DYNAMIC PROGRAMMING EQUATIONS FOR MDPS WITH AVAR CRITERIA FOR UNBOUNDED COSTS

KEREM UĞURLU

Abstract. In this paper we derive dynamic programming equations to minimize the Average-Value-at-Risk (AVaR) of the possibly unbounded $L^p$-costs in finite and infinite horizon which is generated by a Markov Decision Process (MDP). We show that with state aggregation and by using the convex analytic formulation of the optimization problem we can solve the problem as in risk-neutral case. To our knowledge, this is the first work of deriving dynamic programming equations with $L^p$-unbounded costs via AVaR-operator.

1. Introduction

Beginning with Bellman [6] dynamic programming has seen huge development both in theory and practice since then. In classical dynamic programming models, the optimization problem has been solved by expected performance criteria. However, in practice expected values are not appropriate to measure the performance criteria. Due to that, risk aversion has been investigated by seperable utility functions as in [8, 10]. To put risk-averse preferences into an axiomatic framework, with the seminal paper of Artzner et al. [2], the risk assessment gained new aspects for random outcomes. We will use this framework to measure risk aversion. Particular tools relevant for us are the duality theory [16] and dynamic programming equations with risk-averse operators [18, 1].

The rest of the paper is as follows. In Section 2, we give the preliminary theoretical framework. In Section 3, we derive the dynamic programming equations for MDP using AVaR criteria for the finite time horizon. In Section 4 we solve the question in infinite case and conclude the paper.

2. Theoretical Background

2.1. Controlled Markov Processes. We quickly review the main concepts of controlled Markov models and introduce the relevant notation. Our presentation closely follows [4]. We consider a controlled Markov state process $(X_n)$ in discrete time with values in Borel set $E$, with nonnegative cost process $(C_n)$. The
random variables are defined on a common probability space \((\Omega, \mathcal{F}, \mathbb{F}_n, P_0)\), where \(P_0\) is the atomless reference probability space. We have the action space \(A\), and denote by \(D \subset E \times A\), the set of all admissible state action combinations. The set \(D(x) := \{a \in A : (x,a) \in D\}\) gives the admissible state actions in state \(x \in E\). Based on the action \(a \in D(x)\) chosen at time \(t\), then we assume that \(A_n\) is \(\mathcal{F}_n = \sigma(X_0, A_0, C_0, ..., X_n)\)-measurable, i.e. our decision might depend entirely on the history. Thus, the sets of histories are as follows:

\[
H_0 := E, \quad H_{k+1} := H_k \times A \times \mathbb{R} \times E
\]

where \(h_k = (x_0, a_0, c_0, x_1, ..., a_k, c_k, x_k) \in H_k\) is the history up to time \(k\). A history dependent policy \(\pi = (g_k)_{k \in \mathbb{N}_0}\) is a mapping \(g_k : H_k \rightarrow A\) such that \(g_k(h_k) \in D(x_k)\). The set of policies is denoted by \(\Pi\). A policy \(\pi \in \Pi\) induces a probability measure \(P^\pi\) on \((\Omega, \mathcal{F})\). The distribution of \(C_n\) and \(X_{n+1}\) is given by the transition kernel \(Q\) from \(E \times A\) to \(E \times \mathbb{R}\) as follows:

\[
P^\pi(X_{n+1} \in B_x, C_n \in B_c | X_0, g_0(X_0), C_0, ..., X_n, g_n(X_0, A_0, C_0, ..., X_n))
\]

\[
= P^\pi(X_{n+1} \in B_x, C_n \in B_c | X_n, g_n(X_0, A_0, C_0, ..., X_n))
\]

\[
= Q(B_x \times B_c | X_n, g_n(X_0, A_0, C_0, ..., X_n))
\]

for measurable sets \(B_x \subset E\) and \(B_c \subset \mathbb{R}\). Suppose, we are given a policy \(\sigma = \{\pi_t\}_{t=0}^\infty\), then by Ionescu Tulcea theorem \([7]\), there exists a unique probability measure \(P^\sigma\) on \((\Omega, \mathcal{F})\), such that for every measurable set \(B \subset \mathcal{F}_n\) and all \(h_t \in \mathcal{H}_t\), \(n = 0, 1, 2, ...
\]

\[
P^\sigma(x_1 \in B) = P(B)
\]

\[
P^\sigma(x_{t+1} \in B|h_t) = Q(B|x_t, \pi_t(h_t))
\]

