Distributionally Robust Modeling of Optimal Control

Alexander Shapiro

School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, Georgia 30332-0205,
e-mail: ashapiro@isye.gatech.edu

Abstract. The aim of this paper is to formulate several questions related to distributionally robust Stochastic Optimal Control modeling. As an example the distributionally robust counterpart of the classical inventory model is discussed in details. Finite and infinite horizon stationary settings are considered.

Key Words: optimal control, infinite horizon, distributional robustness, risk measures, rectangularity, duality, inventory model, basestock policy

1 Introduction

The classical Stochastic Optimal Control (SOC) (discrete time) model (e.g., [3]):

$$\min_{\pi\in\Pi} \ \mathbb{E}^{\pi}\big[\sum_{t=1}^{T}\gamma^{t-1}f(x_t,u_t,\xi_t)\big], \tag{1.1}$$

Here variables $x_t \in \mathcal{X} \subset \mathbb{R}^n$ represent state of the system, $\mathcal{U}$ is a nonempty closed subset of $\mathbb{R}^m$, $u_t \in \mathcal{U}$ are controls, $\xi_t \in \Xi \subset \mathbb{R}^d$ are random vectors, $f : \mathcal{X} \times \mathcal{U} \times \Xi \to \mathbb{R}$ is cost function, $\gamma \in (0,1)$ is the discount factor and $F : \mathcal{X} \times \mathcal{U} \times \Xi \to \mathcal{X}$ is a (measurable) mapping. The optimization (minimization) in (1.1) is performed over the set Π of feasible policies satisfying (w.p.1)

$$u_t \in \mathcal{U} \ \text{ and } \ x_{t+1} = F(x_t,u_t,\xi_t), \ t \geq 1. \tag{1.2}$$

We consider cases when the horizon T is finite or infinite.

We refer to (1.1) as the *risk neutral* SOC and consider its distributionally robust counterpart in a nested form. The risk neutral SOC model was thoroughly investigated. On the other hand, the analysis of its distributionally robust counterpart is more involved and still there are delicate questions which are not clearly understood. By duality arguments the distributionally robust formulations are closely related to the risk averse settings, we will discuss this later. The aim of this paper is to describe and formulate certain related questions which in some cases could be quite nonintuitive. As an example we discuss in details the distributionally robust counterpart of the classical inventory model (cf., [12]).

The paper is organized as follows. In the next section we briefly discuss nested formulations of distributionally robust - risk averse functionals in finite and infinite horizon settings. A general discussion of distributionally robust SOC problems is presented in section 3. Section 4 is devoted to a detailed discussion of the inventory model. Finally in section 5 the related questions are formulated and discussed.

We use the following notation and terminology throughout the paper. By $\mathcal{F}$ we denote the Borel sigma algebra of the (closed) set $\Xi \subset \mathbb{R}^d$, and by $\mathfrak{P}$ the set of probability measures on $(\Xi,\mathcal{F})$. For a probability measure $P \in \mathfrak{P}$, the space $L_p(\Xi,\mathcal{F},P)$, $p \in [1,\infty)$, is formed by measurable functions $Z : \Xi \to \mathbb{R}$ such that $\int_\Xi |Z|^p dP < \infty$, and equipped with the norm $\|Z\|_p = \big(\int_\Xi |Z|^p dP\big)^{1/p}$. Assuming that the set $\Xi \subset \mathbb{R}^d$ is compact, the space $C(\Xi)$ consists of continuous functions $Z : \Xi \to \mathbb{R}$ equipped with the sup-norm $\|Z\|_\infty := \sup_{\xi\in\Xi}|Z(\xi)|$. By $\mathbb{B}(\mathcal{X})$ we denote the space of bounded functions $g : \mathcal{X} \to \mathbb{R}$ equipped with the respective sup-norm. These spaces equipped with the respective norms are Banach spaces. For a process $\xi_1, ...,$ we denote by $\xi_{[t]} = (\xi_1, ..., \xi_t)$ its history up to time t. It is said that the random process $\xi_1, ...,$ is stagewise independents if ξ_{t+1} is independent of $\xi_{[t]}$ for all $t \geq 1$. For $a \in \mathbb{R}$, $[a]_+ := \max\{0, a\}$.

2 Distributionally robust functionals

Consider a nonempty set $\mathcal{M} \subset \mathfrak{P}$ of probability measures and the associated functional

$$\mathcal{R}(Z) := \sup_{Q\in\mathcal{M}} \left\{ \mathbb{E}_Q[Z] = \int_\Xi Z(\xi)dQ(\xi) \right\} \tag{2.1}$$

defined on a space of measurable functions $Z : \Xi \to \mathbb{R}$. In order for the functional $\mathcal{R}(Z)$ to be well defined we consider a linear space $\mathcal{Z}$ of measurable functions $Z : \Xi \to \mathbb{R}$ and assume that for every $Z \in \mathcal{Z}$ the expectation $\mathbb{E}_Q[Z]$ is well defined and finite valued for every $Q \in \mathcal{M}$, and moreover that the supremum in (2.1) is finite. That is, $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$ is finite valued. We refer to $\mathcal{M}$ as the *ambiguity* set and to $\mathcal{R}$ as the *distributionally robust* functional.

We deal with the following examples of the linear space $\mathcal{Z}$. Suppose that every $Q \in \mathcal{M}$ is absolutely continuous with respect to a given probability measure $P \in \mathfrak{P}$, referred to as the reference probability measure. Then with every $Q \in \mathcal{M}$ is associated its density function $\zeta = dQ/dP$. Consider a set $\mathfrak{A}$ of density functions (with respect to the reference measure P), and the corresponding ambiguity set $\mathcal{M} := \{Q : dQ/dP \in \mathfrak{A}\}$. Consider the space $\mathcal{Z} := L_p(\Xi, \mathcal{F}, P)$, $p \in [1, \infty)$. Its dual space is $\mathcal{Z}^* = L_q(\Xi, \mathcal{F}, P)$, with $q \in (1, \infty]$ and $1/p+1/q = 1$. The respective bilinear form is $\langle Z, \zeta \rangle = \int_\Xi Z\zeta dP$, for $Z \in \mathcal{Z}$, $\zeta \in \mathcal{Z}^*$. We assume that $\mathfrak{A}$ is a *bounded* subset of the dual space $\mathcal{Z}^*$, and refer to this framework as the L_p-setting. With some abuse of the notation we also use $\mathcal{Z}^*$ to denote the space of finite signed measures $dQ = \zeta dP$, $\zeta \in L_q(\Xi, \mathcal{F}, P)$.

