In Section 4, we prove our main theorem and present an ϵ algorithm to our control problem. In Section 5, we illustrate our results with two examples, one on an optimal investment problem, and the other on an LQ regulator control problem.

2 Theoretical Background

In this section, we recall the necessary background on static coherent risk measures, and then we extend this kind of operators to the dynamic setting in controlled Markov chain framework in discrete time.

2.1 Coherent Risk Measures

Consider an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the space $\mathcal{Z} := L^1(\Omega, \mathcal{F}, \mathbb{P})$ of measurable functions $Z : \Omega \to \mathbb{R}$ (random variables) having finite first order moment, i.e. $\mathbb{E}^{\mathbb{P}}[|Z|] < \infty$, where $\mathbb{E}^{\mathbb{P}}[\cdot]$ stands for the expectation with respect to the probability measure $\mathbb{P}$. A mapping $\rho : \mathcal{Z} \to \mathbb{R}$ is said to be a *coherent risk measure*, if it satisfies the following axioms

- (A1)(Convexity) $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$ $\forall \lambda \in (0,1)$, $X, Y \in \mathcal{Z}$.
- (A2)(Monotonicity) If $X \preceq Y$, then $\rho(X) \leq \rho(Y)$, for all $X, Y \in \mathcal{Z}$.
- (A3)(Translation Invariance) $\rho(c + X) = c + \rho(X)$, $\forall c \in \mathbb{R}$, $X \in \mathcal{Z}$.
- (A4)(Homogeneity) $\rho(\beta X) = \beta\rho(X)$, $\forall X \in \mathcal{Z}$. $\beta \geq 0$.

The notation $X \preceq Y$ means that $X(\omega) \leq Y(\omega)$ for $\mathbb{P}$-a.s. Risk measures $\rho : \mathcal{Z} \to \mathbb{R}$, which satisfy (A1)-(A3) only, are called convex risk measures. We remark that under the fourth property (homogeneity), the first property (convexity) is equivalent to sub-additivity. We call the risk measure $\rho : \mathcal{Z} \to \mathbb{R}$ law invariant, if $\rho(X) = \rho(Y)$, whenever X and Y have the same distributions. We pair the space $\mathcal{Z} = L^1(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathcal{Z}^* = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$, and the corresponding scalar product

$$\langle \zeta, Z \rangle = \int_\Omega \zeta(\omega) Z(\omega) dP(\omega), \ \zeta \in \mathcal{Z}^*, Z \in \mathcal{Z}. \tag{2.1}$$

By [6], we know that real-valued law-invariant convex risk measures are continuous, hence lower semi-continuous (l.s.c.), in the norm topology of the space $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Hence, it follows by Fenchel-Moreau theorem that

$$\rho(Z) = \sup_{\zeta \in \mathcal{Z}^*} \{\langle \zeta, Z \rangle - \rho^*(\zeta)\}, \text{ for all } Z \in \mathcal{Z}, \tag{2.2}$$

where $\rho^*(Z) = \sup_{Z\in\mathcal{Z}}\{\langle\zeta, Z\rangle - \rho(Z)\}$ is the corresponding conjugate functional (see [20]). If the risk measure ρ is convex and positively homogeneous, hence coherent, then ρ^* is an indicator function of a convex and closed set $\mathfrak{A} \subset \mathcal{Z}^*$ in the respective paired topology. The dual representation in Equation 2.2 then takes the form

$$\rho(Z) = \sup_{\zeta\in\mathfrak{A}}\langle\zeta, Z\rangle,\ Z \in \mathcal{Z}, \tag{2.3}$$

where the set $\mathfrak{A}$ consists of probability density functions $\zeta : \Omega \to \mathbb{R}$, i.e. with $\zeta \succeq 0$ and $\int \zeta dP = 1$.

A fundamental example of law invariant coherent risk measures is Average- Value-at-Risk measure (also called the Conditional-Value-at-Risk or Expected Shortfall Measure). Average-Value- at-Risk at the level of α for $Z \in \mathcal{Z}$ is defined as

$$\mathsf{AV@R}_\alpha(Z) = \frac{1}{1-\alpha}\int_\alpha^1 \mathsf{V@R}_p(Z)dp, \tag{2.4}$$

where

$$\mathsf{V@R}_p(Z) = \inf\{z \in \mathbb{R} : \mathbb{P}(Z \le z) \ge p\} \tag{2.5}$$

is the corresponding left side quantile. The corresponding dual representation for $\mathsf{AV@R}_\alpha(Z)$ is

$$\mathsf{AV@R}_\alpha(Z) = \sup_{m\in\mathcal{A}}\langle m, Z\rangle, \tag{2.6}$$

with

$$\mathcal{A} = \{m \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}) : \int_\Omega m d\mathbb{P} = 1, 0 \le \|m\|_\infty \le \frac{1}{\alpha}\}. \tag{2.7}$$

Next, we give a representation characterizing any law invariant coherent risk measure, which is first presented in Kusuoka [19] for random variables in $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$, and later further investigated in $\mathcal{Z}^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$ for $1 \le p < \infty$ in [17].

Lemma 2.1. *[19] Any law invariant coherent risk measure $\rho : \mathcal{Z}^p \to \mathbb{R}$ can be represented in the following form*

$$\rho(Z) = \sup_{\nu\in\mathfrak{M}}\int_0^1 \mathsf{AV@R}_\alpha(Z)d\nu(\alpha), \tag{2.8}$$

where $\mathfrak{M}$ is a set of probability measures on the interval [0,1].

2.2 Controlled Markov Chain Framework

Next, we introduce the controlled Markov chain framework that we are going to study our problem on. We take the control model $\mathcal{M} = \{\mathcal{M}_n, n \in \mathbb{N}_0\}$, where for each $n \ge 0$, we have

$$\mathcal{M}_n := (X_n, A_n, \mathbb{K}_n, Q_n, F_n, c_n) \tag{2.9}$$

with the following components:

- X_n and A_n denote the state and action (or control) spaces,which are assumed to be complete seperable metric spaces with their corresponding Borel σ-algebras $\mathcal{B}(X_n)$ and $\mathcal{B}(A_n)$.

- For each $x_n \in X_n$, let $A_n(x_n) \subset A_n$ be the set of all admissible controls in the state x_n. Then
$$\mathbb{K}_n := \{(x_n, a_n) : x_n \in X_n,\ a_n \in A_n\} \tag{2.10}$$
stands for the set of feasible state-action pairs at time n.

- We let
$$x_{i+1} = F_i(x_i, a_i, \xi_i), \tag{2.11}$$
for all $i = 0, 1, ...$ with $x_i \in X_i$ and $a_i \in A_i$ as described above, with independent random variables $(\xi_i)_{i\geq 0}$ on the atomless probability space
$$(\Omega^i, \mathcal{G}^i, \mathbb{P}^i). \tag{2.12}$$
We take that $\xi_i \in S_i$, where S_i are Borel spaces. Moreover, we assume that the system equation
$$F_i : \mathbb{K}_i \times S_i \to X_i \tag{2.13}$$
as in Equation (2.11) is continuous.