We consider the following cost functions: \(C^N := \sum_{k=0}^N C_k\), for some \(N \in \mathbb{N}_0\) a fixed terminal time being a positive integer for the finite planning horizon and \(C^\infty := \sum_{k=0}^\infty C_k\) for the infinite planning horizon. We take that the cost functions \(C_k\) are non-negative and \(C^N\) and \(C^\infty\) belong to space \(L^p(\Omega, \mathcal{F}, P_0)\), where \(p \in [1, \infty]\). We start from the following two well-studied optimization problems for controlled Markov processes. The first one is called finite horizon expected value problem, where we want to find a policy \(\Pi = \{\pi_0, \pi_1, ..., \pi_N\}\) with the minimization of the expected cost:

\[
\min_{\Pi} \mathbb{E} \left[ \sum_{t=0}^T c_t(x_t, a_t) \right]
\]

where \(a_t = \pi_t(x_0, x_1, ..., x_t)\) and \(c_t(x_t, a_t)\) is measurable for each \(t = 0, ..., T\). The second problem is the infinite horizon expected value problem. The objective is to
find a policy $\Pi = \{\pi_t\}_{t=0}^{\infty}$ with the minimization of the expected cost:

$$\min_{\Pi} \mathbb{E} \left[ \sum_{t=0}^{\infty} c_t(x_t, a_t) \right]$$

Under some assumptions these two optimization problems have solutions in form of Markov policies. Moreover, in the infinite horizon case, when the underlying Markov model is stationary, we can find a Markovian stationary optimal policy. In both cases, the optimal policies can be found by solving corresponding dynamic programming equations. Our goal is to study the same questions where we use a risk-averse operator $\rho$ instead of the expectation operator and derive the corresponding dynamic equations and look for Markovian optimal solutions under some conditions.

2.2. Coherent risk measures and their representation on $L^p$. We introduce the corresponding risk averse operators that we will be working on throughout the rest of the paper.

**Definition 2.1.** A function $\rho: L^p \rightarrow \mathbb{R}$ is said to be a coherent risk measure if it satisfies the following axioms [2]

- $\rho(\lambda Z + (1 - \lambda)W) \leq \lambda \rho(Z) + (1 - \lambda)\rho(W)$ for all $\lambda \in (0, 1)$, $Z, W \in L^p$;
- If $Z \leq W$ $\mathbb{P}$-a.s. then $\rho(Z) \leq \rho(W)$ for all $Z, W \in L^p$;
- $\rho(c + W) = c + \rho(W)$ for all $c \in \mathbb{R}$ and $W \in L^p$;
- $\rho(\beta Z) = \beta \rho(Z)$ for all $Z \in L^p$ and $\beta \geq 0$.

The particular risk averse operator that we will be working with is the AVaR $\alpha$

**Definition 2.2.** Let $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ be a real-valued random variable and let $\alpha \in (0, 1)$.

- We define the Value-at-Risk of $X$ at level $\alpha$, $\text{VaR}_\alpha(X)$, by

$$\text{VaR}_\alpha(X) = \inf \{ x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq \alpha \} \quad (2.2.1)$$

- We define the coherent risk measure, the Average-Value-at-Risk of $X$ at level $\alpha$, denoted by $\text{AVaR}_\alpha(X)$ by

$$\text{AVaR}_\alpha(X) = \frac{1}{1 - \alpha} \int_0^1 \text{VaR}_t(X) dt \quad (2.2.2)$$

The basic tool for the representation of coherent risk measures is the Fenchel-Moreau theorem, we restate it here (see [3])

**Theorem 2.3.** Let $(E, \tau)$ be a locally convex topological vector space with topological dual $E^*$. Let $f : E \rightarrow \mathbb{R} \cup \{\infty\}$ and $f \neq 0$ being convex and lower
semi-continuous. Then $f$ is identical to the doubly conjugate $f^{**}$, i.e.

$$f(x) = \sup_{x^* \in E^*} (\langle x^*, x \rangle - f^*(x^*)), \forall x \in E,$$

where $f^*(x^*) = \sup_{x \in E} (\langle x^*, x \rangle - f(x)), \forall x^* \in E^*$, is the conjugate of $f$.

Based on this representation, it is shown in [19] that we have the following representation for AVaR$_\alpha(X)$.

**Theorem 2.4.** Let $X \in L^p(\Omega, \mathcal{F}, P)$, where $p \in [1, \infty]$. Denote $(L^p)^* := L^q$ with the topological dual of $L^p$, where $\frac{1}{p} + \frac{1}{q} = 1$. We define

$$\langle X, \mu \rangle = \int_{\Omega} X d\mu$$

Then

$$(2.4.1) \quad \text{AVaR}_\alpha(X) = \sup_{\mu \in C} \langle X, \mu \rangle$$

where $C$ is a weak*-convex sets of probability densities $\mu$ in $L^q$ with

$\{0 \leq \frac{d\mu}{dP} \leq \frac{1}{1-\alpha}\}$. For $1 \leq p < \infty$, the set $C$ is weak*-compact, as well.

