Toward a Scalable Upper Bound for a CVaR-LQ Problem

Margaret P. Chapman[†] and Laurent Lessard[‡]

***Abstract*— We study a linear-quadratic, optimal control problem on a discrete, finite time horizon with distributional ambiguity, in which the cost is assessed via Conditional Value-at-Risk (CVaR). We take steps toward deriving a scalable dynamic programming approach to upper-bound the optimal value function for this problem. This dynamic program yields a novel, tunable risk-averse control policy, which we compare to existing state-of-the-art methods.**

***Index Terms*— Stochastic optimal control, LMIs, Linear systems**

I. Introduction

THE standard approach to stochastic optimal control is to evaluate a random cumulative cost in expectation. However, this approach is not designed to protect against worst-case circumstances. This limitation motivates robust optimal control [1], [2] and related methods, such as minimax model predictive control [3] and mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control [4].

Robust methods typically assume bounded disturbances, which excludes certain common noise models, such as Gaussian noise. A technique to alleviate this restriction is to use a *risk-averse* formulation, in which a random cost is assessed via *exponential utility*. Here, the objective takes the form $\mathcal{J}_\gamma(x,\pi) := \frac{1}{\gamma}\log\left(E_x^\pi(e^{\gamma Z/2})\right)$, where Z is a random cumulative cost, π is a control policy, x is an initial condition, and $\gamma > 0$ is a risk-aversion parameter.[1] This problem has been studied in increasing levels of generality from the 1970s to the 2010s, e.g., see [5]–[10]. As γ increases, the criterion $\mathcal{J}_\gamma(x,\pi)$ represents a more risk-averse perspective, while as γ approaches zero, $\mathcal{J}_\gamma(x,\pi)$ tends to the usual expected cost.

In the case of linear dynamics with Gaussian noise and quadratic costs, the problem of optimizing $\mathcal{J}_\gamma(x,\pi)$ is commonly called LEQR control. For a fixed $\gamma > 0$, a Riccati recursion is used to derive the optimal value functions and the optimal control law, which is linear state-feedback [7]. At each step t of the recursion, it must be the case that the matrix $\Sigma^{-1} - \gamma\bar{P}_{t+1}$ is positive definite, where Σ is the covariance of the process noise, and $\bar{P}_{t+1}$ is the matrix obtained from step $t+1$. If γ is chosen too large, then the above condition may be violated, and the controller synthesis procedure breaks down. While it is known that $\mathcal{J}_\gamma(x,\pi)$ approximates a weighted sum of the expectation $E_x^\pi(Z)$ and the variance $\text{var}_x^\pi(Z)$ if $\gamma\text{var}_x^\pi(Z)$ is "small" [7], a more precise interpretation of $\mathcal{J}_\gamma(x,\pi)$ has not been established.

[†]M.P.C. is with the University of Toronto, Toronto, Canada. Contact email: mchapman@ece.utoronto.ca

[‡]L.L. is with Northeastern University, Boston, Massachusetts, USA. Contact email: l.lessard@northeastern.edu

[1]One may consider $\gamma < 0$, which corresponds to a *risk-seeking* perspective. We focus on the *risk-averse* perspective here, which assumes that noise leads to harm rather than benefit.

The *Conditional Value-at-Risk* (CVaR) functional, which was invented in the early 2000s by the financial engineering community [11], has potential to alleviate the above issues. The CVaR of Z at level $\alpha \in (0,1]$ represents the expectation of the $\alpha\cdot 100\%$ largest values of Z. The intuitive interpretation of CVaR and its quantitative characterization of risk aversion (in terms of a *fraction* of worst-case outcomes) are two reasons for its popularity in financial engineering (see [12] and the references therein) and its emerging popularity in control (e.g., see [13], [14]). In addition to financial applications, CVaR may be a useful tool for the design of stormwater systems [14], which are required to satisfy precise regulatory specifications, and for the operation of robotic systems [15].

However, the optimization of CVaR is computationally expensive in general. Unlike the expectation of a random (cumulative) cost, the CVaR of a random cost, subject to the dynamics of a Markov decision process, does *not* satisfy a dynamic programming (DP) recursion on the state space. One way to resolve this issue and make DP valid is via a suitable state augmentation [16].

Here, we study a linear-quadratic optimal control problem with *distributional ambiguity*, where the cost is assessed via CVaR. Our first step is to derive an upper bound to the optimal value of this problem. This derivation (Thm. 1) and additional analysis (Thm. 2) motivate the formulation of an interesting dynamic programming algorithm (Thm. 3). While the associated value functions are defined on an augmented state space, they are computed in a *scalable* fashion since their parameters come from a Riccati-like recursion. Moreover, our algorithm provides a risk-averse controller, in which a risk-aversion level is parameterized in a novel way through a positive definite matrix. While our controller synthesis procedure is more computationally complex than LEQR, it does not involve a condition that is analogous to the positive definiteness of $\Sigma^{-1} - \gamma\bar{P}_{t+1}$ for all t.

II. A CVaR-Linear-Quadratic Problem

A. Notation

If $M \in \mathbb{R}^{n\times n}$, then $M \geq 0$ ($M > 0$) means that M is symmetric and positive semi-definite (positive definite). Upper-case letters denote random variables (e.g., X_t), and lower-case letters denote values of random variables (e.g., x_t). If $\mathcal{E}$ is a metric space, $\mathcal{B}(\mathcal{E})$ is the Borel sigma algebra on $\mathcal{E}$. We define $\bar{\mathbb{R}} := \mathbb{R}\cup\{-\infty,+\infty\}$, $\mathbb{R}_+ := \mathbb{R}\cap[0,\infty)$, and $\mathbb{R}_+^n := \{z\in\mathbb{R}^n : z_i \in \mathbb{R}_+,\ i=1,\dots,n\}$. $0_{n\times m}$ is the $n\times m$ zero matrix. I_n is the $n\times n$ identity matrix. The trace of a matrix $M\in\mathbb{R}^{n\times n}$ is $\text{tr}(M)$. $\mathcal{P}(\mathbb{R}^n)$ is the set of probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

B. Linear-Quadratic System Model

Consider a fully observable, linear time-invariant system:

$$X_{t+1} = AX_t + BU_t + W_t \quad \forall t \in \{0, 1, \ldots, N-1\}, \quad (1)$$

where X_t is a $\mathbb{R}^n$-valued random state, U_t is a $\mathbb{R}^m$-valued random control, and W_t is a $\mathbb{R}^n$-valued random disturbance at time t. The matrices $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ and the length of the time horizon $N \in \mathbb{N}$ are given. The initial state X_0 is fixed at an arbitrary $x \in \mathbb{R}^n$. For convenience, define $f(x_t, u_t, w_t) := Ax_t + Bu_t + w_t$ for all $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^m$, and $w_t \in \mathbb{R}^n$.