- We let
$$\Omega = \otimes_{i=1}^{\infty} X^i \tag{2.14}$$
where X^i is as defined in Equation (2.13). For $n \geq 0$, we let
$$\mathcal{F}_n = \sigma(\sigma(\cup_{i=0}^{n} \mathcal{G}^i) \cup \sigma(X_0, A_0, X_1, A_1 \ldots, A_{n-1}, X_n)) \tag{2.15}$$
$$\mathcal{F} = \sigma(\cup_{i=0}^{\infty} \mathcal{F}_i) \tag{2.16}$$
be the filtration of increasing σ-algebras. Furthermore, we define the corresponding probability measures $(\Omega, \mathcal{F})$ as
$$\mathbb{P} = \prod_{i=1}^{\infty} \mathbb{P}^i, \tag{2.17}$$

where the existence of $\mathbb{P}$ is justified by Kolmogorov extension theorem (see [24]). We assume that for any $n \geq 0$, the random vector $\xi_{[n]} = (\xi_0, \xi_1, \ldots, \xi_n)$ and ξ_{n+1} are independent on $(\Omega, \mathcal{F}, \mathbb{P})$.

- The transition law is denoted by $Q_{n+1}(B_{n+1}|x_n, a_n)$, where $B_{n+1} \in \mathcal{B}(X_{n+1})$ is the Borel σ-algebra on X_n, and $(x_n, a_n) \in X_n \times A_n$ is a stochastic kernel on X_n given $\mathbb{K}_n$ (see [23, 24] for further details). We remark here that at each $n \geq 0$ the stochastic kernel depends only on (x_n, a_n) rather than $\mathcal{F}_n$. That is, for each pair $(x_n, a_n) \in \mathbb{K}_n$, $Q_{n+1}(\cdot|x_n, a_n)$ is a probability measure on X_{n+1}, and for each $B_{n+1} \in \mathcal{B}_{n+1}(X_{n+1})$, $Q_{n+1}(B_{n+1}|\cdot, \cdot)$ is a measurable function on $\mathbb{K}_n$. Let $x_0 \in X_0$ be given with the corresponding policy $\Pi = (\pi_n)_{n\geq 0}$. By the Ionescu Tulcea theorem (see e.g. [24]), we know that there exists a unique probability measure $\mathbb{P}^\pi$ on $(\Omega, \mathcal{F})$ such that given $x_0 \in X_0$, a measurable set $B_{n+1} \subset X_{n+1}$ and $(x_n, a_n) \in \mathbb{K}_n$, for any $n \geq 0$, we have

$$\mathbb{P}^{\Pi}_{n+1}(x_{n+1} \in B_{n+1}) \triangleq Q_{n+1}(B_{n+1}|x_n, a_n). \tag{2.18}$$

- Let $\mathbb{F}_n$ be the family of measurable functions $\pi_n : X_n \to A_n$ for $n \geq 0$. A sequence $(\pi_n)_{n\geq 0}$ of functions $\pi_n \in \mathbb{F}_n$ for $n \geq 0$ is called a control policy (or simply a policy), and the function $\pi_n(\cdot)$ is called the decision rule or control at time $n \geq 0$. We denote by Π the set of all control policies. For notational convenience, for every $n \in \mathbb{N}_0$ and $(\pi_n)_{n\geq 0} \in \Pi$, we write

$$\begin{aligned} c_n(x_n, \pi_n) &:= c_n(x_n, \pi_n(x_n)) \\ &:= c_n(x_n, a_n). \end{aligned}$$

  We denote by $\mathfrak{P}(A_n(x_n))$ as the set of probability measures on $A_n(x_n)$ for each time $n \geq 0$. A randomized Markovian policy $(\pi_n)_{n\geq 0}$ is a sequence of measurable functions such that $\pi_n(x_n) \in \mathfrak{P}(A_n(x_n))$ for all $x_n \in X_n$, i.e. $\pi_n(x_n)$ is a probability measure on $A_n(x_n)$. $(\pi_n)_{n\geq 0}$ is called a deterministic policy, if $\pi_n(x_n) = a_n$ with $a_n \in A_n(x_n)$.

- $c_n(x_n, a_n) : \mathbb{K}_n \to \mathbb{R}_+$ is the real-valued cost-per-stage function at stage $n \in \mathbb{N}_0$ with $(x_n, a_n) \in \mathbb{K}_n$.

Definition 2.1. *A real valued function v on $\mathbb{K}_n$ is said to be inf-compact on $\mathbb{K}_n$, if the set*

$$\{a_n \in A_n(x_n) | v(x_n, a_n) \leq r\} \tag{2.19}$$

is compact for every $x_n \in X_n$ and $r \in \mathbb{R}$. As an example, if the sets $A_n(x_n)$ are compact and $v(x_n, a_n)$ is l.s.c. in $a_n \in A_n(x_n)$ for every $x_n \in X_n$, then $v(\cdot, \cdot)$ is inf-compact on $\mathbb{K}_n$. Conversely, if v is inf-compact on $\mathbb{K}_n$, then v is l.s.c. in $a_n \in A_n(x_n)$ for every $x_n \in X_n$.

We make the following assumption about the transition law $(Q_n)_{n\geq 1}$.

Assumption 2.1. *For any $n \geq 0$, the transition law Q_n is weakly continuous; i.e. for any continuous and bounded function $u(\cdot)$ on X_{n+1}, the map*

$$(x_n, a_n) \to \int_{X_{n+1}} u(y) dQ_n(y|x_n, a_n) \tag{2.20}$$

is continuous on $\mathbb{K}_n$.

Furthermore, we make the following assumptions on the one step cost functions and action sets.

Assumption 2.2. *For every $n \geq 0$,*

- *the real valued non-negative cost function $c_n(\cdot,\cdot)$ is l.s.c. in (x_n, a_n). That is for any $(x_n, a_n) \in X_n \times A_n$, we have*

$$c_n(x_n, a_n) \leq \liminf_{(x_n^k, a_n^k) \to (x_n, a_n)} c_n(x_n^k, a_n^k), \tag{2.21}$$

  as $k \to \infty$.

- *The multifunction (also known as a correspondence or point-to-set function) $x_n \to A_n(x_n)$, from X_n to A_n, is upper semicontinuous (u.s.c.) that is, if $\{x_n^l\} \subset X_n$ and $\{a_n^l\} \subset A_n$ are sequences such that $\{x_n^l\} \to \bar{x}_n$ with $\{a_n^l\} \subset A_n$ for all l, and $a_n^l \to \bar{a}_n$, then $\bar{a}_n$ is in $A_n(\bar{x}_n)$.*

- *For every state $x_n \in X_n$, the admissible action set $A_n(x_n)$ is compact.*

2.3 Conditional Risk Mappings

In order to construct dynamic models of risk, we extend the concept of static coherent risk measures to dynamic setting. For any $n \geq 1$, we denote the space $\mathcal{Z}_n := L^1(\Omega, \mathcal{F}_n, \mathbb{P}_n^\pi)$ of measurable functions with $Z : \Omega \to \mathbb{R}$ (random variables) having finite first order moment, i.e. $\mathbb{E}^{\mathbb{P}_n^\pi}[|Z|] < \infty$ $\mathbb{P}_n^\pi$-a.s., where $\mathbb{E}^{\mathbb{P}_n^\pi}$ stands for the conditional expectation at time n with respect to the conditional probability measure $\mathbb{P}_n^\pi$ as defined in Equation (2.18).

Definition 2.2. *Let $X, Y \in \mathcal{Z}_{n+1}$. We say that a mapping $\rho_n : \mathcal{Z}_{n+1} \to \mathcal{Z}_n$ is a one step conditional risk mapping, if it satisfies following properties*

- *(a1) Let $\gamma \in [0, 1]$. Then,*

$$\rho_n(\gamma X + (1-\gamma)Y) \preceq \gamma\rho_n(X) + (1-\gamma)\rho_n(Y) \tag{2.22}$$

- *(a2) If $X \preceq Y$, then $\rho_n(X) \preceq \rho_n(Y)$*
- *(a3) If $Y \in \mathcal{Z}_n$ and $X \in \mathcal{Z}_{n+1}$, then $\rho_n(X+Y) = \rho_n(X) + Y$.*
- *(a4) For $\lambda \succeq 0$ with $\lambda \in \mathcal{Z}_n$ and $X \in \mathcal{Z}_{n+1}$, we have that $\rho_{n+1}(\lambda X) = \lambda \rho_{n+1}(X)$.*

Here, the relation $Y(\omega) \preceq X(\omega)$ stands for $Y \leq X$ $\mathbb{P}^\pi_n$-a.s. We next state the analogous results for representation theorem for conditional risk mappings as in Equation (2.3) (see also [18]).