**Remark 2.5.** In the subsequent sections, the weak*-compact convex set $C$ of absolutely continuous probability densities $\frac{d\mu}{dP}$ with respect to the reference probability measure $P$ will depend on the state $x_n$ and the current action $a_n$.

Hence, $a_n \rightarrow C(x_n, Q(B_x \times B_c | X_n, a_n(X_0, A_0, C_0, ..., X_n)))$ will be treated as a set valued function and its continuity properties will be vital for subsequent analysis.

We will also need the continuity of the coherent risk measure AVaR$_\alpha(X)$ with respect to $L^p$ norm. For that we state the following theorem [9].

**Theorem 2.6.** Let $(E, \tau)$, be a Frechet-lattice and $f : E \rightarrow \mathbb{R} \cup \{\infty\}$ be a proper, convex, increasing function. Then $f$ is continuous on $I_f := \text{int}(\text{dom}f)$, where $\text{dom}f = \{x \in E : f(x) < \infty\}$ denotes the domain of $f$.

As a consequence of this theorem, we get the continuity with respect to the $L^p$ as a corollary, see also [11] [9].

**Corollary 2.7.** AVaR$_\alpha(X)$, where $X \in L^p$, $1 \leq p \leq \infty$, is continuous with respect to norm topology.

We will also need the following alternative representation for AVaR$_\alpha(X)$ as shown in [15].

**Lemma 2.8.** Let $X \in L^p(\Omega, \mathcal{F}, P)$ be a real-valued random variable and let $\alpha \in (0,1)$. Then it holds that

$$(2.8.1) \quad \text{AVaR}_\alpha(X) = \min_{s \in \mathbb{R}} \left\{ s + \frac{1}{1-\alpha} \mathbb{E}[(X-s)^+] \right\},$$
where the minimum is attained at \( s = \text{VaR}_\alpha(X) \).

3. Finite Horizon Problem

First, we are interested in solving the following optimization problem in the finite horizon.

\[
\min_{\Pi} \text{AVaR}_\alpha \left( \sum_{n=0}^{N} c(x_n, a_n) \right)
\]

We use the representation given in Lemma 2.8 and rewrite the problem as follows:

\[
\inf_{\pi} \text{AVaR}_\pi \left( C_N \arrowvert X_0 = x \right) = \inf_{\pi \in \Pi} \inf_{s \in \mathbb{R}} \left\{ s + \frac{1}{1 - \alpha} \mathbb{E}_x \left[ (C^T - s)^+ \right] \right\}
\]

Based on this, we will be interested in the inner optimization problem as in [4]. Let \( n = 0, 1, 2, ..., N \). We define

\[
w_{n\pi}(x, s) := \mathbb{E}_x \left[ (C^n - s)^+ \right], \quad x \in E, s \in \mathbb{R}, \pi \in \Pi,
\]

\[
w_n(x, s) := \inf_{\pi \in \Pi} w_{n\pi}(x, s), \quad x \in E, s \in \mathbb{R}
\]

We work with the Markov Decision Model with a 2-dimensional state space \( \tilde{E} = E \times \mathbb{R} \), action space \( A \) and admissible actions in \( D \). The second component of the state \( (x, s) \in \tilde{E} \) gives the relevant information of the history of the process. We denote the disturbance variables \( Z_n = (X_n, C_{n-1}) \) with values in \( E \times \mathbb{R}_+ \) influencing the transition. If the state of the Markov Decision Process is \( (x, s) \) at time \( n \) and action \( a \) is chosen, then the distribution of \( Z_{n+1} \) is given by the transition kernel \( Q(\cdot | x, a) \). The transition function \( F : \tilde{E} \times A \times E \times \mathbb{R}_+ \rightarrow \tilde{E} \) determining the new state, is given by

\[
F((x, s), a, (z_1, z_2)) = (z_1, s - z_2)
\]

where the first component of \((z_1, s-z_2)\) is the new state of the original state process and the necessary information update is given by the second component. We take that there is no running cost and we assume that the terminal cost function is given by \( V_{-1}(x, s) := \text{VaR}_\alpha(X) \). We take the decision rules \( f : \tilde{E} \rightarrow A \) such that \( f(x, s) \in D(x) \) and denote by \( \Pi^M \) the set of Markov policies \( \sigma = (f_0, f_1, \ldots, \) \( ), \) where \( f_n \) are decision rules. By Markov policy, we mean that the decision at time \( n \) depends only on the current state \( x \) and \( s \). We denote for \( v \in \mathcal{M}(\tilde{E}) := \{ v : \tilde{E} \rightarrow \mathbb{R}_+ : \text{vismeasurable} \} \) the operators:

\[
Lv(x, a) := \int v(x', s - c)Q(dx' \times dc|x, a), \quad (x, s) \in \tilde{E}, a \in D(x)
\]
and
\[ T_f v(x, s) := \int v(x', s - c) Q(dx' \times dc|x, f(x, s)), \quad (x, s) \in \tilde{E} \]