We make the following assumptions about the $\mathbb{R}^n$-valued disturbance process $(W_0, W_1, \ldots, W_{N-1})$. W_t and W_s are independent for all $t \neq s$, and W_t is independent of the initial state X_0 for each t. For each t, the exact distribution of W_t is not known. However, the first and maximal second moment of W_t are known, which we specify below.

Definition 1 (Ambiguity Set): We define $\mathcal{P}_W \subseteq \mathcal{P}(\mathbb{R}^n)$ to be the set of probability measures with zero mean and covariance upper-bounded by $\Sigma > 0$. Each disturbance W_t has a distribution $\nu_t \in \mathcal{P}_W$. In other words, ν_t satisfies $\int_{\mathbb{R}^n} w_t \, \nu_t(\mathrm{d}w_t) = 0_{n\times 1}$ and $\int_{\mathbb{R}^n} w_t w_t^T \, \nu_t(\mathrm{d}w_t) \leq \Sigma$.

As the system evolves, a random cumulative quadratic cost is incurred. The random cost-to-go for time $t \in \{0, 1, \ldots, N-1\}$ is defined as

$$Z_t := \underbrace{X_N^T Q_f X_N}_{Z_N} + \sum_{j=t}^{N-1} \underbrace{X_j^T Q X_j + U_j^T R U_j}_{C_j}. \quad (2)$$

C_j (Z_N) is the random stage (terminal) cost at time j (N). $Q \in \mathbb{R}^{n\times n}$, $R \in \mathbb{R}^{m\times m}$, and $Q_f \in \mathbb{R}^{n\times n}$ satisfy $Q > 0$, $R > 0$, and $Q_f > 0$, respectively. We define $Z := Z_0$ and $c(x_t, u_t) := x_t^T Q x_t + u_t^T R u_t$ for all $x_t \in \mathbb{R}^n$ and $u_t \in \mathbb{R}^m$.

C. CVaR-Risk-Averse Optimal Control Problem

Consider a CVaR optimal control problem on a discrete, finite time horizon with distributional ambiguity:

$$J_\alpha^*(x) := \inf_{\pi\in\Pi} \sup_{\gamma\in\Gamma} \mathrm{CVaR}_{\alpha,x}^{\pi,\gamma}(Z), \quad (3)$$

subject to the linear dynamics (1), where $x \in \mathbb{R}^n$ is an initial condition and $\alpha \in (0, 1]$ is a risk-aversion level. The objective $\mathrm{CVaR}_{\alpha,x}^{\pi,\gamma}(Z)$ is the CVaR of Z at level α, when the system is initialized at x and evolves according to a control policy $\pi \in \Pi$ and a disturbance strategy $\gamma \in \Gamma$. (γ provides a distribution for W_t for each t. Π and Γ will be defined in this section.) The CVaR of Z represents the expectation of the $\alpha \cdot 100\%$ largest values of Z.

While the problem (3) does not satisfy a dynamic programming (DP) recursion on $\mathbb{R}^n$, there is a useful DP recursion on $\mathbb{R}^n \times \mathbb{R}$ (Lemma 3). A CVaR optimal control problem *without* distributional ambiguity has been solved by defining an augmented state space [16]. Taking inspiration from [16], we use a $\mathbb{R}^n \times \mathbb{R}$-valued, random *augmented state* (X_t, S_t). X_t is defined by (1). S_t is a $\mathbb{R}$-valued random variable

$$S_{t+1} = S_t - C_t \quad \text{for all } t \in \{0, 1, \ldots, N-1\}. \quad (4)$$

S_t keeps track of the random cumulative cost up to time t. The initial augmented state (X_0, S_0) is fixed at an arbitrary $(x, s) \in \mathbb{R}^n \times \mathbb{R}$. We use the *augmented state space* $\mathbb{R}^n \times \mathbb{R}$ to define Π, the class of history-dependent control policies that summarize the history through (X_t, S_t).

Definition 2 (Control Policies Π): A control policy $\pi \in \Pi$ takes the form $\pi := (\pi_0, \pi_1, \ldots, \pi_{N-1})$, such that for each t, π_t is a stochastic kernel on $\mathbb{R}^m$ given $\mathbb{R}^n \times \mathbb{R}$.

Definition 3 (Disturbance Strategies Γ): Every disturbance strategy $\gamma \in \Gamma$ takes the form $\gamma := (\nu_0, \nu_1, \ldots, \nu_{N-1})$, such that $\nu_t \in \mathcal{P}_W$ is the unknown distribution of W_t for each t.

D. Probability Space for Random Cumulative Cost

For any $(x, s) \in \mathbb{R}^n \times \mathbb{R}$, $\pi \in \Pi$, and $\gamma \in \Gamma$, the random cost Z is defined on a probability space $(\Omega, \mathcal{B}(\Omega), P_{x,s}^{\pi,\gamma})$, where the sample space is $\Omega := (\mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}_+)^N \times \mathbb{R}^n \times \mathbb{R}$, and every $\omega \in \Omega$ takes the form $\omega = (x_0, s_0, u_0, c_0, \ldots, x_{N-1}, s_{N-1}, u_{N-1}, c_{N-1}, x_N, s_N)$, where $(x_t, s_t) \in \mathbb{R}^n \times \mathbb{R}$, $u_t \in \mathbb{R}^m$, and $c_t \in \mathbb{R}_+$ are values of (X_t, S_t), U_t, and C_t, respectively. We have specified implicitly that the coordinates of ω have causal dependencies via (1), (2), and (4). The random state at time t is a function $X_t : \Omega \to \mathbb{R}^n$, such that if $\omega \in \Omega$ is as above, then $X_t(\omega) := x_t$, which is Borel measurable. S_t, U_t, and C_t are defined analogously. The probability measure $P_{x,s}^{\pi,\gamma}$ is used to evaluate expectations, e.g., $E_{x,s}^{\pi,\gamma}(Z) := \int_\Omega Z(\omega) \, \mathrm{d}P_{x,s}^{\pi,\gamma}(\omega)$. The form of $P_{x,s}^{\pi,\gamma}$ depends on the dynamics of the augmented state (1) (4), an initial augmented condition $(x, s) \in \mathbb{R}^n \times \mathbb{R}$, a control policy $\pi \in \Pi$, and a disturbance strategy $\gamma \in \Gamma$. E.g., see [17, Prop. C.10, Remark C.11] for a similar construction.