Theorem 2.1. *Let $\rho_n : \mathcal{Z}_{n+1} \to \mathcal{Z}_n$ be a law-invariant conditional risk mapping satisfying assumptions as stated in Definition 2.2. Let $Z \in \mathcal{Z}_{n+1}$. Then*

$$\rho_n(Z) = \sup_{\mu \in \mathfrak{A}_{n+1}} \langle \mu, Z \rangle, \tag{2.23}$$

where $\mathfrak{A}_{n+1}$ is a convex closed set of conditional probability measures on $(\Omega, \mathcal{F}_{n+1})$, that are absolutely continuous with respect to $\mathbb{P}^\pi_{n+1}$.

Next, we give the Kusuoka representation for conditional risk mappings analogous to Lemma 2.1.

Lemma 2.2. *Let $\rho_n : \mathcal{Z}_{n+1} \to \mathcal{Z}_n$ be a law invariant one-step conditional risk mapping satisfying Assumptions (a1)-(a4) as in Definition 2.2. Let $Z \in \mathcal{Z}_{n+1}$. Then, conditional Average-Value-at-Risk at the level of $0 < \alpha < 1$ is defined as*

$$\mathsf{AV@R}^n_\alpha(Z) \triangleq \frac{1}{1-\alpha} \int_\alpha^1 \mathsf{V@R}^n_p(Z) dp, \tag{2.24}$$

where

$$\mathsf{V@R}^n_p(Z) \triangleq \operatorname{ess\,inf}\{z \in \mathbb{R} : \mathbb{P}^\pi_{n+1}(Z \leq z) \geq p\}. \tag{2.25}$$

Here, we note that $\mathsf{V@R}^n_p(Z)$ is $\mathcal{F}_n$-measurable by definition of essential infimum (see [8] for a definition of essential infimum and essential supremum). Then, we have

$$\rho_n(Z) \triangleq \operatorname*{ess\,sup}_{\nu \in \mathfrak{M}} \int_0^1 \mathsf{AV@R}^n_\alpha(Z) d\nu(\alpha), \tag{2.26}$$

where $\mathfrak{M}$ is a set of probability measures on the interval [0,1].

Remark 2.1. *By Equations (2.24),(2.25) and (2.26), it is easy to see that the corresponding optimal controls at each time $n \geq 0$ is deterministic, if the one step conditional risk mappings are $\mathsf{AV@R}^n_\alpha : \mathcal{Z}_{n+1} \to \mathcal{Z}_n$ as defined in (2.24). On the other hand, by Kusuoka representation, Equation (2.26), it is clear that for other coherent risk randomized policies might be optimal. In this paper, we restrict our study to deterministic policies.*

Definition 2.3. *A policy $\pi \in \Pi$ is called admissible, if for any $n \geq 0$, we have*

$$c_n(x_n, a_n) + \lim_{N\to\infty} \gamma\rho_n\big(c_{n+1}(x_{n+1}, a_{n+1}) \tag{2.27}$$

$$+ \gamma\rho_{n+1}(c_{n+2}(x_{n+2}, a_{n+2}) \ldots + \gamma\rho_{N-1}(c_N(x_N, a_N)))\big) < \infty,\ \mathbb{P}_n^\pi \text{ a.s.} \tag{2.28}$$

The set of all admissible policies is denoted by Π_{ad}.

3 Main Problem

Under Assumptions 2.1, 2.2, our control problem reads as

$$\inf_{\pi\in\Pi_{\text{ad}}} \Big(c_0(x_0, a_0) + \lim_{N\to\infty} \gamma\rho_0(c_1(x_1, a_1) + \gamma\rho_1(c_2(x_2, a_2) \tag{3.29}$$

$$\ldots + \gamma\rho_{N-1}(c_N(x_N, a_N)))\Big) \tag{3.30}$$

Namely, our objective is to find a policy $(\pi_n^*)_{n\geq 0}$ such that the value function in Equation (3.29) is minimized. For convenience, we introduce the following notations that are to be used in the rest of the paper

$$\varrho_{n-1}(\sum_{t=n}^{\infty} c_t(x_t, \pi_t)) := \lim_{N\to\infty} \gamma\rho_{n-1}(c_n(x_n, a_n) + \gamma\rho_n(c_{n+1}(x_{n+1}, a_{n+1})$$

$$... + \rho_{N-1}(c_N(x_N, a_N)))$$

$$V_n(x, \pi) := c_n(x_n, a_n) + \varrho_n(\sum_{t=n+1}^{\infty} c_t(x_t, a_t))$$

$$V_n^*(x) := \inf_{\pi\in\Pi_{\text{ad}}} c_n(x_n, a_n) + \varrho_n(\sum_{t=n+1}^{\infty} c_t(x_t, a_t))$$

$$V_{n,N}(x, \pi) := c_n(x_n, a_n) + \varrho_n(\sum_{t=n+1}^{N-1} c_t(x_t, a_t))$$

$$V_{N,\infty}(x, \pi) := c_N(x_N, a_N) + \varrho_N(\sum_{t=N+1}^{\infty} c_t(x_t, a_t))$$

$$V_{n,N}^*(x) := \inf_{\pi\in\Pi_{\text{ad}}} c_N(x_N, a_N) + \varrho_n(\sum_{t=n+1}^{N} c_t(x_t, a_t))$$

For the control problem to be nontrivial, we need the following assumption on the existence of the policy.

Assumption 3.1. *There exists a policy* $\pi \in \Pi_{ad}$ *such that*

$$c_0(x_0, a_0) + \varrho_0(x_0) < \infty. \tag{3.31}$$

We are now ready to state our main theorem.

Theorem 3.1. *Let* $0 < \gamma < 1$. *Suppose that Assumptions 2.1, 2.2 and 3.1 are satisfied. Then,*

(a) the optimal cost functions V_n^* *are the pointwise minimal solutions of the optimality equations: that is, for every* $n \in \mathbb{N}_0$ *and* $x_n \in X_n$,

$$V_n^*(x_n) = \inf_{a \in A(x_n)} \Big(c_n(x_n, a_n) + \gamma \rho_n(V_{n+1}^*(x_{n+1})) \Big). \tag{3.32}$$

(b) There exists a policy $\pi^* = (\pi_n^*)_{n \geq 0}$ *such that for each* $n \geq 0$, *the control attains the minimum in (3.32), namely for* $x_n \in X_n$

$$V_n^*(x_n) = c_n(x_n, \pi_n^*) + \gamma \rho_n(V_{n+1}^*(x_{n+1})). \tag{3.33}$$

4 Proof of Main Result

Lemma 4.1. *[3] Fix an arbitrary* $n \in \mathbb{N}_0$. *Let* $\mathbb{K}$ *be defined as*

$$\mathbb{K} := \{(x, a) | x \in X, a \in A(x)\}, \tag{4.34}$$

where X *and* A *are complete seperable metric Borel spaces and let* $v : \mathbb{K} \to \mathbb{R}$ *be a given* $\mathcal{B}(X \times A)$ *measurable function. For* $x \in X$, *define*

$$v^*(x) := \inf_{a \in A(x)} v(x, a). \tag{4.35}$$

If v *is non-negative, l.s.c. and inf-compact on* $\mathbb{K}$ *as defined in Definition 2.1, then for any* $x \in X$, *there exists a measurable mapping* $\pi_n : X \to A$ *such that*

$$v^*(x) = v(x, \pi_n) \tag{4.36}$$

and $v^*(\cdot) : X \to \mathbb{R}$ *is measurable, and l.s.c.*

Lemma 4.2. *For any* $n \geq 1$, *let* $c_n(x_n, a_n)$ *be in* $\mathcal{Z}_n$. *Then* $\rho_{n-1}(c_n(x_n, a_n))$ *is an element of* $\mathcal{Z}_{n-1} = L^1(\Omega, \mathcal{F}_n, \mathbb{P}_n^\pi)$.