The minimal cost operator of the Markov Decision Model is given by
\[(3.0.5) \quad T v(x, s) = \inf_{a \in D(x)} L v(x, s, a). \]

For a policy \( \sigma = (f_0, f_1, f_2, \ldots) \in \Pi^M \). We denote by \( \tilde{\sigma} = (f_1, f_2, \ldots) \) the shifted policy. We define for \( \sigma \in \Pi^M \) and \( n = -1, 0, 1, \ldots, N \):
\[ V_{n+1, \sigma} := T_{f_0} V_{n\sigma}, V_{n+1} := \inf_{\sigma} V_{n+1\sigma} = TV_n \]

A decision rule \( f_n^* \) with the property that \( V_n = T_{f_n^*} V_{n-1} \) is called the minimizer of \( V_n \). We have Markovian policies \( \Pi^M \subset \Pi \) in the following sense: Given the global variable \( s \), for every \( \sigma = (f_0, f_1, \ldots) \in \Pi^M \) we find a policy \( \pi = (g_0, g_1, \ldots) \in \Pi \) such that
\[ g_0(x_0) := f_0(x_0, s) \]
\[ g_0(x_0) := f_0(x_0, s) \]
\[ : = : \]

We remark here that a Markovian policy \( \sigma = (f_0, f_1, \ldots) \in \Pi^M \) also depends on the history of the process but not on the whole information. The necessary information at time \( n \) of the history \( h_n = (x_0, a_0, c_0, x_1, \ldots, a_{n-1}, c_{n-1}, x_n) \) is \( x_n \) and \( s - c_0 - c_1 - \ldots - c_{n-1} \). It will be shown below that the optimal policy \( \pi^* \) of problem \( 3.0.2 \). This dependence of the past and the optimality of the Markovian policy is shown in the following theorem.

**Theorem 3.1.** [4] For a given policy \( \sigma \), the only necessary information at time \( n \) of the history \( h_n = (x_0, a_0, c_0, x_1, \ldots, a_{n-1}, c_{n-1}, x_n) \) are the followings

- the state \( x_n \)
- the value \( s_n = s - c_0 - c_1 - \ldots - c_{n-1} \) for \( n = 0, 1, \ldots, N \).

Moreover, it holds for \( n = 0, 1, \ldots, N \) that

- \( w_{n\sigma} = V_{n\sigma} \) for \( \sigma \in \Pi^M \).
- \( w_n = V_n \)

If there exist minimizers \( f_n^*(x, s) \) of \( V_n(x, s) \) on all stages, then the Markov policy \( \sigma^* = (f_N^*, \ldots, f_0^*) \) is optimal for the problem \( 3.0.2 \).
Proof. For \( n = 0 \), we obtain

\[
V_{0\sigma}(x, s) = T_{f_0} V_{-1}(x, s)
\]

\[
= \int V_{-1}(x', s - c)Q(dx' \times dc|x, f_0(x, s))
\]

\[
= \int (s - c)^- Q(dx' \times dc|x, f_0(x, s))
\]

\[
= \int (c - s)^+ Q(dx' \times dc|x, f_0(x, s))
\]

\[
= E_x^\pi [(C_0 - s)^+] = w_{0\sigma}(x, s)
\]

Next by induction argument

\[
V_{n+1\sigma} = T_{f_0} V_{-1}(x, s)
\]

\[
= \int V_{n\sigma}(x', s - c)Q(dx' \times dc|x, f_0(x, s))
\]

\[
= \int E_x^{\pi_n} [(C^n - (s - c))^+]Q(dx' \times dc|x, f_0(x, s))
\]

\[
= \int E_x^{\pi_n} [(c + C^n - s)^+]Q(dx' \times dc|x, f_0(x, s))
\]

\[
= \int E_x^{\pi_n} [(C^{n+1} - s)Q(dx' \times dc|x, f_0(x, s)) = w_{n+1\sigma}(x, s)
\]

We note that the history of the Markov Decision Process \( \tilde{h}_n = (x_0, s_0, a_0, c_0, x_1, s_1, a_1, ..., x_n, s_n) \) contains history \( h_n = (x_0, a_0, c_0, x_1, a_1, ..., x_n) \). We denote by \( \Pi \) the history dependent policies of the Markov Decision Process. By ([5, Theorem 2.2.3]) that

\[
\inf_{\sigma \in \Pi^H} V_{n\sigma}(x, s) = \inf_{\pi \in \Pi} V_{n\pi}(x, s).
\]