Definition 2.1 *It said that $\zeta', \zeta \in \mathcal{Z}^*$ are distributionally equivalent (with respect to P), denoted $\zeta' \stackrel{\mathcal{D}}{\sim} \zeta$, if $P(\zeta' \le z) = P(\zeta \le z)$ for any $z \in \mathbb{R}$. It is said that set $\mathfrak{A} \subset \mathcal{Z}^*$ is* law invariant *if the following condition holds: $\zeta \in \mathfrak{A}$ and $\zeta' \stackrel{\mathcal{D}}{\sim} \zeta$ implies that $\zeta' \in \mathfrak{A}$.*

It could be noted that the functional $\mathcal{R}$ is law invariant iff the set $\mathfrak{A}$ is law invariant (cf., [10, Proposition 6.30]).

The above L_p-setting assumes existence of a reference probability measure. However, there are important example where there is no naturally defined reference measure. Therefore we also consider the following framework. Suppose that the set $\Xi \subset \mathbb{R}^d$ is compact and consider the space $\mathcal{Z} := C(\Xi)$. Its dual space $\mathcal{Z}^*$ is formed by the linear space of finite signed measures μ with the respective bilinear form $\langle Z, \mu \rangle = \int_\Xi Z d\mu$, for $Z \in \mathcal{Z}$, $\mu \in \mathcal{Z}^*$ (Riesz representation). The dual norm of $\mu \in \mathcal{Z}^*$ is given by the total variation of μ and is equal to one for any probability measure μ. Therefore for any (nonempty) set $\mathcal{M} \subset \mathcal{Z}^*$ the corresponding distributionally robust functional $\mathcal{R}(Z)$ is well defined and finite valued for every $Z \in \mathcal{Z}$. We refer to this framework as the $C(\Xi)$-setting.

By duality arguments the functional $\mathcal{R}$ can be viewed as a risk averse measure. For a discussion of these frameworks and a relation to risk averse measures we can refer to [10, Chapter 7]. We assume throughout the paper that either the L_p or $C(\Xi)$ settings hold. Note that in both frameworks we can assume that the set $\mathcal{M}$ is convex and closed in the weak* topology of the dual space $\mathcal{Z}^*$. Moreover since $\mathcal{M}$ is bounded, $\mathcal{M}$ is weakly* compact.

We can introduce the respective partial order in the space $\mathcal{Z}$. That is, for $Z, Z' \in \mathcal{Z}$ we write $Z \succeq Z'$ if $Z \ge Z'$ almost surely, with respect to the reference measure P, in the L_p-setting; and $Z(\xi) \ge Z'(\xi)$ for all $\xi \in \Xi$ in the $C(\Xi)$-setting. The distributionally robust functional $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$ is monotone, i.e., if $Z \succeq Z'$, then $\mathcal{R}(Z) \ge \mathcal{R}(Z')$. Moreover, $\mathcal{R}$ is subadditive, positively homogeneous and translation equivarint, and hence satisfies the axioms of coherent risk measures (cf., [2]). In both L_p and $C(\Xi)$ settings, the space $\mathcal{Z}$ is a Banach lattice. It follows that if $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$ is a real valued functional satisfying the axioms of coherent risk measures, then it can be represented in the dual form (2.1) for some set of probability measures in the dual space (cf., [6]). It is said that $\mathcal{R}$ is *strictly* monotone, if $Z \succeq Z'$ and $Z \ne Z'$ implies that $\mathcal{R}(Z) > \mathcal{R}(Z')$. We refer to [10, section 9.3.4] for a discussion of necessary and sufficient conditions for strict monotonicity.

Note that if $Z \in \mathcal{Z}$ is bounded by constant υ, i.e., $|Z| \le \upsilon$ almost surely (with respect to the reference measure P) in the L_p-setting, and $|Z(\xi)| \le \upsilon$ for all $\xi \in \Xi$ in the $C(\Xi)$-setting, then $|\mathcal{R}(Z)| \le \upsilon$. In both settings the space $\mathcal{Z}$ is a Banach space, $\mathcal{Z}^*$ is its dual space of continuous linear functionals, and

$$\langle Z, \mu\rangle \le \|Z\|\,\|\mu\|_*,\ Z \in \mathcal{Z},\ \mu \in \mathcal{Z}^*, \tag{2.2}$$

where $\|\cdot\|$ is the norm of $\mathcal{Z}$ and $\|\cdot\|_*$ is its dual norm of $\mathcal{Z}^*$. It follows that

$$|\mathcal{R}(Z)| \le \kappa\|Z\|,\ Z \in \mathcal{Z}, \tag{2.3}$$

where $\kappa = \sup_{\zeta\in\mathfrak{A}} \|\zeta\|_*$ in the L_p-setting, and $\kappa = 1$ in the $C(\Xi)$-setting. Note that in the L_p-setting the constant κ is finite since it is assumed that the set $\mathfrak{A} \subset \mathcal{Z}^*$ is bounded.

2.1 Parametric family of ambiguity sets.

Suppose now that the ambiguity set depends on parameter vector θ varying in set $\Theta \subset \mathbb{R}^k$. That is, consider a set $\mathcal{M}_\theta \subset \mathfrak{P}$ of probability measures parameterized by $\theta \in \Theta$ and the corresponding functional

$$\mathcal{R}_\theta(Z) := \sup_{Q_\theta \in \mathcal{M}_\theta} \int_\Xi Z(\xi) dQ_\theta(\xi). \tag{2.4}$$

We can apply the change of variables approach to construction of the parametric family $\mathcal{M}_\theta$ of probability distributions. Let Ξ' be a (nonempty) closed subset of a finite dimensional vector space and $\mathfrak{P}'$ be the set of Borel probability measures on Ξ'. Consider a family of measurable mappings $G_\theta : \Xi' \to \Xi$, $\theta \in \Theta$. With every probability measure $Q' \in \mathfrak{P}'$ is associated[1] probability measure $Q_\theta = Q' \circ G_\theta^{-1} \in \mathfrak{P}$. Let $\mathcal{M}' \subset \mathfrak{P}'$ be a specified set of probability measures. Consequently define

$$\mathcal{M}_\theta := \{Q_\theta = Q' \circ G_\theta^{-1} : Q' \in \mathcal{M}'\},\ \theta \in \Theta. \tag{2.5}$$

That is, the parametric family $\mathcal{M}_\theta$ of ambiguity sets is determined by the mappings G_θ and the set $\mathcal{M}' \subset \mathfrak{P}'$.