Proof. Let $\mu \in \mathfrak{A}_n$ be as in Theorem 2.1. By non-negativity of the one step cost function $c_n(\cdot,\cdot)$ and by Fatou Lemma, we have

$$\langle \mu, c_n(x_n, a_n)\rangle \leq \liminf_{(x_n^k, a_n^k)\to(x_n,a_n)} \langle \mu, c_n(x_n^k, a_n^k)\rangle. \tag{4.37}$$

Hence, $\langle \mu, c_n(x_n, a_n)\rangle$ is l.s.c. for $\mathbb{P}^\pi_{n-1}$-a.s. Then, by Equation (2.23), we have

$$\rho_{n-1}(c_n(x_n, a_n)) = \operatorname*{ess\,sup}_{\mu\in\mathfrak{A}_n} \langle \mu, c_n(x_n, a_n)\rangle. \tag{4.38}$$

Hence, by Equation (4.37) and by Equation (4.38) taking supremum of l.s.c. functions being still l.s.c., we conclude that for fixed ω, $\rho_{n-1}(c_n(x_n(\omega), a_n(\omega)))$ is l.s.c. with respect to (x_n, a_n).

Next, we show that $\rho_{n-1}(c_n(x_n, a_n))$ is $\mathcal{F}_{n-1}$ measurable. By Lemma 2.2, we have

$$\begin{aligned}
\rho_{n-1}(c_n(x_n, a_n)) &= \operatorname*{ess\,sup}_{\nu\in\mathfrak{M}} \int_{[0,1]} \mathsf{AV@R}^{n-1}_\alpha(c_n(x_n, a_n))d\nu, && (4.39)\\
&= \operatorname*{ess\,sup}_{\nu\in\mathfrak{M}} \int_{[0,1]} \frac{1}{1-\alpha}\int_\alpha^1 \mathsf{V@R}^{n-1}_p(c_n(x_n, a_n))dp\, d\nu && (4.40)\\
&= \operatorname*{ess\,sup}_{\nu\in\mathfrak{M}} \int_{[0,1]} \frac{1}{1-\alpha}\int_\alpha^1 \operatorname{ess\,inf}\big(z\in\mathbb{R} : \mathbb{P}^\pi_n(c_n(x_n, a_n)\leq z)\geq p\big)dp\, d\nu, && (4.41)
\end{aligned}$$

where $\mathfrak{M}$ is a set of probability measures on the interval [0,1]. By noting that for any $p\in[\alpha,1]$, $\operatorname{ess\,inf}\big(z\in\mathbb{R} : \mathbb{P}^\pi_n(c_n(x_n, a_n)\leq z)\geq p\big)$ is $\mathcal{F}_{n-1}$-measurable, and then, by integrating from α to 1 and multiplying by $\frac{1}{1-\alpha}$, $\mathcal{F}_{n-1}$ measurability is preserved. Similarly, in Equation 4.39, integrating with respect to a probability measure ν on $[0,1]$ and taking supremum of the integrals preserve $\mathcal{F}_{n-1}$ measurability. Hence, we conclude the proof. ☐

Corollary 4.1. *Let $n\geq 1$, $x_n\in X_n$ and $a_n\in A_n$, where X_n and A_n are as introduced in Equation (2.9). Then,*

$$\min_{a_n\in\pi(x_n)} \rho_{n-1}(c_n(x_n, a_n)) \tag{4.42}$$

is l.s.c. in x_n $\mathbb{P}^\pi_{n-1}$-a.s. Furthermore, $\min_{a_n\in\pi(x_n)} \rho_{n-1}(c_n(x_n, a_n))$ is $\mathcal{F}_{n-1}$ measurable.

Proof. We know by Lemma 4.2, $\rho_{n-1}(c_n(x_n, a_n))$ is l.s.c. $\mathbb{P}^\pi_{n-1}$-a.s. Hence, by Lemma 4.1,

$$\min_{a_n\in\pi(x_n)} \rho_{n-1}(c_n(x_n, a_n)) \tag{4.43}$$

is l.s.c. in x_n for any $x_n \in X_n$ $\mathbb{P}^{\pi}_{n-1}$-a.s. for $n \geq 1$. Furthermore, by Lemma 4.1, we know that there exists an $\pi^* \in \Pi$ such that

$$\min_{a_n \in \pi(x_n)} \rho_{n-1}(c_n(x_n, a_n)) = \rho_{n-1}(c_n(x_n, \pi^*(x_n))) \tag{4.44}$$

$$= \rho_{n-1}(c_n(F_{n-1}(x_{n-1}, a_{n-1}, \xi_{n-1}), \tag{4.45}$$

$$\pi^*(F_{n-1}(x_{n-1}, a_{n-1}, \xi_{n-1})))), \tag{4.46}$$

$$\tag{4.47}$$

where F_{n-1} is as defined in Equation (2.11), but we know that $\rho_{n-1}(c_n(x_n, \pi^*_n)$ is $\mathcal{F}_{n-1}$ measurable. Hence, the result follows by Lemma 4.2. □

For every $n \geq 0$, let $L_n(X_n)$ and $L_n(X_n, A_n)$ be the family of non-negative mappings on (X_n, A_n), respectively. Denote

$$T_n(v_{n+1}) := \min_{a_n \in A(x_n)} \big\{ c_n(x_n, a_n) + \gamma \rho_n(v_{n+1}(F_n(x_n, a_n, \xi_n))) \big\}. \tag{4.48}$$

Lemma 4.3. *Suppose that Assumption 2.1, 2.2 and 3.1 hold, then for every $n \geq 0$, we have*

(a) T_n maps $L_{n+1}(X_{n+1})$ into $L_n(X_n)$.

(b) For every $v_{n+1} \in L_{n+1}(X_{n+1})$, there exists a policy π^*_n such that for any $x_n \in X_n$, $\pi^*_n(x_n) \in A_n(x_n)$ attains the minimum in (4.48), namely

$$T_n(v_{n+1}) := c_n(x_n, \pi^*_n) + \gamma \rho(v_{n+1}(F_n(x_n, \pi^*_n, \xi_n))) \tag{4.49}$$

Proof. By assumption, our one-step cost functions $c_n(x_n, a_n)$ are in $L_n(X_n)$. By Corollary 4.1, $\gamma \rho_n(v_{n+1}(F_n(x_n, \pi^*_n, \xi_n)))$ is in $L_n(X_n)$. Hence their sum is in $L_n(X_n, A_n)$, as well. Hence, the result follows via Corollary 4.1 again. □

By Lemma 4.3, we express the optimality equations (4.48) as

$$V^*_n = T_n V^*_{n+1} \quad \text{for } n \geq 0. \tag{4.50}$$

Next, we continue with the following lemma.

Lemma 4.4. *Under the Assumptions 2.1 and 2.2, for $n \geq 0$, let $v_n \in L_n(X_n)$ and $v_{n+1} \in L_{n+1}(X_{n+1})$.*

(a) If $v_n \geq T_n(v_{n+1})$, then $v_n \geq V^*_n$.