Hence, we obtain

\[
\inf_{\sigma \in \Pi^H} w_{n\sigma} \geq \inf_{\pi \in \Pi} w_{n\pi} \geq \inf_{\tilde{\pi} \in \Pi^H} = \inf_{\sigma \in \Pi^H} V_{n\sigma} = \inf_{\sigma \in \Pi^H} w_{n\sigma}
\]

We conclude the proof. \( \square \)
Now using the Theorem 2.4 and Theorem 3.1 we represent the optimization problem as

\[(3.1.1)\]

\[
\inf_{\pi \in \Pi} \text{AvR}_\alpha (C^N) = \inf_{\pi \in \Pi} \sup_{\mu \in \mathcal{C}} (C^N, \mu) = \inf_{\pi \in \Pi} \sup_{\mu \in \mathcal{C}} (C^N, \mu) = \inf_{\pi \in \Pi} \sup_{\mu \in \mathcal{C}} \sum_{i=0}^{N} (c(x_i, a_i), \mu)
\]

\[(3.1.2)\]

\[
= \inf_{\pi \in \Pi} \sup_{\mu \in \mathcal{C}} \sum_{i=0}^{N} c(x_i', a_i') \mathbb{E} \left[ \frac{d\mu}{d\pi^*} \right] (dx_i' \, da_i' \mid x_i, f_i(x_0, a_0, c_0, \ldots, x_i))
\]

\[(3.1.3)\]

\[
= \inf_{\pi \in \Pi} \sup_{\mu \in \mathcal{C}} \sum_{i=0}^{N} c(x_i', a_i') \mathbb{E} \left[ \frac{d\mu}{d\pi^*} \right] (dx_i' \, da_i' \mid x_i, f(x_i, s_i)),
\]

where \(\sigma(x_i, s_i)\) is the sigma algebra that is generated by the information of the state \(x_i\) and the necessary information from the history \(s_i\) introduced above. Moreover, we note here that the set of probability measures on the whole path, \(\mathcal{C}\), is updated at each time \(n\), with respect to the controlled transition kernel \(Q(x_i, a_i)\). Namely, we denote for each time \(n\), the multifunctions \(g(x_n, a_n) = C(x_n, Q(x_n, a_n))\) where \(\mu \in C(x_n, Q(x_n, a_n))\) satisfies \(0 \leq \frac{d\mu}{da_n} |_{x_n} f(x_n, s_n) \leq \frac{1}{1+n}\).

Continuity properties of \(\sup_{\mu \in \mathcal{C}} (C^N, \mu)\) is crucial for our analysis, so we quickly review basic properties of the set valued functions. We refer the reader to [3] for further details. A multifunction \(C(x, y)\) is called \emph{upper semi-continuous} at a point \((x_0, y_0)\), if for every neighborhood \(B\) of \(C(x_0, y_0)\) we can find neighborhoods \(X_0\) of \(x_0\) and \(B_0\) of \(y_0\) such that for all \(x \in X_0\) and all \(y \in B_0\) we have \(C(x, y) \subset B\). The multifunction \(C(x, y)\) is called \emph{lower semi-continuous} at a point \((x_0, y_0)\), if for every \(\mu \in C(x_0, y_0)\) and for every sequence \(\{x^k, y^k\}\) in the domain of \(C\) converging to \((x_0, y_0)\) and for every sequence \(\{x^k, y^k\}\) in the domain of \(C\) converging to \((x_0, y_0)\) we can find a sequence \(\mu^k \in C(x^k, y^k)\) converging to \(\mu\). The multifunction \(C\) is called \emph{continuous}, if it is both upper and lower semi-continuous at every point. Let \(F : X \to Y\) be a set-valued map, and \(f\) be a mapping such that \(f : \text{Graph}(F) \to \mathbb{R}\), then \(g(x) := \sup_{y \in F(x)} f(x, y)\) with \(g : X \to \mathbb{R} \cup \{\infty\}\) is called the marginal function. Let metric spaces \(X, Y\), a set-valued map \(F : X \to Y\) and a function \(f : \text{Graph}(F) \to \mathbb{R}\) be given. We define the marginal function \(g : X \to \mathbb{R} \cup \{\infty\}\) as \(g(x) := \sup_{y \in F(x)} f(x, y)\). We continue with the following theorem for further analysis.

**Theorem 3.2.** Let metric spaces \(X, Y\), a set-valued map \(F : X \to Y\) and a function \(f : \text{Graph}(F) \to \mathbb{R}\) be given.

- If \(f\) and \(F\) are lower semicontinuous, so is the marginal function \(g(x)\).
- If \(f\) and \(F\) are upper semicontinuous and if the values of \(F\) are compact, so is the marginal function \(g(x)\).

We apply these concepts in the following lemma.
Lemma 3.3. Suppose the transition kernel $Q(x, \cdot)$ is continuous and the multifunction $C(x, \cdot)$ is lower-semicontinuous then for every fixed cost value $C^N$ the function $a_n \to \sup_{\mu \in C_n} \langle C^N, \mu \rangle$ is lower semicontinuous.