By change of variables $\xi = G_\theta(\eta)$ we can write for $Q_\theta = Q' \circ G_\theta^{-1}$, $Q' \in \mathfrak{P}'$,

$$\int_\Xi Z(\xi) dQ_\theta(\xi) = \int_{\Xi'} Z(G_\theta(\eta)) dQ'(\eta). \tag{2.6}$$

Consequently

$$\mathcal{R}_\theta(Z) = \sup_{Q' \in \mathcal{M}'} \int_{\Xi'} Z'_\theta(\eta) dQ'(\eta) = \sup_{Q' \in \mathcal{M}'} \mathbb{E}_{Q'}[Z'_\theta], \tag{2.7}$$

where $Z'_\theta : \Xi' \to \mathbb{R}$ is defined as $Z'_\theta(\cdot) := Z(G_\theta(\cdot))$. That is, the parameter vector is moved from the probability distributions to the decision variables. This construction will be useful in situations where the considered probability distributions depend on our decisions, we will discuss this later.

[1]For a measurable mapping $G : \Xi' \to \Xi$, the measure $Q = Q' \circ G^{-1}$ is defined as $Q(A) = Q'(G^{-1}(A))$ for measurable set $A \subset \Xi$ with $G^{-1}(A) = \{\eta \in \Xi' : G(\eta) \in A\}$.

Consider now the L_p-setting with reference measures $P_\theta \in \mathfrak{P}$ parameterized by $\theta \in \Theta$. Suppose that $P_\theta = P' \circ G_\theta^{-1}$, where as before $G_\theta : \Xi' \to \Xi$ are measurable mappings and $P' \in \mathfrak{P}'$ is a given probability measure. Let $\mathcal{Z}' := L_p(\Xi, \mathcal{F}, P')$, $p \in [1, \infty)$, and $\mathfrak{A}'$ be a set of density (with respect to P') functions in the dual space of the space $\mathcal{Z}'$. Consider the respective $\mathcal{M}' := \{Q' : dQ'/dP' \in \mathfrak{A}'\} \subset \mathfrak{P}'$ of probability measures absolutely continuous with respect to P' and the corresponding family $\mathcal{M}_\theta \subset \mathfrak{P}$ defined in (2.5). Suppose that for every $\theta \in \Theta$ the mapping $G_\theta : \Xi' \to \Xi$ is *one-to-one*. For $\zeta' = dQ'/dP' \in \mathfrak{A}'$ and measurable set $A \subset \Xi$ we can write by change of variables $\eta = G_\theta^{-1}(\xi)$,

$$Q_\theta(A) = Q'(G_\theta^{-1}(A)) = \int_{G_\theta^{-1}(A)} \zeta'(\eta) dP'(\eta) = \int_A \zeta_\theta(\xi) dP_\theta(\xi), \tag{2.8}$$

where $\zeta_\theta(\xi) = \zeta'(G_\theta^{-1}(\xi))$. We obtain the following.

Proposition 2.1 *For $\theta \in \Theta$ suppose that the (measurable) mapping $G_\theta : \Xi' \to \Xi$ is* one-to-one*. Then the set $\mathcal{M}_\theta$ consists of probability measures Q_θ absolutely continuous with respect to $P_\theta = P' \circ G_\theta^{-1}$ such that $dQ_\theta/dP_\theta \in \mathfrak{A}_\theta$, where*

$$\mathfrak{A}_\theta := \{\zeta_\theta \in \mathcal{Z}^* : \zeta_\theta(\cdot) = \zeta'(G_\theta^{-1}(\cdot)),\ \zeta' \in \mathfrak{A}'\}. \tag{2.9}$$

Moreover, if $\mathfrak{A}'$ is law invariant with respect to P', then $\mathfrak{A}_\theta$ is law invariant with respect to P_θ. Also if the probability measure P' is atomless, then P_θ is atomless.

An important example of law invariant coherent risk measure is the Average Value-at-Risk (also called Conditional Value-at-Risk, Expected Shortfall, Expected Tail Loss),

$$\mathsf{AV@R}_{\alpha,P}(Z) := \inf_{\tau \in \mathbb{R}} \left\{\tau + \alpha^{-1}\mathbb{E}_P[Z - \tau]_+\right\},\ \alpha \in (0, 1], \tag{2.10}$$

defined on the space $\mathcal{Z} = L_1(\Xi, \mathcal{F}, P)$. It has the dual representation with the respective dual set $\mathfrak{A}$ consisting of density functions ζ such that $0 \le \zeta \le \alpha^{-1}$ almost surely with respect to P.

We can apply this to the construction of ambiguity sets. That is, consider the set $\mathfrak{A}'$ of density functions $\zeta' : \Xi' \to \mathbb{R}$ such that $0 \le \zeta' \le \alpha^{-1}$ almost surely with respect to P', corresponding to $\mathsf{AV@R}_{\alpha,P'}$. Then for $\theta \in \Theta$, the respective set $\mathfrak{A}_\theta$ consists of density functions ζ_θ such that $0 \le \zeta_\theta \le \alpha^{-1}$ almost surely with respect to P_θ. That is, $\mathfrak{A}_\theta$ is the dual set of $\mathsf{AV@R}_{\alpha,P_\theta}$. This can be extended to the general class of law invariant coherent risk measures by using their Kusuoka representation.

2.2 Nested distributionally robust functionals.

We will need to deal with nested extensions of the distributionally robust functionals. We restrict our discussion to the following *rectangular* framework. For a positive integer T consider the set $\Xi_T := \Xi \times \cdots \times \Xi \subset \mathbb{R}^{Td}$, equipped with its Borel sigma algebra $\mathcal{F}_T$, and linear space $\mathcal{Z}_T$ of measurable functions $Z_T : \Xi_T \to \mathbb{R}$. In the L_p-setting we consider the probability measure $P_T := P \times \cdots \times P$ given by the product of the reference probability measures, and $\mathcal{Z}_T := L_p(\Xi_T, \mathcal{F}_T, P_T)$. In the $C(\Xi)$ setting we consider $\mathcal{Z}_T := C(\Xi_T)$. In both setting with the Banach space $\mathcal{Z}_T$ is associate its dual space $\mathcal{Z}_T^*$. Consider the set

$$\mathfrak{M}_T := \{Q_1 \times \cdots \times Q_T : Q_t \in \mathcal{M},\ t = 1, ..., T\} \tag{2.11}$$

of product probability measures defined on $(\Xi_T, \mathcal{F}_T)$.