E. Defining CVaR of Random Cumulative Cost

The Conditional Value-at-Risk of $Z := Z_0$ (2) at a risk-aversion level $\alpha \in (0, 1]$ is defined as follows:

$$\mathrm{CVaR}_{\alpha,x}^{\pi,\gamma}(Z) := \begin{cases} \inf\limits_{s\in\mathbb{R}} g_{\alpha,x}^{\pi,\gamma}(s, Z) & \text{if } E_{x,s}^{\pi,\gamma}(Z) < \infty \ \forall s \\ +\infty & \text{otherwise,} \end{cases} \quad (5)$$

where $g_{\alpha,x}^{\pi,\gamma}(s, Z) := s + \frac{1}{\alpha} E_{x,s}^{\pi,\gamma}(\max(Z - S_0, 0))$.

Remark 1: It is standard to define

$$\mathrm{CVaR}_\alpha(Y) := \inf_{s\in\mathbb{R}} s + \tfrac{1}{\alpha} E(\max(Y - s, 0)),$$

where Y is a random variable with finite first moment. In (5), we use an *extended definition* for CVaR to permit a class of policies Π that depends on the augmented state space and need not have a particular analytical form (e.g., linear).

III. Upper Bound for CVaR-LQ Problem

We use the definition of $\mathrm{CVaR}_{\alpha,x}^{\pi,\gamma}(Z)$ (5) to re-express $J_\alpha^*(x)$ (3). For any $x \in \mathbb{R}^n$ and $\alpha \in (0, 1]$, it holds that

$$\begin{aligned} J_\alpha^*(x) &= \inf_{\pi\in\Pi} \sup_{\gamma\in\Gamma} \begin{cases} \inf\limits_{s\in\mathbb{R}} g_{\alpha,x}^{\pi,\gamma}(s, Z) & \text{if } E_{x,s}^{\pi,\gamma}(Z) < +\infty \ \forall s \in \mathbb{R} \\ +\infty & \text{otherwise} \end{cases} \\ &= \underbrace{\inf_{\pi\in\Pi} \sup_{\gamma\in\Gamma}}_{E_{x,s}^{\pi,\gamma}(Z)<+\infty \ \forall s\in\mathbb{R}} \inf_{s\in\mathbb{R}} g_{\alpha,x}^{\pi,\gamma}(s, Z). \end{aligned}$$

$J_{\alpha,\pi}(x)$

In the current section, first we show that there is a policy $\pi \in \Pi$ such that $J_{\alpha,\pi}(x)$ is finite (Lemma 1), which guarantees that the problem (3) is well-defined. Second, we derive an upper bound to $J^*_\alpha(x)$ (Thm. 1):

$$J^*_\alpha(x) \leq \inf_{s\in\mathbb{R}} \; s + \tfrac{1}{\alpha} \underbrace{\inf_{\pi\in\Pi} \sup_{\gamma\in\Gamma} E^{\pi,\gamma}_{x,s}(\max(Z - S_0, 0))}_{V^*_0(x,s)}.$$

Toward the goal of computing V^*_0 scalably, we will define a value iteration algorithm with value functions $V_N, \ldots, V_1, V_0$ (Sec. IV). We will analyze the algorithm in the setting of deterministic policies and finitely many disturbance values. We will show that, under a measurable selection assumption, $\bar{V}^*_0 \leq \bar{V}_0$ (Thm. 2), where $\bar{V}^*_0$ and $\bar{V}_0$ are the versions of V^*_0 and V_0 in the simplified setting, respectively. In Sec. V, we will prove that $V_0 \leq \hat{V}_0$, where

$$\hat{V}_0(x,s) := a_0 + \max(x^T P_0 x - s, 0) \quad \forall (x,s) \in \mathbb{R}^n \times \mathbb{R},$$

such that $a_0 \in \mathbb{R}$ and $P_0 > 0$ are obtained via a Riccati-like recursion (Thm. 3). We will explain how the proof of Thm. 3 provides an algorithm for a novel risk-averse controller. Also, the above analysis takes key steps toward deriving

$$J^*_\alpha(x) \leq \inf_{s\in\mathbb{R}} \; s + \tfrac{1}{\alpha}\hat{V}_0(x,s) \quad \forall x \in \mathbb{R}^n, \; \forall \alpha \in (0,1],$$

a *scalable* upper bound to a CVaR linear-quadratic optimal control problem with distributional ambiguity.

Lemma 1 ($J_{\alpha,\pi}(x)$ is finite for some π): For all $x \in \mathbb{R}^n$ and $\alpha \in (0,1]$, there is a $\pi \in \Pi$ such that $J_{\alpha,\pi}(x) \in \mathbb{R}$.

Proof: Let $\pi \in \Pi$ be an open-loop deterministic policy such that U_t takes the value $0_{m\times 1}$ for each t. By using the quadratic cost, linear dynamics, and the definition of the ambiguity set, it holds that

$$0 \leq E^{\pi,\gamma}_{x,s}(Z) \leq H^{\pi,\Sigma}_x \quad \forall \gamma \in \Gamma, \; \forall s \in \mathbb{R}, \tag{6}$$

where $H^{\pi,\Sigma}_x := x^T Q x + \mathrm{tr}\big(F x x^T F^T \bar{Q}\big) + \mathrm{tr}\big(G\bar{\Sigma}G^T\bar{Q}\big)$. $\bar{\Sigma}$ is a block diagonal matrix containing N copies of Σ. $\bar{Q} := \mathrm{diag}(Q, \ldots, Q, Q_f)$ is a block diagonal matrix with $N-1$ copies of Q. $F \in \mathbb{R}^{Nn\times n}$ and $G \in \mathbb{R}^{Nn\times Nn}$ depend on A and N. The desired statement follows from (6). ∎

By the previous lemma and since $\{J_{\alpha,\pi}(x) : \pi \in \Pi\}$ is bounded below by 0, it holds that $J^*_\alpha(x) \in \mathbb{R}$.

Theorem 1 (Upper bound to $J^*_\alpha(x)$): Define

$$G_\alpha(x) := \inf_{\pi\in\Pi} \inf_{s\in\mathbb{R}} \sup_{\gamma\in\Gamma} g^{\pi,\gamma}_{\alpha,x}(s,Z) \quad \forall x \in \mathbb{R}^n, \forall \alpha \in (0,1],$$
$$V^*_0(x,s) := \inf_{\pi\in\Pi} \sup_{\gamma\in\Gamma} E^{\pi,\gamma}_{x,s}(\max(Z - S_0, 0)) \; \forall x \in \mathbb{R}^n, \forall s \in \mathbb{R}. \tag{7}$$

For all $x \in \mathbb{R}^n$ and $\alpha \in (0,1]$, $J^*_\alpha(x) \leq G_\alpha(x)$, $G_\alpha(x) \in \mathbb{R}$, and $G_\alpha(x) = \inf_{s\in\mathbb{R}} \; s + \frac{1}{\alpha}V^*_0(x,s)$. Moreover, V^*_0 is finite.