Proof. Since the kernel $Q(x, \cdot)$ is continuous and the multifunction $C(x, \cdot)$ is lower semicontinuous, the composition $C(x, Q(x, \cdot))$ is lower semicontinuous as well. Moreover, for fixed $C^N$ the function $\mu \to \langle C^N, \mu \rangle$ is continuous in weak* topology. Moreover, for $p < \infty$ the values of $C_n$ is weak* compact by Fenchel duality theorem. Hence, by Theorem 3.2, the statement of the theorem follows. \[\square\]

3.1. Time Consistency. We need a conceptual requirement of time consistency to further investigate the optimization problem in 3.0.2. We can define time consistency from the point of view of optimal policies (see also [20]). Intuitively, The sequence of optimization problems is said to be dynamically consistent if the optimal strategies obtained when solving the original problem at time $n$ remain optimal for all subsequent problems. More precisely, if a policy $\pi_t$ is optimal on the time interval $[s, T]$, then it is also optimal on the sub-interval $[t, T]$ for every $t$ with $s \leq t \leq T$.

We are now ready to state the main theorem for the finite time horizon optimization problem.

Theorem 3.4. Assume that the following conditions are satisfied:

- For every $x_n \in E$, the controlled transition kernel with respect to the selected control $Q(x_n, \cdot)$, $n = 0, 1, ..., N$ is continuous.
- For all measurable selections $a_n(x_n) \in D(x_n)$ the functions $x_n \to c(x_n, a_n(x_n))$, $n = 0, ..., N$ are elements of $L^p(\Omega, F_n, \mathbb{P})$.
- For every $x_n \in E$, the non-negative cost functions $c(x_n, \cdot) \mathbb{E} \left[ \frac{d\mu}{d\mathbb{P}}|x_n, s_n \right]$, $n = 0, ..., N$ are lower semi-continuous.
- For every $x_n \in E$, the action sets $D(x_n)$, $n = 0, ..., N$ are compact.
- For every $x_n$, the multifunction $C(x_n, \cdot)$ is lower semicontinuous.

Then problem (3.0.2) has an optimal Markovian policy $\Pi^*_N(x_0)$.

Proof. Rewriting the problem (3.0.2) explicitly

\begin{equation}
\min_{\pi_0, \pi_N} \max_{0 \leq \frac{d\mu}{d\mathbb{P}} \leq \frac{1}{N}} \sum_{i=0}^{N} \int c(x_i, a_i) \, d\mu = \min_{\pi_0, \pi_N} \max_{0 \leq \frac{d\mu}{d\mathbb{P}} \leq \frac{1}{N}} \sum_{i=0}^{N-1} \int c(x_i, a_i) \, d\mu + \min_{\pi_N} \max_{0 \leq \frac{d\mu}{d\mathbb{P}} \leq \frac{1}{N}} \int c(x_N, a_N) \mathbb{E} \left[ \frac{d\mu}{d\mathbb{P}}|x_N, s_N \right] \, d\mathbb{P} \tag{3.4.1}
\end{equation}

\begin{equation}
\min_{\pi_N} \max_{0 \leq \frac{d\mu}{d\mathbb{P}} \leq \frac{1}{N}} \int c(x_N, a_N) \mathbb{E} \left[ \frac{d\mu}{d\mathbb{P}}|x_N, s_N \right] \, d\mathbb{P} \tag{3.4.2}
\end{equation}

The interchangeability of optimization is due to time-consistency concept described above. By Theorem 3.1 we can consider the Markovian policies on the augmented
state space as defined above. Hence, the innermost optimization problem
\[(3.4.3) \min_{\pi^M} \max_{0 \leq \mu_{x_N,s_N} \leq 1} \int c(x_N, a_N)E\left[\frac{d\mu}{dP}\big|_{x_N,s_N}\right]dP\]
is searched among the Markovian policies \(\pi^M\) at time \(N\). We note here that \(E[\frac{d\mu}{dP}|F_N]\), depends only on the current state on the augmented space by Theorem 3.1. Hence, we have \(E[\frac{d\mu}{dP}|x_N,s_N]\). Moreover, \(c(x_N,.)E[\frac{d\mu}{dP}|x_N,s_N]\) is lower-semicontinuous by assumption and
\[0 \leq \mu_{x_N,s_N} \leq 1 - \alpha \int c(x_N, a_N)E\left[\frac{d\mu}{dP}\big|_{x_N,s_N}\right]dP\]
is lower-semicontinuous by Lemma 3.3 above. As the action set \(a_N(x)\) is compact, the subproblem (3.4.3) for \(n = N\) has for every \(x_N \in B_{x_N}\) an optimal solution \(\pi_N^*(x_N)\), which is a measurable function of \(x_N\) by [16], Theorem 14.37. By backward iteration for \(t = 0, 1, \ldots, N - 1\), we conclude the proof. \[\Box\]