The finite horizon *nested* counterpart of $\mathcal{R}$ is defined in the following way. For $Z_t \in \mathcal{Z}_t$ define: $\rho_1(Z_1) := \mathcal{R}(Z_1)$ for $t = 1$, and for $t \geq 2$ define mapping $\rho_t : \mathcal{Z}_t \to \mathcal{Z}_{t-1}$ as $\rho_t(Z_t)(\xi_{[t-1]}) = \rho_{t|\xi_{[t-1]}}(Z_t)$ with

$$\rho_{t|\xi_{[t-1]}}(Z_t) := \sup_{Q \in \mathcal{M}} \int_{\Xi} Z_t(\xi_{[t-1]}, \xi_t) Q(d\xi_t). \tag{2.12}$$

The nested T-horizon functional is defined as the composite functional $\mathfrak{R}_T := \rho_1 \circ \cdots \circ \rho_T$, i.e.,

$$\mathfrak{R}_T(Z_T) := \rho_1\big(\rho_2(\cdots \rho_T(Z_T)\big). \tag{2.13}$$

Note that by (2.3), for any $t \geq 2$ and $Z_t \in \mathcal{Z}_t$ we have that

$$\left|\rho_{t|\xi_{[t-1]}}(Z_t)\right| \leq \kappa \| Z_t^{\xi_{[t-1]}} \|, \tag{2.14}$$

where $Z_t^{\xi_{[t-1]}}(\cdot) := Z_t(\xi_{[t-1]}, \cdot)$ is an element of the space $\mathcal{Z}$ for a given $\xi_{[t-1]}$. Applying this bound backward in time for $t = T, T-1, ...$, we obtain that the nested functional $\mathfrak{R}_T : \mathcal{Z}_T \to \mathbb{R}$ is finite valued. Consequently by duality arguments, there exists a set $\widehat{\mathfrak{M}}_T$ of probability measures on $(\Xi_T, \mathcal{F}_T)$, corresponding to the dual representation of $\mathfrak{R}_T$, such that the nested functional has the following representation

$$\mathfrak{R}_T(Z_T) = \sup_{\mathcal{Q}_T \in \widehat{\mathfrak{M}}_T} \left\{ \mathbb{E}_{\mathcal{Q}_T}[Z_T] = \int Z_T(\xi_1, ..., \xi_T) \mathcal{Q}_T(d\xi_1, ..., d\xi_T) \right\}, \ Z_T \in \mathcal{Z}_T. \tag{2.15}$$

The set $\widehat{\mathfrak{M}}_T$ is a subset of the dual space $\mathcal{Z}_T^*$, and can be taken to be convex, weakly* closed. Moreover, since $\mathfrak{R}_T$ is finite valued, the set $\widehat{\mathfrak{M}}_T \subset \mathcal{Z}_T^*$ is bounded, and hence is weakly* compact. Note that for any $\mathcal{Q}_T = Q_1 \times \cdots \times Q_T \in \mathfrak{M}_T$,

$$\mathfrak{R}_T(Z_T) \geq \mathbb{E}_{\mathcal{Q}_T}[Z_T]. \tag{2.16}$$

The set $\widehat{\mathfrak{M}}_T$ has a complicated structure. By (2.16) the set $\widehat{\mathfrak{M}}_T$ is larger than the set $\mathfrak{M}_T$ of the respective product measures. In general the set $\widehat{\mathfrak{M}}_T$ does not have the rectangular structure, i.e., could contain probability distributions which are not stagewise independent. We refer to [4] and [10, section 7.6.2] for a discussion of construction and properties of such nested functionals and the corresponding sets $\widehat{\mathfrak{M}}_T$.

Consider now a sequence $Z_t \in \mathcal{Z}_t$, $t = 1, ...,$ of bounded functions, i.e., there is a constant υ such that $|Z_t(\cdot)| \leq \upsilon$ for all t. Let $\gamma \in (0, 1)$ and $S_T := \sum_{t=1}^T \gamma^{t-1} Z_t$. Then

$$\mathfrak{R}_T(S_T) = \rho_1\Big(Z_1 + \gamma \rho_{2|\xi_{[1]}}\big(Z_2 + \cdots + \gamma \rho_{T|\xi_{[T-1]}}(Z_T)\big)\Big), \tag{2.17}$$

and the limit

$$\mathfrak{R}_\infty(S_\infty) := \lim_{T \to \infty} \mathfrak{R}_T(S_T) \tag{2.18}$$

exists and is finite (cf., [10, section 8.4]).

3 Distributionally robust SOC

In this section we discuss a distributionally robust counterpart of problem (1.1). *Unless stated otherwise we assume that the probability distributions of the process ξ_t, $t = 1, ...,$ are independent of our decisions.* Let us recall that (assuming that the random process ξ_t is iid) with the infinite horizon (when $T = \infty$) risk neutral problem (1.1) is associated Bellman equation for the value function:

$$V(x) = \inf_{u \in \mathcal{U}} \mathbb{E}\left[f(x, u, \xi) + \gamma V\big(F(x, u, \xi)\big)\right]. \tag{3.1}$$

Assuming further that the cost function is bounded, we have that equation (3.1) has a unique solution $\bar{V}(\cdot)$, and an optimal policy for problem (1.1) is given by $u_t = \pi(x_t)$ with

$$\pi(x) \in \arg\min_{u \in \mathcal{U}} \mathbb{E}\left[f(x, u, \xi) + \gamma \bar{V}\big(F(x, u, \xi)\big)\right]. \tag{3.2}$$

Let us consider a distributionally robust counterpart of problem (1.1) starting with the finite horizon case. Consider the nested functional $\mathfrak{R}_T$, of the form (2.17), and following distributionally robust problem

$$\min_{\pi \in \Pi} \ \mathfrak{R}_T\left(\textstyle\sum_{t=1}^T \gamma^{t-1} f(x_t, u_t, \xi_t)\right). \tag{3.3}$$

The optimization (minimization) in (3.3) is performed over feasible policies $\pi \in \Pi$ satisfying the constraints (1.2). In the L_p-setting these constraints should be satisfied almost surely with respect to the reference probability measure, and in the $C(\Xi)$-setting these constraints should be satisfied for all $\xi_t \in \Xi$.

Dynamic programming equations for problem[2] (3.3) are $V_{T+1}(\cdot) \equiv 0$, and for $t = T, ..., 1$,

$$V_t(x_t) = \inf_{u_t \in \mathcal{U}} \sup_{Q \in \mathcal{M}} \mathbb{E}_Q\left[f(x_t, u_t, \xi_t) + \gamma V_{t+1}\big(F(x_t, u_t, \xi_t)\big)\right]. \tag{3.4}$$

An optimal policy for problem (3.3) is given by $u_t = \pi_t(x_t)$, $t = 1, ..., T$, with

$$\pi_t(x_t) \in \arg\min_{u_t \in \mathcal{U}} \sup_{Q \in \mathcal{M}} \mathbb{E}_Q\left[f(x_t, u_t, \xi_t) + \gamma V_{t+1}\big(F(x_t, u_t, \xi_t)\big)\right]. \tag{3.5}$$

If the functional $\mathcal{R}$ is *strictly* monotone, then (3.5) are necessary as well as sufficient conditions for optimality of solutions of the distributionally robust problem (3.3). Without strict monotonicity it could happen that the distributionally robust problem possesses optimal solutions (optimal policies) which do not satisfy the dynamic programming equations (cf., [8]).