(b) If $v_n \leq T_n(v_{n+1})$ and in addition,

$$\lim_{N\to\infty} v_N(x_{N+1}(\omega)) = 0, \tag{4.51}$$

$\mathbb{P}$-a.s., then $v_n \leq V_n^*$.

Proof. (a) By Lemma 4.3, there exists a policy $\pi = (\pi_n)_{n\geq 0}$ such that for all $n \geq 0$,

$$v_n(x_n) \geq c_n(x_n, \pi_n) + \rho_n(v_{n+1}(F_n(x_n, \pi_n, \xi_n))). \tag{4.52}$$

By iterating the right hand side and by monotonicity of $\varrho_n(\cdot)$, we get

$$v_n(x_n) \geq c_n(x_n, \pi_n) + \varrho_n(\sum_{i=n+1}^{N-1} c_i(x_i, \pi_i) + v_N(x_N)). \tag{4.53}$$

Since $v_N(x_N) \geq 0$, we have

$$v_n(x_n) \geq c_n(x_n, \pi_n) + \varrho_n(\sum_{i=n+1}^{N-1} c_i(x_i, \pi_i)), \text{ a.s.} \tag{4.54}$$

Hence, letting $N \to \infty$, we obtain $v_n(x) \geq V_n(x, \pi)$ and so $v_n(x) \geq V_n^*(x)$.

(b) Suppose that $v_n \leq T_n v_{n+1}$ for $n \geq 0$, so that

$$v_n(x_n) \leq c_n(x_n, \pi_n) + \rho_n(c_{n+1}(x_{n+1}, \pi_{n+1}) + v_{n+1}(x_{n+1})) \tag{4.55}$$

for any $\pi \in \Pi_{\text{ad}}$, $\mathbb{P}_n^\pi$-a,s. Summing from $i = 1$ to $i = N - 1$ gives

$$v_n(x_n) \leq c_n(x_n, a_n) + \varrho_n(\sum_{i=1}^{N-1} c_{n+i}(x_{n+i}, a_{n+i}) \tag{4.56}$$

$$+ \varrho_N(\sum_{i=n+N}^{\infty} c_i(x_i, a_i))) \tag{4.57}$$

Letting $N \to \infty$ and by $\pi \in \Pi_{\text{ad}}$, we get that

$$\lim_{N\to\infty} \varrho_n(v_{n+N}) = 0 \tag{4.58}$$

so that we have

$$v_n(x_n) \leq V_n(x_n, \pi), \tag{4.59}$$

Taking infimum, we have

$$v_n(x_n) \leq V_n^*(x_n) \tag{4.60}$$

Thus, we conclude the proof. □

To further proceed, we need the following technical lemma.

Lemma 4.5. *[24] For every $N > n \geq 0$, let X_n, A_n be complete, seperable metric spaces and $\mathbb{K}_n := \{(x_n, a_n) : x_n \in X_n, a_n \in A_n\}$ with w_n and $w_{n,N}$ be functions on $\mathbb{K}_n$ that are non-negative, l.s.c. and inf-compact on $\mathbb{K}_n$. If $w_{n,N} \uparrow w_n$ as $N \to \infty$, then*

$$\lim_{N\to\infty} \min_{a_n \in A_n} w_{n,N}(x_n, a_n) = \min_{a_n \in A_n} w_n(x_n, a_n), \tag{4.61}$$

for all $x_n \in X$.

The next result gives the validity of the convergence of value iteration.

Theorem 4.1. *Suppose that Assumptions 2.1 and 2.2 are satisfied. Then, for every $n \geq 0$ and $x_n \in X_n$,*

$$V^*_{n,N}(x_n) \uparrow V^*_n(x_n) \ \mathbb{P}\text{-a.s. } N \to \infty \tag{4.62}$$

and $V^*_n(x_n)$ l.s.c. $\mathbb{P}$-a.s.

Proof. We obtain $V^*_{n,N}$ by the usual dynamic programming. Indeed, let $J_{N+1}(x_{N+1}) \equiv 0$ for all $x_{N+1} \in X_{N+1}$ a.s. and going backwards in time for $n = N, N-1, \ldots$, let

$$J_n(x_n) := \inf_{a_n \in A(x_n)} c_n(x_n, a_n) + \rho_n(J_{n+1}(F_n(x_n, a_n, \xi_n))). \tag{4.63}$$

Since $J_{N+1}(\cdot) \equiv 0$ is l.s.c., by backward induction, J_N is l.s.c. $\mathbb{P}$-a.s. and $\mathcal{F}_N$-measurable. Moreover, by Corollary 4.1, for every $t = N-1, ..., n$, there exists π^N_t such that $\pi^N_t(x_t) \in A_t(x_t)$ attains the minimum in Equation (4.63). Hence $\{\pi^N_{N-1}, ..., \pi^N_n\}$ is an optimal policy. We note that $c_n(x_n, a_n)$ as well as $\rho_n(J_{n+1}(F_n(x_n, a_n, \xi_n)))$ is l.s.c., $\mathcal{F}_n$ measurable, inf-compact and non-negative. Hence their sum preserves those properties. Furthermore, J_n is the optimal $(N-n)$ cost by construction. Hence, $J_n(x) = V^*_{n,N}(x)$ and since $J_n(x)$ is l.s.c. so is $V^*_{n,N}(x_n)$ with

$$V^*_{n,N}(x_n) := \inf_{a_n \in A(x_n)} \Big(c_n(x_n, a_n) + \rho_n(V^*_{n+1,N}(x_{n+1})) \Big). \tag{4.64}$$

By the non-negativity assumption on $c_n(\cdot, \cdot)$ for all $n \geq 0$, the sequence $N \to V^*_{n,N}$ is non-decreasing and $V^*_{n,N}(x_n) \leq V^*_n(x_n)$, for every $x_n \in X_n$ and $N > n$. Hence, denoting

$$v_n(x_n) := \sup_{N > n} V^*_{n.N}(x_n) \text{ for all } x_n \in X_n. \tag{4.65}$$

and v_n being supremum of l.s.c. functions is itself l.s.c. $\mathbb{P}$-a.s. and $\mathcal{F}$-measurable. Letting $N \to \infty$ in (4.64) by Lemma 4.5, we have that

$$v_n(x_n) := \inf_{a_n \in A(x_n)} \Big(c_n(x_n, a_n) + \rho_n(V_{n+1}(x_{n+1})) \Big) \tag{4.66}$$

for all $n \in \mathbb{N}_0$ and $x_n \in X_n$. Hence, v_n are solutions of the optimality equations, $v_n = T_n v_{n+1}$, and so by Lemma 4.3, $v_n(x_n) \geq V_n^*(x_n)$. This gives $v_n(x) = V_n^*(x)$. Hence, $V_{n,N}^* \uparrow V_n^*$ and V_n^* is l.s.c. □

Now, we are ready to prove our main theorem.

Proof of Theorem 2.1. (a) By Theorem 4.1, the sequence $(V_n^*)_{n\geq 0}$ is a solution to the optimality equations. By Lemma 4.3, it is the minimal such solution.