**Remark 3.5.** We note here that the optimal policy \(\pi^*\) is Markov, where we mean the two-dimensional Markov Decision Process containing the current state \(x_n\) and the quantity \(s_n\), which stands for the threshold beyond which the costs matter. We recall that \(s_{n+1} = s_n - C_N\). The global variable \(s_n\) thus contains the information of the history necessary to take the decision, hence as named in [4], \(s_n\) is the sufficient statistic in risk averse optimization problem using the AVaR-criteria.

### 4. Infinite Horizon Problem

Throughout this section, we assume that the infinite sum of nonnegative cost functions, \(\sum_{n=0}^{\infty} c(s_n, a_n)\), belongs to \(L^p(\Omega, F, \mathbb{P})\). We define the riskiness of a position of infinite horizon with a prespecified constant \(\alpha\)
\[(4.0.1) \quad AVaR_\alpha\left(\sum_{n=0}^{\infty} c(s_n, a_n)\right)\]

**Theorem 4.1.** We have the following:
\[(4.1.1) \quad \lim_{N \to \infty} AVaR_\alpha\left(\sum_{n=0}^{N} c(s_n, a_n)\right) = AVaR_\alpha\left(\sum_{n=0}^{\infty} c(s_n, a_n)\right)\]
Hence \(AVaR_\alpha(\sum_{n=0}^{\infty} c(s_n, a_n))\) is a well-defined law-invariant coherent risk measure.

**Proof.** By assumption \(\sum_{n=0}^{N} c(s_n, a_n)\) converge to \(\sum_{n=0}^{\infty} c(s_n, a_n)\) in \(L^p\)-norm. Then by Corollary [27] we conclude that \(\lim_{N \to \infty} AVaR_\alpha\left(\sum_{n=0}^{N} c(s_n, a_n)\right) = AVaR_\alpha(\sum_{n=0}^{\infty} c(s_n, a_n))\) is well defined. Secondly, by passing to the limit, the coherent risk measure properties are straightforward to check. \(\Box\)
Now we state the optimization problem.

\[
(4.1.2) \quad \min_{\Pi} \text{AVaR}_\alpha \left( \sum_{n=0}^{\infty} c(x_n, a_n) \right)
\]

**Theorem 4.2.** Assume the following.

- For every \( x_n \in E \), the controlled transition kernel with respect to the selected control \( Q(x_n, \cdot) \), \( n = 0, 1, \ldots, N \) is continuous.
- For all measurable selections \( a_n(x_n) \in D(x_n) \) the functions \( x_n \to c(x_n, a_n(x_n)) \), \( n = 0, \ldots, N \) are elements of \( L^p(\Omega, \mathcal{F}_n, \mathbb{P}) \).
- For every \( x_n \in E \), the non-negative cost functions \( c(x_n, \cdot) \) are lower semi-continuous.
- For every \( x_n \in E \), the action sets \( D(x_n) \), \( n = 0, \ldots, N \) are compact.
- For every \( x_n \in E \), the multifunction \( C(x_n, \cdot) \) is lower semicontinuous.

Then, problem (4.1.2) has an optimal Markovian policy \( \Pi^*_\infty(x_0) = \lim_{T \to \infty} \Pi^*_T(x_0) \).

**Proof.** We recall that

\[
\min_{\pi_0, \ldots, \pi_N} \text{AVaR}_\alpha \left( \sum_{n=0}^{\infty} c(x_n, a_n) \right) = \lim_{N \to \infty} \min_{\pi_0, \ldots, \pi_N} \text{AVaR}_\alpha \left( \sum_{n=0}^{N} c(x_n, a_n) \right)
\]

\[
= \lim_{N \to \infty} \min_{\pi_0, \ldots, \pi_N} \max_{0 \leq \mu \leq \frac{1}{N}} \sum_{i=0}^{N} \int c(s_i, a_i) d\mu
\]

\[
= \lim_{N \to \infty} \min_{\pi_0, \ldots, \pi_{N-1}} \max_{0 \leq \mu \leq \frac{1}{N}} \sum_{i=0}^{N-1} \int c(s_i, a_i) d\mu
\]

\[
+ \min_{\pi_N} \max_{0 \leq \mu \leq \frac{1}{N}} \int c(s_N, a_N) \mathbb{E}[\frac{d\mu}{dP_\pi}]|x_N, s_N| dP_\pi
\]

where in the last line we have used the fact that, due to state aggregation, the necessary information is already contained at the current state \( x_N \) and \( s_N \) without the necessity to appealing to whole information \( \mathcal{F}_N \), and in the second line we have used the Fenchel-Rockafellar representation. We know by (Theorem 3.4) that there exists a Markovian optimal policy for each finite time horizon \( N \). We need to justify the iteration of the optimal policy \( \pi_0, \pi_1, \ldots, \pi_N \) as \( N \to \infty \). We know by Lemma 3.1 that \( \sup_{\mu \in C(N, \mu)} \) is lower-semicontinuous. By [22], Lemma 4.2.4, (see also [12]), the iteration of optimal Markovian policies and passing to the limit is legitimate, hence the optimal policy is exactly the iteration of the Markovian optimal policy by letting \( N \to \infty \). Thus, we conclude the proof. \( \square \)
5. Conclusion