Using the dual representation (2.15) of $\mathfrak{R}_T$, we can write problem (3.3) as

$$\min_{\pi \in \Pi} \sup_{\mathcal{Q}_T \in \widehat{\mathfrak{M}}_T} \ \mathbb{E}^{\pi}_{\mathcal{Q}_T}\left[\textstyle\sum_{t=1}^T \gamma^{t-1} f(x_t, u_t, \xi_t)\right], \tag{3.6}$$

A dual of problem (3.6) is obtained by interchanging the 'min' and 'max' operators in (3.6), that is

$$\max_{\mathcal{Q}_T \in \widehat{\mathfrak{M}}_T} \inf_{\pi \in \Pi} \ \mathbb{E}^{\pi}_{\mathcal{Q}_T}\left[\textstyle\sum_{t=1}^T \gamma^{t-1} f(x_t, u_t, \xi_t)\right], \tag{3.7}$$

[2] Recall that $\mathcal{R}(\cdot) = \sup_{Q \in \mathcal{M}} \mathbb{E}_Q[\,\cdot\,]$.

The optimal value of the dual problem (3.7) is less than or equal to the optimal value of problem (3.6). Suppose that there is no duality gap between problems (3.6) and (3.7), i.e., their optimal values are equal to each other and finite. If the considered problem is convex, then this holds under mild regularity conditions. Since the set $\widehat{\mathfrak{M}}_T$ is weakly* compact, the dual problem has an optimal solution $\mathcal{Q}_T^*$, provided $\sum_{t=1}^T \gamma^{t-1} f(x_t, u_t, \xi_t) \in \mathcal{Z}_T$ for all $(x_t, u_t) \in \mathcal{X} \times \mathcal{U}$. It follows that any optimal solution (optimal policy) of problem (3.3) is also an optimal solution of the problem

$$\min_{\pi \in \Pi} \ \mathbb{E}^{\pi}_{\mathcal{Q}_T^*} \left[\sum_{t=1}^T \gamma^{t-1} f(x_t, u_t, \xi_t) \right]. \tag{3.8}$$

As it was mentioned before, the set $\widehat{\mathfrak{M}}_T$ could contain probability distributions which are not stagewise independent. Therefore there is no guarantee that the probability distribution $\mathcal{Q}_T^*$ of $(\xi_1, ..., \xi_T)$ is formed by independent of each other marginal distributions of $\xi_1, ..., \xi_T$.

Infinite horizon case. Suppose that the cost function $f : \mathcal{X} \times \mathcal{U} \times \Xi \to \mathbb{R}$ is bounded. Consider Bellman operator (mapping) $\mathfrak{T} : \mathbb{B}(\mathcal{X}) \to \mathbb{B}(\mathcal{X})$, defined as

$$\mathfrak{T}(g)(\cdot) := \inf_{u \in \mathcal{U}} \mathcal{R}\big(f(\cdot, u, \xi) + \gamma g(F(\cdot, u, \xi))\big), \ g \in \mathbb{B}(\mathcal{X}). \tag{3.9}$$

Note that indeed $\mathfrak{T}$ maps $g \in \mathbb{B}(\mathcal{X})$ into $\mathfrak{T}(g) \in \mathbb{B}(\mathcal{X})$ since it is assumed that the cost function is bounded. It is straightforward to verify that the operator $\mathfrak{T}$ is monotone and has the constant shift property (e.g., [10, Proposition 8.9]), and thus (cf., [3])

$$\|\mathfrak{T}(g) - \mathfrak{T}(g')\|_\infty \le \gamma \|g - g'\|_\infty, \ \ g, g' \in \mathbb{B}(\mathcal{X}). \tag{3.10}$$

That is, $\mathfrak{T}$ is a contraction mapping, and hence has unique fixed point. Consequently the equation

$$V(x) = \inf_{u \in \mathcal{U}} \sup_{Q \in \mathcal{M}} \mathbb{E}_Q \left[f(x, u, \xi) + \gamma V\big(F(x, u, \xi)\big) \right] \tag{3.11}$$

has unique solution $\bar{V} \in \mathbb{B}(\mathcal{X})$. The above equation (3.11) is the distributionally robust counterpart of Bellman equation (3.1).

The dynamic programming equations (3.4) can be written as

$$V_t = \mathfrak{T}(V_{t+1}), \ t = 1, ..., T. \tag{3.12}$$

Since $\mathfrak{T}$ has the contraction property (3.10), it follows that for $t < T$,

$$\|V_t - \bar{V}\|_\infty \le (1 - \gamma)^{-1} \gamma^{T-t} \|V_T - V_{T+1}\|_\infty. \tag{3.13}$$

Note that since the cost function is bounded, $\|V_T - V_{T+1}\|_\infty = \sup_{x \in \mathcal{X}} \big| \inf_{u \in \mathcal{U}} \mathcal{R}(f(x, u, \xi)) \big|$ is a finite constant. We also use notation $V_{t,T} = V_t$ to emphasize that the value functions $V_{t,T}$, $t = 1, ..., T$, are computed for the horizon T. It follows that for any fixed $t \ge 1$,

$$\lim_{T \to \infty} \|V_{t,T} - \bar{V}\|_\infty = 0. \tag{3.14}$$

Remark 3.1 (Decision dependent distributions) So far we assumed that the probability distributions of the process ξ_t are independent of our decisions. Consider now the situations where the distributions of the data process can be functions of state and control variables. We still assume the rectangular setting, but now the ambiguity set $\mathcal{M}_{x,u}$ could be a function of $(x, u) \in \mathcal{X} \times \mathcal{U}$. We still can write the respective dynamic programming equations of the form (3.4) with the set $\mathcal{M}$ replaced by $\mathcal{M}_{x_t,u_t}$, that is

$$V_t(x_t) = \inf_{u_t \in \mathcal{U}} \sup_{Q \in \mathcal{M}_{x_t,u_t}} \mathbb{E}_Q \left[f(x_t, u_t, \xi_t) + \gamma V_{t+1}\big(F(x_t, u_t, \xi_t)\big) \right]. \tag{3.15}$$

The corresponding nested functional $\mathfrak{R}^\pi_T$ then depends on policy $\pi \in \Pi$ and is defined on the respective histories of the decision process. This observation was used in [5] for formulation of a risk averse approach to Markov Decision Processes (MDPs) by employing counterparts of coherent risk measures defined on the histories of the decision process.