(b) By Theorem 4.1, the functions V_n^* are l.s.c. $\mathbb{P}$-a.s. and $\mathcal{F}_n$-measurable. Therefore,

$$c_n(x_n, \pi_n^*) + \rho_n(V_{n+1}^*(x_{n+1})) \tag{4.67}$$

is non-negative, l.s.c. $\mathbb{P}$-a.s., $\mathcal{F}_n$-measurable and inf-compact on $\mathbb{K}_n$ for any $a_n \in A_n$, for every $n \geq 0$. Thus, the existence of optimal policy π_n^* follows from Lemma 4.1. Iterating Equation (4.67) gives

$$\begin{aligned} V_n^*(x_n) &= c_n(x_n, \pi_t^*) + \varrho_n\bigg(\sum_{t=n+1}^{N-1} c_t(x_t, \pi_t^*) + V_N^*(x_N)\bigg) && (4.68)\\ &\geq V_{n,N}(x_n, \pi_n^*). && (4.69) \end{aligned}$$

Letting $N \to \infty$, we conclude that $V_n^*(x) \geq V_n(x, \pi^*)$. But by definition of $V_n^*(x)$, we have $V_n^*(x) \leq V_n(x, \pi^*)$. Hence, $V_n^*(x) = V_n(x, \pi^*)$, and we conclude the proof. □

4.1 An ϵ-Optimal Approximation to Optimal Value

We note that our iterative scheme via validity of convergence of value iterations in Theorem 2.1 is computationally not effective for large horizon N problem, since we have to calculate the dynamic programming equations for each time horizon $n \leq N$. To overcome this difficulty, we propose the following methodology, which requires only one time calculation of dynamic programming equations of the optimal control problem and is able to give an ϵ-optimal approximation to the original problem.

By Assumption 3.1, we have after some N_0

$$\varrho_{N_0}\Big(\sum_{n=N_0+1}^{\infty} c_n(x_n, a_n)\Big) < \epsilon \ \mathbb{P}\text{-a.s.} \tag{4.70}$$

But, then this means for the theoretical optimal policy $(\pi_n^*)_{n\geq 0}$, justified in Theorem 2.1, we have

$$\varrho_{N_0}\Big(\sum_{n=N_0+1}^{\infty} c_n(x_n, \pi_n^*)\Big) \leq \epsilon \ \mathbb{P}\text{-a.s.} \tag{4.71}$$

Then, by monotonicity of ϱ, for the optimal policy π^* we have

$$\varrho_{N_0}(\sum_{n=N_0+1}^{\infty} c_n(x_n, \pi_n^*)) \leq \varrho_{N_0}(\sum_{n=N_0+1}^{\infty} c_n(x_n, \pi_n)) \leq \epsilon \ \mathbb{P}\text{-a.s.} \tag{4.72}$$

Hence, we have proved the following theorem.

Theorem 4.2. *Suppose that Assumptions 2.1 and 2.2 hold. Let $\pi_0 \in \Pi_{ad}$ be the policy in Assumption 3.1 such that*

$$\varrho_{N_0}\big(\sum_{n=N_0+1}^{\infty} c_n(x_n, a_n)\big) < \epsilon \ \mathbb{P}\text{-a.s..} \tag{4.73}$$

Then, we have for the optimal policy

$$\varrho_{N_0}\big(\sum_{n=N_0+1}^{\infty} c_n(x_n, a_n^*)\big) \leq \epsilon \ \mathbb{P}\text{-a.s.} \tag{4.74}$$

Hence $\pi^* = \{\pi_0^*, \pi_1^*, \pi_2^*, ..., \pi_{N_0}^*, \pi_{N+1}^0, \dots\}$ is an ϵ-optimal policy for the original problem.

5 Applications

5.1 An Optimal Investment Problem

In this section, we are going to study a variant of mean-variance utility optimization (see e.g. [21]). The framework is as follows. We consider a financial market on an infinite time horizon $[0, \infty)$. The market consists of a risky asset S_n and a riskless asset R_n, whose dynamics are given by

$$\begin{aligned} S_{n+1} - S_n &= \mu S_n + \sigma S_n \xi_n \\ R_{n+1} - R_n &= r R_n \end{aligned}$$

with $R_0 = 1, S_0 = s_0$, where $(\xi_n)_{n\geq 0}$ are i.i.d standard normal random variables having distribution functions Φ on $\mathbb{R}$ with $\mathcal{Z} = L^1(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \Phi)$ and $\mu, r, \sigma > 0$. We consider a self-financing portfolio composed of S and R. We let $(\widetilde{\pi}_n)_{n\geq 0}$ denote the amount of money invested in risky asset S_n at time n and X_n denote the investor's wealth at time n. Namely,

$$X_n^{\widetilde{\pi}} = \widetilde{\pi}_n S_n + R_n \tag{5.75}$$

$$X_{n+1}^{\widetilde{\pi}} - X_n^{\widetilde{\pi}} = \widetilde{\pi}_n (S_{n+1} - S_n) + (X_n^{\widetilde{\pi}} - \widetilde{\pi}_n) r R_n \tag{5.76}$$

For each $n \geq 0$, we denote $\widetilde{\pi}_n = X_n^{\widetilde{\pi}} \pi_n$ so that π_n stands for the fraction of wealth that is put in risky asset. Hence, the wealth dynamics are governed by

$$X_{n+1}^{\pi} - X_n^{\pi} = [rZ_n^{\pi} + (\mu - r)\pi_n] + \sigma \pi_n \xi_n \tag{5.77}$$

with initial value $x_0 = S_0 + B_0$. We further assume $|\pi_n| \leq C$ for some constant $C > 0$ at each time $n \geq 0$.

The particular coherent risk measure used in this example is the mean-deviation risk measure that is in *static setting* defined on $\mathcal{Z}$ as

$$\varrho(X) := \mathbb{E}^{\mathbb{P}}[X] + \gamma g(X), \tag{5.78}$$

with $\gamma > 0$ with

$$g(X) := \mathbb{E}^{\mathbb{P}}\big(|X - \mathbb{E}^{\mathbb{P}}[X]|\big), \tag{5.79}$$

for $X \in \mathcal{Z}$, where $\mathbb{E}^{\mathbb{P}}$ stands for the expectation taken with respect to the measure $\mathbb{P}$. Hence γ determines our *risk averseness* level. For ϱ to satisfy the properties of a coherent risk measure, it is necessary that γ is in $[0, 1/2]$. In fact, γ being in $[0, 1/2]$ is both necessary and sufficient for ϱ to satisfy monotonicity (see [6]). Hence, for fixed $0 \leq \gamma \leq 1/2$ with $X \in \mathcal{Z}$, we have that

$$\rho(X) = \sup_{m \in \mathfrak{A}} \langle m, X \rangle, \tag{5.80}$$

where $\mathfrak{A}$ is a subset of the probability measures, that are of the form (identifying them with their corresponding densities)

$$\mathfrak{A} = \bigg\{ m \in L^{\infty}(\mathbb{R}, \mathcal{B}(\mathbb{R}), \Phi) : \int_{\mathbb{R}} m(x) d\Phi(x) = 1, \tag{5.81}$$

$$m(x) = 1 + h(x) - \int_{\mathbb{R}} h(x) d\Phi(x), \ \|h\|_{\infty} \leq \gamma \ \Phi\text{-a.s.} \bigg\} \tag{5.82}$$

for some $h \in L^{\infty}(\mathbb{R}, \mathcal{B}(\mathbb{R}), \Phi)$. Then, we *define* for each time $n \geq 0$, the dynamic correspondent of ρ as $\rho_n : \mathcal{Z}_{n+1} \to \mathcal{Z}_n$ with

$$\rho_n(X_{n+1}) = \sup_{m_n \in \mathfrak{A}_{n+1}} \langle m_n, X_{n+1} \rangle, \tag{5.83}$$

as in Equation (2.14), (2.15), (2.17) using $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \Phi)$. Hence, the controlled one step conditional risk mapping has the following representation