In this work, we have given a solution to risk-averse optimization using the AVaR\(\alpha\) operator for the possibly unbounded cost functions that are in \(L^p\), \(1 \leq p \leq \infty\). For this, we have extended state space taking into account the necessary information for further time periods and referred to convex analytic representation of the coherent risk measure. We have shown in this extended space that we keep a sufficient statistic of the form \(s_n\) introduced in [4] and the state information \(x_n\). We have shown, by keeping track of this sufficient statistic, there exist Markovian optimal policies both in finite and infinite time horizon cases in possibly unbounded \(L^p\) -cost functions.

References

[1] Acciaio, B., Penner, I. (2011). Dynamic convex risk measures., In G. Di Nunno and B. ksendal (Eds.), Advanced Mathematical Methods for Finance, Springer, 1-34.
[2] Artzner, P., Delbaen, F., Eber, J.M., Heath, D. (1999). Coherent measures of risk, Math. Finance 9, 203-228.
[3] Aubin, J.-P., Frankowska, H. (1978). Set-Valued Analysis Birkhauser, Boston, 1990.
[4] Bauerle, N., Ott J. (2011). Markov Decision Processes with Average-Value-at-Risk Criteria, Mathematical Methods of Operations Research, 74, 361-379.
[5] Bauerle, N., Rieder, U. (2011). Markov Decision Processes with applications to finance, Springer.
[6] Bellman, R. (1952). On the theory of dynamic programming Proc. Natl. Acad. Sci 38, 716.
[7] Bertsekas, D., Shreve, S.E. (1978). Stochastic Optimal Control. The Discrete Time Case, Math. Program. Ser. B 125:235-261.
[8] Chung, K.J., Sorel, M.J. (1987). Discounted MDPs: distribution functions and exponential utility maximization SIAM J. Control Optimization., 25, 49-62.
[9] Ekeland, I., Temam, R. (1974). R. Convex Analysis and Variational Problems, Dunmod.
[10] Fleming, W., Sheu, S. (1999). Optimal long term growth rate of expected utility of wealth Ann. Appl. Prob., 9, 871-903.
[11] Filipovic, D. and Svindland, G. (2012). The canonical model space for law-invariant convex risk measures is L1, Mathematical Finance 22(3), 585-589.
[12] Hernandez-Lerma, O. (2012). Nonstationary discrete-time deterministic and stochastic control systems with infinite horizon, International Journal of Control, vol. 83, pp 1751-1757.
[13] Hernandez-Lerma, O., Lasserre, J.B. (1996). Discrete-time Markov Control Processes, Basic Optimality Criteria., Springer, New York.
[14] Kupper, M., Schachermayer, W. (2009). Representation results for law invariant time consistent functions, Mathematics and Financial Economics 189-210.
[15] Rockafellar, R.T., Uryasev, S. (2002). Conditional-Value-at-Risk for general loss distributions, Journal of Banking and Finance 26, 1443-1471.
[16] Rockafellar, R.T., Wets, R.J.-B. (1998). Variational Analysis., Springer, Berlin.
[17] Ruschendorf, L., Kaina, M. (2009). On convex risk measures on Lp-spaces, Mathematical Methods in Operations Research, 475-495.
[18] Ruszczynski, A. (1999). Risk-averse dynamic programming for Markov decision processes, Math. Program. Ser. B 125:235-261.
[19] Ruszczyński, A. and Shapiro, A. (2006). *Optimization of convex risk functions*, Mathematics of Operations Research, vol. 31, pp. 433-452.

[20] Shapiro, A. (2012). *Time consistency of dynamic risk measures*, Operations Research Letters, vol. 40, pp. 436-439.

[21] Xin, L., Shapiro, A. (2009). *Bounds for nested law invariant coherent risk measures*, Operations Research Letters, vol. 40, pp. 431-435.

[22] Xin, L., Shapiro, A. (2012). *Representation results for law invariant time consistent functions*, Mathematics and Financial Economics 189-210.

Department of Mathematics, University of Southern California, Los Angeles, CA
E-mail address: kugurlu@usc.edu