On the other hand, we can use the change of variables approach discussed in section 2.1. That is, for $\theta = (x, u)$ and $\Theta = \mathcal{X} \times \mathcal{U}$, suppose that the ambiguity set $\mathcal{M}_{x,u}$ is given in the form (2.5) with the respective mappings $G_{x,u} : \Xi' \to \Xi$. Then by (2.7) the dynamic programming equations (3.15) can be written as

$$V_t(x_t) = \inf_{u_t \in \mathcal{U}} \sup_{Q' \in \mathcal{M}'} \mathbb{E}_{Q'} \left[f(x_t, u_t, G_{x_t,u_t}(\eta_t)) + \gamma V_{t+1}\big(F(x_t, u_t, G_{x_t,u_t}(\eta_t))\big) \right]. \tag{3.16}$$

Consequently the problem can be reformulated in the distributionally robust form with the cost function $f'(x, u, \eta) := f(x, u, G_{x,u}(\eta))$ and state mapping $F'(x, u, \eta) := F(x, u, G_{x,u}(\eta))$, $(x, u, \eta) \in \mathcal{X} \times \mathcal{U} \times \Xi'$. The reformulated problem is of the form (3.3) with probability distributions of the process $\eta_t \in \Xi'$ being independent of our decisions. In the L_p-setting and law invariant cases it suffices to make change of variables for the reference probability measure alone.

This makes the analysis simpler since then there is no need to develop a separate theory for the case of decision dependent distributions. It should be noted, however, that there is a cost involved since such reformulation could destroy a useful structure of the considered problem. We will discuss this further in Remark 4.2 of the next section.

4 Inventory model

In this section we discuss the classical inventory model (cf., [12]). It can be formulated in the framework of control model with the cost function

$$f(x, u, D) := cu + \psi(x + u - D), \tag{4.1}$$

and the mapping $F(x, u, D) := x + u - D$. Here

$$\psi(z) := b[-z]_+ + h[z]_+, \tag{4.2}$$

$x \in \mathbb{R}$ is the current inventory level, $u \in \mathbb{R}_+$ is the order quantity, D is the demand and $\gamma \in (0, 1)$ is the discount factor. We assume that $b > c \geq 0$, $h \geq 0$ and that $D \in \Xi$, where $\Xi \subset \mathbb{R}_+$ is an interval.

The T-stage distributionally robust inventory model can be written as

$$\min_{\pi \in \Pi} \; \mathfrak{R}_T \left(\sum_{t=1}^T \gamma^{t-1} \big(cu_t + \psi(x_t + u_t - D_t) \big) \right), \tag{4.3}$$

where Π is the set of (feasible) policies satisfying the feasibility constraints

$$u_t \geq 0 \;\; \text{and} \;\; x_{t+1} = x_t + u_t - D_t. \tag{4.4}$$

As before, we denote by $\mathcal{M}$ the ambiguity set of probability distributions, of the demand D, supported on Ξ, and by $\mathfrak{R}_T$ the corresponding nested functional. The rectangularity assumption, used in the construction of $\mathfrak{R}_T$, is the distributionally robust counterpart of the assumption that the demand process D_t is stagewise independent, used in the risk neutral setting.

Problem (4.3) is a convex problem, the corresponding value functions

$$V_t(x_t) = \inf_{u_t \geq 0} \sup_{Q_t \in \mathcal{M}} \mathbb{E}_{Q_t} \big[cu_t + \psi(x_t + u_t - D_t) + \gamma V_{t+1}(x_t + u_t - D_t) \big] \tag{4.5}$$

are convex. Suppose that the interval Ξ is bounded. Then the value functions $V_t(\cdot)$ are finite valued and hence continuous. By using convexity of the value functions it is possible to show, in a way similar to the risk neutral case, that the basestock policy is optimal for problem (4.3) (cf., [1]). That is, let us make change of variables $v_t = x_t + u_t$ and consider a point $v_t^* \in \mathfrak{V}_t$, $t = 1, ..., T$, where

$$\mathfrak{V}_t := \arg\min_{v_t \in \mathbb{R}} \sup_{Q_t \in \mathcal{M}} \mathbb{E}_{Q_t} \big[cv_t + \psi(v_t - D_t) + \gamma V_{t+1}(v_t - D_t) \big]. \tag{4.6}$$

Then an optimal policy $u_t = \pi_t(x_t)$, satisfying the dynamic programming equations, is given by

$$\pi_t(x_t) := [v_t^* - x_t]_+. \tag{4.7}$$

That is, if the inventory level x_t, at time t, is smaller than the critical value v_t^*, then the basestock policy suggests to order such amount as to fill the inventory level to the critical value. On the other hand, if $x_t \geq v_t^*$, then order nothing. Note that the critical values v_t^* cannot be negative and cannot be larger than $d_{\max}$. That is, under the specified assumptions the set $\mathfrak{V}_t$ is nonempty and is contained in the interval Ξ, $t = 1, ..., T$.

Remark 4.1 It could be noted that if the distributionally robust functional $\mathcal{R}$ is not *strictly* monotone, then there may exist many optimal policies which are not basestock and do not satisfy the dynamic programming equations (cf., [11]).

Duality considerations Consider the dual of the right hand side of (4.6):

$$\max_{Q_t \in \mathcal{M}} \inf_{v_t \in \mathbb{R}} \mathbb{E}_{Q_t} \big[cv_t + \psi(v_t - D_t) + \gamma V_{t+1}(v_t - D_t) \big]. \tag{4.8}$$

Since under the specified assumptions, the set of minimizers in the right hand side of (4.6) is nonempty and bounded, by Sion's minimax theorem there is no duality gap between problems

(4.6) and (4.8). Also since the set $\mathcal{M}$ is weakly* compact, the dual problem (4.8) possesses an optimal solution $\bar{Q}_t \in \mathcal{M}$. It follows that

$$\mathfrak{V}_t \subset \operatorname*{arg\,min}_{v_t \in \mathbb{R}} \mathbb{E}_{\bar{Q}_t}\big[cv_t + \psi(v_t - D_t) + \gamma V_{t+1}(v_t - D_t)\big]. \tag{4.9}$$

Consequently the basestock policy (4.7) is also optimal for the risk neutral problem

$$\min_{\pi \in \Pi} \quad \mathbb{E}^{\pi}_{\bar{Q}_t}\left[\sum_{t=1}^{T} \gamma^{t-1}\big(cu_t + \psi(x_t + u_t - D_t)\big)\right], \tag{4.10}$$

with D_t being stagewise independent having distribution $\bar{Q}_t$ supported on the interval Ξ. However, by (2.16) the optimal value of the above risk neutral problem (4.10) could be smaller than the optimal value of the distributionally robust problem (4.3).

