Journal of Computational and Applied Mathematics 344 (2018) 275–287

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Robust optimal control using conditional risk mappings in infinite horizon

Kerem Uğurlu

Department of Applied Mathematics, University of Washington, Seattle, WA 98195, United States

ARTICLE INFO

Article history:
Received 23 August 2017
Received in revised form 9 April 2018

Keywords:
Coherent/convex risk measures
Dynamic risk measures
Optimal control

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 [1–3] 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 [4]. This approach and the corresponding problems continue to be an active research area in various scenarios (see e.g. the recent works [5–7] 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. [8,9] 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 have been introduced in the seminal paper [10]. [11] has removed the positive homogeneity assumption of a coherent risk measure and named it as a convex risk measure (see [12] 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 [13–17] 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 [18–21].

To overcome this deficit, dynamic extensions of convex/coherent risk measures so called conditional risk measures are introduced in [22] and studied extensively in [23]. In [24], 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 [25,26] and to unbounded costs with

E-mail address: keremu@uw.edu.

https://doi.org/10.1016/j.cam.2018.05.030
0377-0427/

w-weighted bounds in [27–29] and to so called *process-based* measures in [30] and to partially observable Markov chain frameworks in [31].

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 [32], 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 [33]). 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 Eq. (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

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