Proof: We have $J^*_\alpha(x) \leq G_\alpha(x)$ because

$$\sup_{\gamma\in\Gamma^\pi_x} \inf_{s\in\mathbb{R}} g^{\pi,\gamma}_{\alpha,x}(s,Z) \leq \inf_{s\in\mathbb{R}} \sup_{\gamma\in\Gamma^\pi_x} g^{\pi,\gamma}_{\alpha,x}(s,Z) \quad \forall \pi \in \Pi,$$

where $\Gamma^\pi_x := \{\gamma \in \Gamma : E^{\pi,\gamma}_{x,s}(Z) < +\infty \; \forall s \in \mathbb{R}\}$, and since $\Gamma^\pi_x \subseteq \Gamma$. $G_\alpha(x) \in \mathbb{R}$ because $\{\sup_{\gamma\in\Gamma} \; g^{\pi,\gamma}_{\alpha,x}(s,Z) : \pi \in \Pi, s \in \mathbb{R}\}$ is bounded below and there exist $s \in \mathbb{R}$ and $\pi \in \Pi$ s.t. $\sup_{\gamma\in\Gamma} \; g^{\pi,\gamma}_{\alpha,x}(s,Z) \in \mathbb{R}$. Indeed, let $s = 0$ and let π assign the value $0_{m\times 1}$ to each U_t. Then, we have

$$0 \leq g^{\pi,\gamma}_{\alpha,x}(0,Z) = \tfrac{1}{\alpha}E^{\pi,\gamma}_{x,0}(Z) \leq \tfrac{1}{\alpha}H^{\pi,\Sigma}_x \qquad \forall \gamma \in \Gamma.$$

We have $G_\alpha(x) = \inf_{s\in\mathbb{R}} \; s + \frac{1}{\alpha}V^*_0(x,s)$ because one may exchange the order of infima. V^*_0 is finite because 1) for any $(x,s) \in \mathbb{R}^n \times \mathbb{R}$, there is a $\pi \in \Pi$ s.t. $\sup_{\gamma\in\Gamma} E^{\pi,\gamma}_{x,s}(\max(Z - S_0, 0)) \in \mathbb{R}$, and 2) $\{\sup_{\gamma\in\Gamma} \; E^{\pi,\gamma}_{x,s}(\max(Z - S_0, 0)) : \pi \in \Pi\}$ is bounded below by 0. For the first property, one may choose the policy that assigns the value $0_{m\times 1}$ to each U_t. ∎

IV. Analysis of a Value Iteration Algorithm

To estimate V^*_0 (7) in a scalable fashion, we propose a value iteration algorithm on $\mathbb{R}^n \times \mathbb{R}$.

Algorithm 1 (Value Iteration for General Setting): Let the functions $V_N, V_{N-1}, \ldots, V_0$ be defined recursively as follows. For all $(x,s) \in \mathbb{R}^n \times \mathbb{R}$ and for $t = N-1, \ldots, 0$,

$$V_N(x,s) := \max(x^T Q_f x - s, 0)$$
$$V_t(x,s) := \inf_{u\in\mathbb{R}^m} \sup_{\nu\in\mathcal{P}_W} \textstyle\int_{\mathbb{R}^n} V_{t+1}(f(x,u,w), s - c(x,u))\nu(\mathrm{d}w).$$

Conjecture 1: The functions $V_N, V_{N-1}, \ldots, V_0$ are Borel measurable and bounded below by 0.

We use the Conjecture in the proof of Thm. 3, which requires the Lebesgue integrals in Algorithm 1 to exist. The Conjecture will be proved formally in future work by using properties of convex functions.

In this work, we will analyze Algorithm 1 in the setting of finitely many disturbance values and deterministic policies.

Definition 4 ($\bar{\Pi}$): $\bar{\Pi}$ is the set of deterministic policies such that every $\pi \in \bar{\Pi}$ takes the form $\pi = (\pi_0, \pi_1, \ldots, \pi_{N-1})$, where each $\pi_t : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$ is Borel measurable.

Definition 5 ($\bar{\mathcal{P}}_W$): Let W_t be supported on the N_W points $\{w^1, w^2, \ldots, w^{N_W}\} \subseteq \mathbb{R}^n$, and let $p^j_t \in [0,1]$ be the (unknown) probability that the value of W_t is w^j. In this case, the ambiguity set of distributions is

$$\bar{\mathcal{P}}_W := \left\{ p \in \mathbb{R}^{N_W}_+ \;\middle|\; \begin{array}{l} \sum_{j=1}^{N_W} p^j = 1, \; \sum_{j=1}^{N_W} w^j p^j = 0_{n\times 1}, \\ \sum_{j=1}^{N_W} w^j (w^j)^T p^j \leq \Sigma \end{array} \right\}.$$

Definition 6 ($\bar{\Gamma}$): The set of disturbance strategies in the setting of finitely many disturbance values is $\bar{\Gamma} := \big\{\gamma = (p_0, p_1, \ldots, p_{N-1}) : p_t \in \bar{\mathcal{P}}_W \; \forall t\big\}$.

The version of V^*_0 (7) in the setting of finitely many disturbance values and deterministic policies is

$$\bar{V}^*_0(x,s) := \inf_{\pi\in\bar{\Pi}} \sup_{\gamma\in\bar{\Gamma}} E^{\pi,\gamma}_{x,s}(\max(Z - S_0, 0)) \tag{8}$$

for all $(x,s) \in \mathbb{R}^n \times \mathbb{R}$. The version of Algorithm 1 in the setting of finitely many disturbance values follows.

Algorithm 2 (Value Iteration in Finite Case): Let the functions $\bar{V}_N, \bar{V}_{N-1}, \ldots, \bar{V}_0$ be defined recursively as follows. For all $(x,s) \in \mathbb{R}^n \times \mathbb{R}$ and for $t = N-1, \ldots, 0$,

$$\bar{V}_N(x,s) := \max(x^T Q_f x - s, 0)$$
$$\bar{V}_t(x,s) := \inf_{u\in\mathbb{R}^m} \underbrace{\sup_{p_t\in\bar{\mathcal{P}}_W} \textstyle\sum_{j=1}^{N_W} p^j_t \bar{V}_{t+1}(f(x,u,w^j), s - c(x,u))}_{\bar{\psi}_{t+1}(x,s,u)}.$$

The next theorem specifies properties of Algorithm 2.