$$\sup_{m_n \in \mathfrak{A}_{n+1}} \langle m_n, X_n^{\pi} \rangle, \tag{5.84}$$

and our optimization problem reads as

$$\min_{\pi_n \in \Pi_{\mathrm{ad}}} \sup_{m_n \in \mathfrak{A}_{n+1}} \langle m_n, X_n^\pi \rangle, \tag{5.85}$$

where $\mathfrak{A}_{n+1}$ are the sets of conditional probabilities analogous to Equation (5.81) with Π_{ad} as defined in Definition 2.3. Namely, $\mathfrak{A}_{n+1}$ is a subset of the conditional probability measures at time $n+1$ that are of the form (identifying them with their corresponding densities)

$$\mathfrak{A}_{n+1} = \Big\{ m_{n+1} \in L^\infty(\Omega, \mathcal{F}_{n+1}, \mathbb{P}_{n+1}^\pi) : \int_\Omega m_{n+1} d\mathbb{P}_{n+1}^\pi = 1, \tag{5.86}$$

$$m_{n+1} = 1 + h - \int_\Omega h d\mathbb{P}_{n+1}^\pi,\ \|h\|_\infty \leq \gamma\ \mathbb{P}_n^\pi\text{-a.s.} \Big\} \tag{5.87}$$

for some $h \in L^\infty(\Omega, \mathcal{F}_{n+1}, \mathbb{P}_{n+1}^\pi)$, where $\mathbb{P}_{n+1}^\pi$ stands for the conditional probability measure on Ω at time $n+1$ as constructed in (2.18).

Our one step cost functions are $c_n(x_n, a_n) = x_n$ for $n \geq 0$ for some discount factor $0 < \gamma < 1$ that are l.s.c. (in fact continuous) in (x_n, a_n) for $n \geq 0$. Hence, starting with initial wealth at time 0, denoted by x_0, investor's control problem reads as

$$x_0 + \min_{\pi \in \Pi_{\mathrm{ad}}} \varrho_0 \Big(\sum_{n=1}^\infty X_n^\pi \Big) \tag{5.88}$$

$$\triangleq x_0 + \min_{\pi \in \Pi_{\mathrm{ad}}} \lim_{N \to \infty} \Big(c_0(x_0, a_0) + \gamma \rho_0 (c_1, (x_1, a_1) + \ldots + \gamma \rho_{N-1} (c_N(x_N, a_N)) \ldots) \Big) \tag{5.89}$$

We note that Π_{ad} is not empty so that our example satisfies Assumption 3.1. Indeed, by choosing $a_n \equiv 0$ for $n \geq 0$, i.e. investing all the current wealth into riskless asset R_n for $n \geq 0$, we have that

$$\varrho \Big(\sum_{n=0}^\infty \gamma^n x_0 \Big) = \frac{x_0}{1-\gamma} \tag{5.90}$$

Hence, as in Theorem 4.2, we find N_0 such that

$$x_0 \sum_{n=N_0}^\infty \gamma^n < \epsilon. \tag{5.91}$$

Thus, we write the corresponding *robust* dynamic programming equations as follows. Starting with $V_{N_0+1}^* \equiv 0$ for $n = 1, 2, ..., N_0$, we have by Equation (5.85)

$$V_n^*(X_n^\pi) = \min_{|\pi_n| \leq C} X_n^\pi + \gamma \rho_n (V_{n+1}^*(X_{n+1}^\pi)) \tag{5.92}$$

$$= \min_{|\pi_n| \leq C} X_n^\pi + \gamma \sup_{m_{n+1} \in \mathfrak{A}_{n+1}} \langle m_n, V_{n+1}^*(X_{n+1}^\pi) \rangle \tag{5.93}$$

going backwards iteratively at first stage, the problem to solve is then

$$V_0^*(x_0) = \min_{|a_0|\le C} x_0 + \gamma\rho_0(V_1^*(X_1^\pi)) \tag{5.94}$$

$$= x_0 + \gamma \min_{|a_0|\le C} \sup_{m_1\in\mathfrak{A}_1} \langle m_1, V_1^*(X_1^\pi)\rangle \tag{5.95}$$

Hence, the corresponding policy

$$\widetilde{\pi} = \{\pi_0^*, \pi_1^*, \pi_2^*, \dots, \pi_{N_0}^*, 0, 0, 0, \dots, \} \tag{5.96}$$

is ϵ-optimal with the optimal value $V_0^\pi(x_0)$ for our example optimization problem (5.88).

5.2 The Discounted LQ-Problem

We consider the linear-quadratic regulator problem in infinite horizon. We refer the reader to [24] for its study using expectation performance criteria. Instead of the expected value, we use the $\mathsf{AV@R}$ operator to evaluate total discounted performance.

For $n \ge 0$, we consider the scalar, linear system

$$x_{n+1} = x_n + a_n + \xi_n, \tag{5.97}$$

with $X_0 = x_0$, where the disturbances $(\xi_n)_{n\ge 0}$ are independent, identically distributed random variables on $\mathcal{Z}_n^2 = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}^n)$ with mean zero and $\mathbb{E}^{\mathbb{P}^n}[\xi_n^2] < \infty$. The control problem reads as

$$x_0 + \min_{\pi\in\Pi_{\text{ad}}} \varrho_0\left(\sum_{n=1}^{\infty} x_n^\pi\right) \tag{5.98}$$

$$\triangleq x_0 + \min_{\pi\in\Pi_{\text{ad}}} \lim_{N\to\infty} \Big((x_0^2 + a_0^2) + \gamma\rho_0((x_1^2 + a_1^2) \tag{5.99}$$

$$+ \dots + \gamma^N \rho_{N-1}((x_N^2 + a_N^2)) \dots)\Big), \tag{5.100}$$

where $\rho_n(\cdot) : \mathcal{Z}_{n+1}^2 \to \mathcal{Z}_n^2$ is the dynamic $\mathsf{AV@R}_\alpha : \mathcal{Z}_{n+1}^2 \to \mathcal{Z}_n^2$ operator defined as

$$\rho_n(Z) \triangleq \sup_{m_{n+1}\in\mathfrak{A}_{n+1}} \langle m_{n+1}, Z\rangle, \tag{5.101}$$

with

$$\mathfrak{A}_n = \Big\{m_n \in L^\infty(\Omega, \mathcal{F}_n, \mathbb{P}_n^\pi) : \int_\Omega m_n d\mathbb{P}_n^\pi = 1, \tag{5.102}$$

$$0 \le \|m_n\|_\infty \le \frac{1}{\alpha}, \mathbb{P}_{n-1}^\pi - \text{a.s.}\Big\} \tag{5.103}$$

We note that Π_{ad} is not empty. Indeed, choose $\pi_n \equiv 0$ for $n \geq 0$ so that

$$x_n = x_0 + \sum_{i=0}^{n-1} \xi_i, \tag{5.104}$$

with

$$\begin{aligned} \varrho(\sum_{n=0}^{\infty} x_n^2) &\leq 2x_0^2 + 2\varrho(\sum_{n=0}^{\infty} \xi_n^2) && (5.105)\\ &\leq 2x_0^2 + 2\sum_{n=0}^{\infty} \gamma^n \mathsf{AV@R}_\alpha(\xi_n^2) && (5.106)\\ &\leq 2x_0^2 + 2\sum_{n=0}^{\infty} \gamma^n \frac{1}{\alpha}\mathbb{E}^{\mathbb{P}}[\xi_i^2] && (5.107)\\ &\leq 2x_0^2 + \frac{2\sigma^2}{\alpha(1-\gamma)} && (5.108)\\ &< \infty, && (5.109) \end{aligned}$$

where we used Equation (5.102) in the third inequality. Hence, we find N_0 such that

$$\frac{2\sigma^2}{\alpha} \sum_{n=N_0}^{\infty} \gamma^n < \epsilon. \tag{5.110}$$