It is also possible to consider dual of the form (3.7) with $\widehat{\mathfrak{M}}_T$ being the respective set of probability distributions of $(D_1, ..., D_T)$ supported on $\Xi \times \cdots \times \Xi$. By convexity of problem (4.3), there is no duality gap in this dual setting. Let $\mathcal{Q}^*_T$ be an optimal solution of the dual problem. Then any optimal solution of problem (4.3) is also an optimal solution of problem (compare with (3.8))

$$\min_{\pi \in \Pi} \quad \mathbb{E}^{\pi}_{\mathcal{Q}^*_T}\left[\sum_{t=1}^{T} \gamma^{t-1}\big(cu_t + \psi(x_t + u_t - D_t)\big)\right], \tag{4.11}$$

and the optimal values of problems (4.3) and (4.11) are equal to each other.

Infinite horizon inventory model Consider now the infinite horizon version of problem (4.3) with $T = \infty$. The corresponding Bellman equation can be written as

$$V(x) = \inf_{u \geq 0} \mathcal{R}\big(cu + \psi(x + u - D) + \gamma V(x + u - D)\big). \tag{4.12}$$

Note that since the demand is bounded, we can assume the required boundedness condition. Let $\bar{V}$ be the solution (fixed point) of equation (4.12). The base basestock policy is optimal here as well. This follows by going to the limit as $T \to \infty$ in the finite horizon case and using (3.14). That is, consider

$$v^* \in \operatorname*{arg\,min}_{v \in \mathbb{R}} \mathcal{R}\big(cv + \psi(v - D) + \gamma \bar{V}(v - D)\big). \tag{4.13}$$

Then $u = \pi(x)$ with

$$\pi(x) := [v^* - x]_+ \tag{4.14}$$

defines an optimal policy for the considered stationary problem.

Proposition 4.1 *The basestock policy* (4.14) *is optimal for the risk averse stationary inventory problem with the critical value v^* given by*

$$v^* \in \operatorname*{arg\,min}_{v \in \mathbb{R}} \mathcal{R}\big(\gamma cD + (1 - \gamma)cv + \psi(v - D)\big). \tag{4.15}$$

Proof Substituting $\bar{u}(x) = [v^* - x]_+$ into the right hand side of (4.12) we obtain,

$$\begin{aligned} V(x) &= c(v^* - x) + \mathcal{R}\big(\psi(v^* - D) + \gamma V(v^* - D)\big), \text{ for } x \le v^*, && (4.16)\\ V(x) &= \mathcal{R}\big(\psi(x - D) + \gamma V(x - D)\big), \text{ for } x \ge v^*. && (4.17)\end{aligned}$$

Note that by (4.13) we have for any $v \in \mathbb{R}$ and $V = \bar{V}$ that

$$\mathcal{R}\big(cv + \psi(v - D) + \gamma V(v - D)\big) \ge \mathcal{R}\big(cv^* + \psi(v^* - D) + \gamma V(v^* - D)\big),$$

and hence by the translation equivariance

$$\mathcal{R}\big(\psi(v - D) + \gamma V(v - D)\big) \ge c(v^* - v) + \mathcal{R}\big(\psi(v^* - D) + \gamma V(v^* - D)\big).$$

It follows that

$$V(x) \le V(x') \text{ for any } x \le v^* \text{ and } \ x' \ge v^*. \tag{4.18}$$

Since D is nonnegative we have that $v^* - D \le v^*$, and hence by (4.16) that

$$V(v^* - D) = cD + \mathcal{R}\big(\psi(v^* - D) + \gamma V(v^* - D)\big). \tag{4.19}$$

Consequently by using the translation equivariance property of $\mathcal{R}$, we can write for $x \le v^*$:

$$\begin{aligned} V(x) &= c(v^* - x) + \mathcal{R}\left(\psi(v^* - D) + \gamma\big(cD + \mathcal{R}\big(\psi(v^* - D) + \gamma V(v^* - D)\big)\big)\right)\\ &= c(v^* - x) + \mathcal{R}\big(\gamma cD + \psi(v^* - D)\big) + \gamma\mathcal{R}\big(\psi(v^* - D) + \gamma V(v^* - D)\big)\\ &= c(v^* - x) + (1 + \gamma)\mathcal{R}\big(\gamma cD + \psi(v^* - D)\big) + \gamma^2\mathcal{R}\big(\psi(v^* - D) + \gamma V(v^* - D)\big).\end{aligned}$$

By continuing this process we obtain in the limit that for $x \le v^*$,

$$V(x) = c(v^* - x) + (1 - \gamma)^{-1}\mathcal{R}\big(\gamma cD + \psi(v^* - D)\big). \tag{4.20}$$

Thus for $v \le v^*$

$$\mathcal{R}\big(cv + \psi(v - D) + \gamma V(v - D)\big) = \mathcal{R}\big(\gamma cD + (1 - \gamma)cv + \psi(v - D)\big) + \mathcal{C},$$

where $\mathcal{C} := \gamma(1 - \gamma)^{-1}\mathcal{R}\big(\gamma cD + (1 - \gamma)cv^* + \psi(v^* - D)\big)$ does not depend on v. Together with (4.18) this implies (4.15). ∎

In the risk neutral case when $\mathcal{R} = \mathbb{E}$, the right hand side of (4.15) becomes $\gamma c\mathbb{E}[D] + (1 - \gamma)cv + \mathbb{E}[\psi(v - D)]$, and hence

$$v^* \in \operatorname*{arg\,min}_{v\in\mathbb{R}} \big\{(1 - \gamma)cv + \mathbb{E}[\psi(v - D)]\big\}. \tag{4.21}$$

Consequently in the risk neutral stationary case $v^* = F^{-1}(\kappa)$, where $\kappa := \frac{b-(1-\gamma)c}{b+h}$ (cf., [12, Section 9.4.4]). Here $F(z) := \mathsf{Pr}(D \le z)$ is the cumulative distribution function of the demand D and $F^{-1}(\kappa) = \inf\{z : F(z) \ge \kappa\}$ is the corresponding quantile. Note that $\kappa \in (0, 1)$ since $b > c \ge 0$ and $h \ge 0$.

Remark 4.2 (Decision dependent distributions) So far we assumed that the ambiguity set of demand distributions does not depend on our decisions. Suppose now that the ambiguity set $\mathcal{M}_{x,u}$ is a function of (x, u). That is, current level of the inventory and our decision of how much to order could change probability distribution of the demand. Suppose that the set $\mathcal{M}_{x,u}$ is determined by mappings $D = G_{x,u}(D')$ and a set $\mathcal{M}'$ of probability distributions of D' supported on an interval $\Xi' \subset \mathbb{R}_+$. Consequently, as it was discussed in Remark 3.1, the problem can be reformulated into a problem with the decision independent distributions by changing the cost function to $cu+\psi(x+u-G_{x,u}(D'))$ and the state (balance) equation to $x_{t+1} = x_t+u_t-G_{x,u}(D')$. Typically such change of variables will destroy convexity of the considered problem. An exception, when the convexity is preserved, is when the mapping $G_{x,u}$ is linear (affine) in (x, u).