Theorem 2 (Analysis of Algorithm 2): For $t = 0, \dots, N$, the value function $\bar{V}_t : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is convex and bounded below by 0, and $\bar{V}_t(x_t, s_t)$ is non-increasing in s_t for each x_t. For $t = 0, 1, \dots, N-1$, for any $(x_t, s_t) \in \mathbb{R}^n \times \mathbb{R}$, there is a $u^*_{x_t,s_t} \in \mathbb{R}^m$ such that

$$\bar{V}_t(x_t, s_t) = \bar{\psi}_{t+1}(x_t, s_t, u^*_{x_t,s_t}). \tag{9}$$

Suppose that there is a Borel measurable function $\pi^*_t : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$ such that for all $(x_t, s_t) \in \mathbb{R}^n \times \mathbb{R}$,

$$\bar{V}_t(x_t, s_t) = \bar{\psi}_{t+1}(x_t, s_t, \pi^*_t(x_t, s_t)). \tag{10}$$

Define $\pi^* := (\pi^*_0, \pi^*_1, \dots, \pi^*_{N-1})$. Then, Algorithm 2 provides an upper bound to $\bar{V}^*_0$ (8), specifically, $\bar{V}^*_0 \le \bar{V}_0$.

Remark 2: Thm. 2 invokes a measurable selection assumption (see also [17, Thm. 3.2.1]), which motivates future study of measurable selection theorems.

To prove Thm. 2, we present two supporting results.

Lemma 2 (Value Function Analysis): Let $v : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be convex and bounded below by 0. Also, let $v(x, s)$ be non-increasing in s for each x. Define $v^* : \mathbb{R}^n \times \mathbb{R} \to \bar{\mathbb{R}}$ as $v^*(x, s) := \inf_{u \in \mathbb{R}^m} \sup_{p \in \bar{\mathcal{P}}_W} \sum_{j=1}^{N_W} p^j \; v\big(f(x, u, w^j), s - c(x, u)\big)$. Then, v^* is finite, convex, and bounded below by 0, and $v^*(x, s)$ is non-increasing in s for each x. Also, for all $(x, s) \in \mathbb{R}^n \times \mathbb{R}$, there is a $u^*_{x,s} \in \mathbb{R}^m$ such that $v^*(x, s) = \sup_{p \in \bar{\mathcal{P}}_W} \sum_{j=1}^{N_W} p^j \; v\big(f(x, u^*_{x,s}, w^j), s - c(x, u^*_{x,s})\big)$.

Proof: Since $f(x, u, w^j)$ is affine in (x, u, s) for each w^j, $s - c(x, u)$ is concave in (x, u, s), v is convex, and $v(x, s)$ is non-increasing in s for each x, $(x, u, s) \mapsto v(f(x, u, w^j), s - c(x, u))$ is convex in (x, u, s) for each w^j. By proceeding step-by-step through the operations that lead to v^* and by using knowledge of the operations that preserve convexity, the desired properties follow. ∎

The next supporting result for Thm. 2 provides properties of conditional expectations and a DP recursion on $\mathbb{R}^n \times \mathbb{R}$. Let $\pi \in \bar{\Pi}$ and $\gamma \in \bar{\Gamma}$. For $t = 0, 1, \dots, N$, denote the (π, γ)-*conditional expectation* of $\max(Z_t - S_t, 0)$ as follows: for all $(x_t, s_t) \in \mathbb{R}^n \times \mathbb{R}$, $W^{\pi,\gamma}_t(x_t, s_t) := E^{\pi,\gamma}\big(\max(Z_t - S_t, 0) | X_t = x_t, S_t = s_t\big)$, where Z_t is defined by (2), and recall that $Z := Z_0$.

Lemma 3 (A DP Recursion): Let $\pi \in \bar{\Pi}$ and $\gamma \in \bar{\Gamma}$. Then, for all $(x, s) \in \mathbb{R}^n \times \mathbb{R}$, we have

$$W^{\pi,\gamma}_0(x, s) = E^{\pi,\gamma}_{x,s}\big(\max(Z - S_0, 0)\big),$$
$$W^{\pi,\gamma}_N(x, s) = \max(x^T Q_f x - s, 0).$$

Also, for $t \in \{N-1, \dots, 0\}$, we have $W^{\pi,\gamma}_t(x_t, s_t) = \sum_{j=1}^{N_W} p^j_t \; W^{\pi,\gamma}_{t+1}\big(f(x_t, \pi_t(x_t, s_t), w^j), s_t - c(x_t, \pi_t(x_t, s_t))\big)$.

Proof: The conclusions follow from the same arguments that are used to prove the DP recursion for expected cumulative costs (when one uses the probability measure $P^{\pi,\gamma}_{x,s}$ and the dynamics of the augmented state). ∎

Next, we use Lemma 2 and Lemma 3 to prove Thm. 2.

Proof: [Thm. 2] The properties of $\bar{V}_t$ hold by verifying the properties of $\bar{V}_N$ and by applying Lemma 2 inductively. By Lemma 3, we have $\bar{V}^*_0(x, s) = \inf_{\pi \in \bar{\Pi}} \sup_{\gamma \in \bar{\Gamma}} W^{\pi,\gamma}_0(x, s)$ for all $(x, s) \in \mathbb{R}^n \times \mathbb{R}$. Denote $\mathbb{T} := \{0, 1, \dots, N\}$. It suffices to show that $W^{\pi^*,\gamma}_t(x_t, s_t) \le \bar{V}_t(x_t, s_t)$ for all $(x_t, s_t) \in \mathbb{R}^n \times \mathbb{R}$, $\gamma \in \bar{\Gamma}$, and $t \in \mathbb{T}$. Indeed, the above statement implies that $\sup_{\gamma \in \bar{\Gamma}} W^{\pi^*,\gamma}_t(x_t, s_t) \le \bar{V}_t(x_t, s_t)$ for all $(x_t, s_t) \in \mathbb{R}^n \times \mathbb{R}$ and $t \in \mathbb{T}$. Since $\pi^* \in \bar{\Pi}$ and by the definition of the infimum, $\bar{V}^*_0(x, s) := \inf_{\pi \in \bar{\Pi}} \sup_{\gamma \in \bar{\Gamma}} W^{\pi,\gamma}_0(x, s) \le \sup_{\gamma \in \bar{\Gamma}} W^{\pi^*,\gamma}_0(x, s) \le \bar{V}_0(x, s)$ for all $(x, s) \in \mathbb{R}^n \times \mathbb{R}$. Now, the base case ($t = N$) holds via Lemma 3 and the definition of $\bar{V}_N$. Assume (the induction hypothesis) that, for some $t \in \{N-1, \dots, 0\}$, we have

$$W^{\pi^*,\gamma}_{t+1}(x_{t+1}, s_{t+1}) \le \bar{V}_{t+1}(x_{t+1}, s_{t+1}) \tag{11}$$

for all $(x_{t+1}, s_{t+1}) \in \mathbb{R}^n \times \mathbb{R}$ and $\gamma \in \bar{\Gamma}$. Let $(x_t, s_t) \in \mathbb{R}^n \times \mathbb{R}$ and $\gamma \in \bar{\Gamma}$. It follows that $W^{\pi^*,\gamma}_t(x_t, s_t) \le \bar{V}_t(x_t, s_t)$ due to π^* being in $\bar{\Pi}$, Lemma 3, and Eqs. (11) and (10). ∎