Starting with $J_{N_0+1} \equiv 0$, the corresponding ϵ-optimal policy for $n = 0, 1, \ldots, N_0$ is found via

$$J_n(x_n) = \min_{|\pi_n| \leq C} \Big((x_n^2 + a_n^2) + \gamma \mathsf{AV@R}_\alpha^n \big(J_{n+1}(x_{n+1} + a_{n+1}) \big) \Big), \tag{5.111}$$

so that at the final stage, we have

$$\begin{aligned} J_0(x_0) &= \min_{|a_0| \leq C} x_0 + \gamma \mathsf{AV@R}_\alpha(J_1(x_1^\pi)) && (5.112)\\ &= x_0 + \gamma \min_{|a_0| \leq C} \sup_{m_1 \in \mathfrak{A}_1} \langle m_1, J_1(x_1) \rangle, && (5.113) \end{aligned}$$

where $\mathfrak{A}$ is as defined in Equation (5.102). Thus, the corresponding policy

$$\widetilde{\pi} = \{\pi_0^*, \pi_1^*, \pi_2^*, \ldots, \pi_{N_0}^*, 0, 0, 0, \ldots, \} \tag{5.114}$$

is ϵ-optimal with the optimal value $V_0^\pi(x_0)$ for problem (5.98).

References

[1] Artzner, P., Delbaen, F., Eber, J.M., Heath, D. (1999). *Coherent measures of risk*, Math. Finance 9, 203-228.

[2] Hernandez-Lerma, O.(1989), *Adaptive Markov Control Processes*, Springer-Verlag. New York.

[3] Rieder, U. (1978). *Measurable Selection Theorems for Optimisation Problems*, Manuscripta Mathematica, 24, 115-131.

[4] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, New York, 1998.

[5] Bellman, R. (1952). *On the theory of dynamic programming* Proc. Natl. Acad. Sci 38, 716.

[6] Ruszczynski, A. and Shapiro, A. (2006). *Optimization of convex risk functions*, Mathematics of Operations Research, vol. 31, pp. 433-452.

[7] H. Follmer and A. Schied, (2002), *Convex measures of risk and trading constraints*, Finance Stochastics, 6 429-447.

[8] H. Follmer and A. Schied, (2011) *Stochastic finance. An introduction in discrete time*, de Gruyter, Berlin.

[9] M.Stadje and P. Cheridito, *Time-inconsistencies of Value at Risk and Time-Consistent Alternatives*, Finance Research Letters. (2009) 6, 1, 40-46.

[10] A. Shapiro, On a time consistency concept in risk averse multi-stage stochastic programming, *Operations Research Letters*. 37 (2009) 143-147.

[11] Goovaerts, M.J. and Laeven, R. (2008), *Actuarial risk measures for financial derivative procing*. Insurance: Mathematics and Economics, 42, 540-547.

[12] Godin, F.(2016), *Minimizing CVaR in global dynamic hedging with transaction costs* (2016), Quantitative Finance, 6, 461-475.

[13] Bauerle, N., Ott, J.(2011)*Markov Decision Processes with Average-Value-at-Risk Criteria* (2011) Mathematical Methods of Operations Research 74, 361-379.

[14] BALBAS, A., BALBAS, R. AND GARRIDO, J.(2010), *Extending pricing rules with general risk functions*, European Journal of Operational Research, 201, 23-33.

[15] ROORDA, B. AND SCHUMACHER, J.(2016), *Weakly time consistent concave valuations and their dual representations.* Finance and Stochastics, 20, 123-151.

[16] RUSZCZYNSKI, A. (2010). *Risk-averse dynamic programming for Markov decision processes*, Math. Program. Ser. B 125:235-261.

[17] PFLUG, G.CH., RÖMISCH, W. (2007) *Modeling, Measuring and Managing Risk.* World Scientific, Singapore.

[18] RUSZCZYNSKI, A., SHAPIRO, A. (2006) *Conditional risk mappings.* Mathematics of Operations Research, 31, 544-561

[19] KUSUOKA S (2001). *On law-invariant coherent risk measures.* Kusuoka S, Maruyama T, eds. Advances in Mathematical Economics, Vol. 3, Springer, Tokyo, 83-95.

[20] ROCKAFELLAR, R.T., WETS, R.J.-B. (1998). *Variational Analysis.*, Springer, Berlin.

[21] BJORK, T., MURGOCI, A., AND ZHOU, X. (2014). *Mean variance portfolio optimization with state dependent risk aversion*, Mathematical Finance 24: 1-24.

[22] BJORK, T., MARIANA KHAPKO, M., AND MURGOCI, A. (2017). *On time-inconsistent stochastic control in continuous time*, Finance and Stochastics, 21(2),331360

[23] SHREVE, S., BERTSEKAS, P.D.(2007). *Stochastic Optimal Control: The Discrete-Time Case*, Athena, Scientific.

[24] HERNANDEZ-LERMA,O., LASSERRE, J.B. (1996). *Discrete-time Markov Control Processes. Basic Optimality Criteria.*, Springer,New York.

[25] RIEDEL, F. (2004). *Dynamic Coherent Risk Measures.* Stochastic Processes Applications, 112, 185-200.

[26] CAVUS, O, RUSZCZYNSKI, A.(2014) *Computational Methods for Risk-averse Undiscounted Transient Markov Models.* Operations Research 62(2), 401-417.

[27] CAVUS, O, RUSZCZYNSKI, A.(2014) *Risk-averse Control of Undiscounted Transient Markov models.* SIAM Journal on Control Optimization 52(6):3935-3966.

[28] Lin, K., Marcus, S. *Dynamic Programming with Non-convex Risk-sensitive Measures.* American Control Conference (ACC), 2013, IEEE. 6778-6783.

[29] Chu, S., Zhang, Y.(2014) *Markov Decision processes with Iterated Coherent Risk Measures* International Journal of Control 87(11):22862293

[30] Shen, Y., Stannat, W., Obermayer, K. (2013) *Risk-sensitive Markov control processes.* SIAM Journal on Control Optimization. 51(5):36523672.

[31] Fan, J., Ruszczynski, A. (2016) *Process-based Risk Measures and Risk-averse Control of Discrete-time Systems* arXiv:1411.2675

[32] J. Fan, J., Ruszczynski, A. (2018) *Risk Measurement and Risk-averse Control of Partially Observable Discrete-time Markov Systems*, Mathematical Methods of Operations Research, 1-24.

[33] Yuksel, S., Saldi, N. *Convex Analysis in Decentralized Stochastic Control, Strategic Measures and Optimal Solutions*(2017) SIAM Journal on Control and Optimization, 55(1):1-28,

[34] Saldi, N., Yuksel, S., and Linder, T.(2017) *Asymptotic Optimality of Finite Approximations to Markov Decision Processes with Borel Spaces* Mathematics of Operations Research, 945-978.

[35] Fleming, W.H., Sheu, S.J.(1999) *Optimal long term growth rate of expected utility of wealth.* Annals of Applied Probability 9, 871-903.

[36] Fleming, W.H., Sheu, S.J.(2000) *Risk-sensitive control and an optimal investment model.* Mathematical Finance 10, 197-213.

[37] Guo, X., Hernandez-del-Valle, A., and Hernandez-Lerma, O.(2010) *Nonstationary discrete- time deterministic and stochastic control systems with infinite horizon*, International Journal of Control, 83:9, 1751-1757.

[38] Shapiro, A.(2016) *Rectangular Sets of Probability Measures*, Operations Research, 64, 528-541.