5 Comments and open questions

Two dual formulations, related to the distributionally robust inventory problem (4.3), were considered in section 4. The dual problem (4.8) was applied to the min-max problem defining the critical value v_t^* at stage $t = 1, ..., T$. An optimal solution (probability measure) of (4.8) determines the corresponding risk neutral problem (4.10). The natural question is whether the optimal value of problem (4.10) is equal to the optimal value of the distributionally robust problem (4.3)? The general considerations can only guarantee that the optimal value of problem (4.10) cannot be larger than the optimal value of problem (4.3).

We also can consider the following dual of the right hand side problem of the dynamic equation (4.5):

$$\max_{Q_t\in\mathcal{M}} \inf_{u_t\geq 0} \mathbb{E}_{Q_t}\big[cu_t + \psi(x_t + u_t - D_t) + \gamma V_{t+1}(x_t + u_t - D_t)\big]. \tag{5.1}$$

Note however that unlike the min-max problem (4.10), an optimal solution (probability measure) of problem (5.1) could depend on x_t. In such cases this duality, related to the value functions, cannot be extended in a reasonable way to duality of the distributionally robust problem (4.3). In some rather exceptional cases, problem (5.1) possesses an optimal solution which does not depend on x_t. One such example is when the set $\mathcal{M}$ consists of probability measures supported on the (bounded) interval $\Xi = [\alpha, \beta]$ and having mean $\mu \in \Xi$. In that case problems (4.8) and (5.1) have the same optimal solution (probability distribution) $\bar{Q}_t$ which is supported on two end points of the interval $[\alpha, \beta]$ with the respective weights (probabilities) $1 - \mu/(\beta - \alpha)$ and $\mu/(\beta - \alpha)$ (cf., [7]). For that set $\mathcal{M}$, the optimal value of the risk neutral problem (4.10) is equal to the optimal value of the distributionally robust problem (4.3). However, although these two problems have the same optimal value, the distributionally robust problem (4.3) could have many optimal solutions which do not satisfy the dynamic programming equations and are not basestock policies (cf., [11]).

The dual formulation of the form (3.7) is applied directly to the respective distributionally robust problem. Given an optimal solution $\mathcal{Q}_T^*$ of that dual problem, the optimal value of problem (4.3) is equal to the optimal value of the risk neutral problem (4.11), and any optimal solution (optimal policy) of problem (4.3) is also optimal for the risk neutral problem (4.11). As it was already mentioned, the corresponding set $\widehat{\mathfrak{M}}_T$ is quite complicated and could contain probability distributions which are not stagewise independent. On the other hand, the basestock policy (4.7) is optimal for the distributionally robust problem and hence is optimal for problem

(4.11) as well. So the natural question: what is the structure of the extreme distribution $\mathcal{Q}_T^*$ of the random vector $(D_1, ..., D_T)$, in particular is it stagewise independent? In principle there is no contradiction between the basestock policy being optimal for problem (4.11) and $\mathcal{Q}_T^*$ being *not* stagewise independent.

Similar questions can be asked for the infinite horizon case with the respective Bellman equation (4.12). We can write the following dual of the right hand side problem of (4.13)

$$\max_{Q\in\mathcal{M}} \inf_{v\in\mathbb{R}} \mathbb{E}_Q\big[cv + \psi(v - D) + \gamma\bar{V}(v - D)\big]. \tag{5.2}$$

Similar to (4.10), let $\bar{Q}$ be an optimal solution of the above dual problem (5.2) and consider the corresponding infinite horizon risk neutral problem

$$\min_{\pi\in\Pi} \quad \mathbb{E}^{\pi}_{\bar{Q}}\left[\sum_{t=1}^{\infty} \gamma^{t-1}\big(cu_t + \psi(x_t + u_t - D_t)\big)\right]. \tag{5.3}$$

It follows that the basestock policy (4.14) is also optimal for the risk neutral infinite horizon problem (5.3) with demand D having distribution $\bar{Q}$. On the other hand, Bellman equation for problem (5.3) is different from Bellman equation (4.12) for the distributionally robust inventory problem. Again the optimal value of problem (5.3) cannot be larger than the optimal value of the distributionally robust inventory problem. The natural question is in what cases these optimal values are equal to each other? As it was mentioned above one such example is when the set $\mathcal{M}$ consists of probability measures supported on the interval $[\alpha, \beta]$ and having mean $\mu \in [\alpha, \beta]$.

The basic condition in the above analysis is that the distributions (the ambiguity sets) of the data process ξ_t do not depend on our actions (controls). Under this assumption we only need to consider distributionally robust functionals and policies which are functions of the data process. If we allow the distributions of ξ_t to depend on decisions (controls), then the analysis becomes considerably more delicate. Also such dependence on our decisions typically destroys the convexity properties of the considered problem (the convexity was crucial for the duality considerations). In the MDP setting the corresponding stochastic structure is determined by transition probabilities which are functions of the state and actions (controls). An approach to risk averse MDP formulation was suggested in [5], with risk measures are build on histories of the decision process which depend on a considered policy (for a brief discussion of this in SOC and MDP settings we also can refer to [9]). This makes the analysis considerably more involved. On the other hand, as it was discussed in Remark 3.1 (see also Remark 4.2) in the SOC setting in some cases it is possible to deal with this by making change of variables. A relation of this to the MDP formalism requires a further clarification.

References

[1] S. Ahmed, U. Cakmak, and A. Shapiro. Coherent risk measures in inventory problems. *European Journal of Operational Research*, 182:226–238, 2007.

[2] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[3] D.P. Bertsekas and S.E. Shreve. *Stochastic Optimal Control, The Discrete Time Case.* Academic Press, New York, 1978.

[4] A. Pichler and A. Shapiro. Mathematical foundations of distributionally multistage optimization. *SIAM J. Optimization*, 31:3044—3067, 2021.

[5] A. Ruszczyński. Risk-averse dynamic programming for markov decision processes. *Mathematical Programming*, 125:235 – 261, 2010.

[6] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31:433–452, 2006.

[7] A. Shapiro. Minimax and risk averse multistage stochastic programming. *European Journal of Operational Research*, 219:719–726, 2012.

[8] A. Shapiro. Interchangeability principle and dynamic equations in risk averse stochastic programming. *Operations Research Letters*, 45:377–381, 2017.

[9] A. Shapiro. Distributionally robust optimal control and MDP modeling. *Operations Research Letters*, 49:809–814, 2021.

[10] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory.* SIAM, Philadelphia, third edition, 2021.

[11] A. Shapiro and L. Xin. Time inconsistency of optimal policies of distributionally robust inventory models. *Operations Research*, 68:1576–1584, 2020.

[12] P.H. Zipkin. *Foundations of Inventory Management.* McGraw-Hil, 2000.