V. A Scalable Upper Bound

Here, we return to the setting where there may be uncountably many disturbance values. We will derive a scalable upper bound to V_0 (Alg. 1) of the form, $\hat{V}_0(x, s) := a_0 + \max(x^T P_0 x - s, 0)$ for all $(x, s) \in \mathbb{R}^n \times \mathbb{R}$, where $a_0 \in \mathbb{R}$ and a positive definite symmetric matrix $P_0 \in \mathbb{R}^{n \times n}$ are obtained through a Riccati-like recursion. The recursion is parameterized by a positive definite symmetric matrix L and provides a risk-averse controller. After the proof of Thm. 3, we will describe the controller synthesis procedure.

Theorem 3: Define $P_N := Q_f$ and $a_N := 0$. Let $L \in \mathbb{R}^{n \times n}$ satisfy $L > 0$. For $t = N-1, \dots, 1, 0$, define the matrices $P_t \in \mathbb{R}^{n \times n}$, such that $P_t > 0$, and the scalars $a_t \in \mathbb{R}$ recursively,

$$P_t := A^T \left(P^{-1}_{t+1} + BR^{-1}B^T - (P_{t+1} + L)^{-1}\right)^{-1} A + Q,$$
$$a_t := a_{t+1} + \operatorname{tr}(\Sigma(P_{t+1} + L)). \tag{12}$$

For all $t \in \{N, \dots, 1, 0\}$, define $\hat{V}_t(x_t, s_t) := a_t + \max(x^T_t P_t x_t - s_t, 0)$ for all $(x_t, s_t) \in \mathbb{R}^n \times \mathbb{R}$. Then, for all $t \in \{N, \dots, 1, 0\}$, we have $V_t \le \hat{V}_t$, provided that V_t is Borel measurable and bounded below by 0.

Remark 3 (About L, P_t, a_t): P_t and a_t (12) are parameterized by L. In the finite-time case above, $L \in \mathbb{R}^{n \times n}$ is only required to be symmetric and positive definite.

Proof: We proceed by induction. The base case holds because $P_N = Q_f$ and $a_N = 0$. Now assume that for some $t \in \{N-1, \dots, 1, 0\}$, for all $(x_{t+1}, s_{t+1}) \in \mathbb{R}^n \times \mathbb{R}$, we have $V_{t+1}(x_{t+1}, s_{t+1}) \le a_{t+1} + \max(x^T_{t+1} P_{t+1} x_{t+1} - s_{t+1}, 0)$, where $P_{t+1} \in \mathbb{R}^{n \times n}$ satisfies $P_{t+1} > 0$ and a_{t+1} is a scalar. It suffices to show that $V_t(x_t, s_t) \le a_t + \max(x^T_t P_t x_t - s_t, 0)$ $\forall (x_t, s_t) \in \mathbb{R}^n \times \mathbb{R}$, where a_t and P_t are defined by (12). Let $(x_t, s_t) \in \mathbb{R}^n \times \mathbb{R}$. Since $\hat{V}_{t+1}$ and V_{t+1} are Borel measurable and $0 \le V_{t+1} \le \hat{V}_{t+1}$, it holds that $V_t(x_t, s_t) \le a_{t+1} + \inf_{u_t \in \mathbb{R}^m} \sup_{\nu_t \in \mathcal{P}_W} \int_{\mathbb{R}^n} \max(\phi^{x_t,s_t}_{t+1,u_t}(w_t), 0) \nu_t(\mathrm{d}w_t)$, where $\phi^{x_t,s_t}_{t+1,u_t}(w_t) := f(x_t, u_t, w_t)^T P_{t+1} f(x_t, u_t, w_t) + c(x_t, u_t) - s_t$. By weak duality (e.g., see [18, Lem. A.1]),

$$V_t(x_t, s_t) \le a_{t+1} + \underbrace{\inf_{u_t \in \mathbb{R}^m} \inf_{M \in \mathcal{M}^{x_t,s_t}_{t+1,u_t}} \operatorname{tr}(\Delta M)}_{\psi(x_t, s_t)}, \tag{13}$$

where $\Delta := \operatorname{diag}(\Sigma, 1)$ and $\mathcal{M}^{x_t,s_t}_{t+1,u_t}$ is the set of matrices $M = \begin{bmatrix} M_{11} & M_{12} \\ M^T_{12} & M_{22} \end{bmatrix} > 0$ s.t. $M_{11} \in \mathbb{R}^{n \times n}$, $M_{22} \in \mathbb{R}$, and

$$\begin{bmatrix} w^T_t & 1 \end{bmatrix} M \begin{bmatrix} w^T_t & 1 \end{bmatrix}^T > \phi^{x_t,s_t}_{t+1,u_t}(w_t) \quad \forall w_t \in \mathbb{R}^n. \tag{14}$$

By matrix algebra, it follows that (14) is equivalent to

$$\Phi^M_{x_t,s_t} + \bar{Q}^T u_t \bar{P} + (\bar{Q}^T u_t \bar{P})^T > 0, \quad \text{where} \tag{15}$$

$$\Phi^M_{x_t,s_t} := \begin{bmatrix} M - G_{s_t} & {K^{x_t}}^T \\ K^{x_t} & H^{-1} \end{bmatrix},\ K^{x_t} := \begin{bmatrix} I_n & 0_{n\times 1} \\ 0_{(n+m)\times n} & \begin{bmatrix} x_t \\ 0_{m\times 1} \end{bmatrix} \end{bmatrix},$$

$$G_{s_t} := \begin{bmatrix} 0_{n\times n} & 0_{n\times 1} \\ 0_{1\times n} & -s_t \end{bmatrix},\ \bar{Q}^T := \begin{bmatrix} 0_{(3n+1)\times m} \\ I_m \end{bmatrix},$$

$$\bar{P} := \begin{bmatrix} 0_{1\times n} & 1 & 0_{1\times(2n+m)} \end{bmatrix},$$

$$H := \begin{bmatrix} P_{t+1} & P_{t+1}\begin{bmatrix} A & B \end{bmatrix} \\ (*)^T & \begin{bmatrix} A & B \end{bmatrix}^T P_{t+1} \begin{bmatrix} A & B \end{bmatrix} + \mathrm{diag}(Q, R) \end{bmatrix}.$$

Here, $(*)$ denotes the appropriate terms for symmetry. By [19, Lemma 3.1], (15) is solvable for $u_t \in \mathbb{R}^m$ if and only if $W_{\bar{P}}^T \Phi^M_{x_t,s_t} W_{\bar{P}} > 0$ and $W_{\bar{Q}}^T \Phi^M_{x_t,s_t} W_{\bar{Q}} > 0$, where the columns of $W_{\bar{P}}$ and $W_{\bar{Q}}$ form bases for the nullspaces of $\bar{P}$ and $\bar{Q}$, respectively. By matrix algebra, it holds that

$$\begin{aligned} &W_{\bar{P}}^T \Phi^M_{x_t,s_t} W_{\bar{P}} > 0 \iff M_{11} > P_{t+1}, \\ &W_{\bar{Q}}^T \Phi^M_{x_t,s_t} W_{\bar{Q}} > 0 \iff M > H_{x_t,s_t}, \quad \text{where} \end{aligned} \tag{16}$$

$$H_{x_t,s_t} := \begin{bmatrix} \tilde{G} & \tilde{G}Ax_t \\ x_t^T A^T \tilde{G} & x_t^T \big(A^T \tilde{G} A + Q\big) x_t - s_t \end{bmatrix}, \text{ and}$$

$$\tilde{G} := P_{t+1} - P_{t+1}B(R + B^T P_{t+1} B)^{-1} B^T P_{t+1}. \tag{17}$$

Therefore, $\psi(x_t, s_t)$ (13) is equivalent to

$$\psi(x_t, s_t) = \inf_{M \in \mathcal{M}^{x_t,s_t}_{t+1}} \mathrm{tr}(\Delta M), \quad \text{where} \tag{18}$$

$$\mathcal{M}^{x_t,s_t}_{t+1} := \left\{ \begin{array}{l|l} M = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix} & M > 0 \\ M_{11} \in \mathbb{R}^{n\times n} & M_{11} > P_{t+1} \\ M_{22} \in \mathbb{R} & M > H_{x_t,s_t} \end{array} \right\}. \tag{19}$$

By (13), (18), and $\Delta = \mathrm{diag}(\Sigma, 1)$, it holds that

$$V_t(x_t, s_t) \le a_{t+1} + \inf_{M \in \mathcal{M}^{x_t,s_t}_{t+1}} \mathrm{tr}(\Sigma M_{11}) + M_{22}. \tag{20}$$

By taking a Schur complement, $M > H_{x_t,s_t}$ is equivalent to $M_{11} > \tilde{G}$ and $M_{22} > h(x_t, s_t, M)$, where

$$\begin{aligned} h(x_t, s_t, M) := {} & x_t^T \big(A^T \tilde{G} A + Q\big) x_t - s_t \\ & + (*)^T (M_{11} - \tilde{G})^{-1} \big(M_{12} - \tilde{G} A x_t\big). \end{aligned} \tag{21}$$

We have $\tilde{G} \le P_{t+1}$ from (17), so $M_{11} > \tilde{G}$ is redundant:

$$\mathcal{M}^{x_t,s_t}_{t+1} = \left\{ \begin{array}{l|l} M = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix} & M > 0 \\ M_{11} \in \mathbb{R}^{n\times n} & M_{11} > P_{t+1} \\ M_{22} \in \mathbb{R} & M_{22} > h(x_t, s_t, M) \end{array} \right\}. \tag{22}$$

To bound the objective, we use the relaxation $M_{12} = 0_{n\times 1}$. Recall that $L > 0$ and define the set $\hat{\mathcal{M}}^{x_t,s_t}_{t+1,L} \subseteq \mathcal{M}^{x_t,s_t}_{t+1}$ as:

$$\hat{\mathcal{M}}^{x_t,s_t}_{t+1,L} := \left\{ \begin{array}{l|l} M = \begin{bmatrix} M_{11} & 0_{n\times 1} \\ 0_{1\times n} & M_{22} \end{bmatrix} & M > 0 \\ M_{11} \in \mathbb{R}^{n\times n} & M_{11} > P_{t+1} + L \\ M_{22} \in \mathbb{R} & M_{22} > \hat{h}(x_t, s_t, M_{11}) \end{array} \right\}, \tag{23}$$

where we define $\hat{h}(x_t, s_t, M_{11}) :=$

$$x_t^T \big(A^T \big(\tilde{G}^{-1} - M_{11}^{-1}\big)^{-1} A + Q\big) x_t - s_t. \tag{24}$$

Thus, we have

$$V_t(x_t, s_t) \le a_{t+1} + \underbrace{\inf_{M \in \hat{\mathcal{M}}^{x_t,s_t}_{t+1,L}} \mathrm{tr}(\Sigma M_{11}) + M_{22}}_{\phi_L(x_t, s_t)}. \tag{25}$$

Let $M^*_{11} := P_{t+1} + L$, $M^*_{22} := \max\big(\hat{h}(x_t, s_t, M^*_{11}), 0\big)$, and $M^*_{x_t,s_t} := \mathrm{diag}(M^*_{11}, M^*_{22})$. Then,

$$\phi_L(x_t, s_t) \le \mathrm{tr}(\Sigma M^*_{11}) + \max\big(\hat{h}(x_t, s_t, M^*_{11}), 0\big). \tag{26}$$

By substituting the definition of M^*_{11}, we have

$$V_t(x_t, s_t) \le a_t + \max\big(\hat{h}(x_t, s_t, P_{t+1} + L), 0\big), \tag{27}$$

where a_t is given by (12). Since $\hat{h}(x_t, s_t, P_{t+1} + L) = x_t^T P_t x_t - s_t$, where P_t is given by (12), we are done. ∎

Remark 4 (Controller Synthesis): Based on the proof of Thm. 3, we can derive a sub-optimal policy as follows. For a fixed $L > 0$, compute the matrices P_t via the recursion (12). Let $x_0 \in \mathbb{R}^n$ be an initial condition. Define $s_0 := x_0^T P_0 x_0$, which depends on L through P_0. For $t = 0, 1, \ldots, N-1$, proceed through the following steps:

1) Compute $M^*_{x_t,s_t}$ as per the proof of Thm. 3, $M^*_{x_t,s_t} := \mathrm{diag}(M^*_{11}, M^*_{22})$, where $M^*_{11} := P_{t+1} + L$, $M^*_{22} := \max(\hat{h}(x_t, s_t, M^*_{11}), 0)$, and $\hat{h}$ is given by (24).
2) Choose a $u_t \in \mathbb{R}^m$ that satisfies (15) when $M = M^*_{x_t,s_t}$. Such a u_t is guaranteed to exist from the choice of $M = M^*_{x_t,s_t}$ and by repeating several steps in the proof above. We note that the u_t satisfying (15) may not be unique.
3) Nature chooses a disturbance value $w_t \in \mathbb{R}^n$.
4) Calculate $x_{t+1} = Ax_t + Bu_t + w_t$ and $s_{t+1} = s_t - c(x_t, u_t)$. Update t by 1. Go to step 1 if $t < N$.

We now identify some interesting similarities and differences between our approach and classical methods.

Remark 5 (Relation to LEQR and LQ games): The Riccati recursion for the LEQR problem in finite time takes the form [7]: for $t = N-1, \ldots, 1, 0$,

$$\bar{P}_t = A^T \big(\bar{P}_{t+1}^{-1} + BR^{-1}B^T - \gamma\Sigma\big)^{-1} A + Q, \tag{28}$$

provided that $\gamma > 0$ is chosen so that $\Sigma^{-1} - \gamma \bar{P}_{t+1}$ is positive definite for each t. Similarly, the Riccati recursion for a soft-constrained LQ game takes the form [1, Eq. 3.4a', p. 53]: for $t = N-1, \ldots, 1, 0$,

$$\hat{P}_t = A^T \big(\hat{P}_{t+1}^{-1} + BR^{-1}B^T - \tfrac{1}{\lambda^2}\Sigma\big)^{-1} A + Q, \tag{29}$$

provided that $\hat{P}_t$ is invertible for each t, $R = I_m$, and λ is a scalar parameter representing a disturbance-attenuation level. The key differences between (12), (28), and (29) appear in the terms $\gamma\Sigma$, $\frac{1}{\lambda^2}\Sigma$, and $(P_{t+1}+L)^{-1}$, respectively. Our recursion (12) encodes a risk-aversion level through the matrix $(P_{t+1} + L)^{-1}$, whereas the classical recursions (28) (29) encode risk aversion by scaling the covariance Σ.

Remark 6 (Relation to minimax MPC): One may interpret an LEQR controller in a model-predictive-control (MPC) setting as an approximate solution to minimax MPC [3, p. 99]. In minimax MPC, a matrix $\mathcal{T} \ge 0$, which depends on a bounded region containing the process noise, appears in the algorithm that provides an optimal control [3, Eq. 8.29,

p. 99]. Our recursion (12) has a similar structure since it is parameterized by a matrix $L > 0$, and it is plausible that a preferable choice of L depends on the maximal covariance Σ (a topic for future investigation). A key distinction between minimax MPC and our approach is the uncertainty model of the process noise. Our approach permits process noise with an unbounded support and a spectrum of possibilities that occur with various probabilities. However, minimax MPC permits process noise that lives in a bounded region with known bounds [3, p. 42]. The "better" uncertainty model may be application-dependent.

VI. Numerical Simulation

Fig. 1 provides example trade-off curves comparing LEQR (as γ varies) with our proposed approach from Section V (as L varies). These results show that for a simple one-state system, our proposed approach (ACVaR) has comparable performance relative to LEQR. This finding is notable given the simplicity of our experiment and that our method avoids the case where γ is too large and the LEQR cost becomes infinite. We also simulated the optimal CVaR controller from [16], which is not distributionally robust. This controller assumes exact prior knowledge of the disturbance distribution, which explains its superior performance. However, this optimal CVaR controller is not scalable to higher-dimensional problem instances, since it requires discretizing the augmented state space.

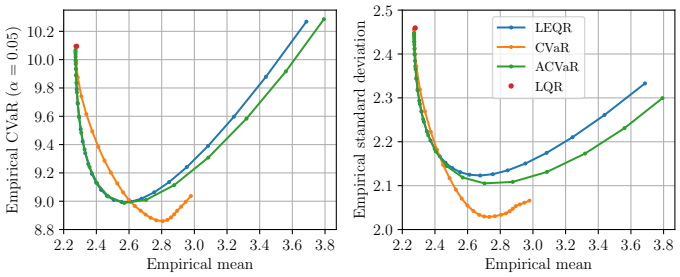

Fig. 1. Trade-offs between empirical mean, standard deviation, and $\text{CVaR}_{0.05}$ of the LQR cost for (i) our controller (**ACVaR**) as L varies, (ii) the LEQR controller as γ varies (**LEQR**), (iii) the exact CVaR_α controller with prior knowledge of the disturbance distribution as α varies (**CVaR**), and (iv) the LQR controller (**LQR**). We used the scalar dynamical system $x_{t+1} = x_t + u_t + w_t$ with $R = Q_f = 1$, $Q = 10^{-3}$, $x_0 = 1$, and $N = 4$. The disturbance w_t is zero-mean Gaussian with unit variance. The parameter ranges were $0.2 \leq L \leq 100$, $\frac{\gamma_c}{10} \leq \gamma \leq \gamma_c$, where γ_c is the critical γ value for LEQR. Each point is the mean of 50,000 trials, where the same schedule of pseudo-random seeds are used across policies. In the limits $L \to \infty$, $\gamma \to 0$, and $\alpha \to 1$ for ACVaR, LEQR, and CVAR_α, respectively, we recover the risk-neutral LQR policy.

VII. Concluding Remarks

We took steps toward deriving a scalable upper bound to a distributionally robust, CVaR optimal control problem for linear systems with quadratic costs. CVaR characterizes the (usually abstract) notion of risk as a fraction of worst-case outcomes, which is intuitive and precise. A result from our analysis is a risk-averse controller with intriguing similarities and differences relative to the state-of-the-art.

Potential areas for future work include studying the infinite-horizon case, characterizing the extent to which the upper bound approximation parameterized by L is tight, and elucidating the connections between the choice of L and the maximal covariance Σ.

Further numerical experiments, potentially with higher-dimensional or more realistic application-specific examples, are needed to ascertain whether the proposed approach may be a superior alternative to LEQR in certain application domains.

Acknowledgment

Both authors would like to thank the reviewers for their valuable feedback regarding connections to existing literature in robust and risk-sensitive control. M.P. Chapman acknowledges support from the University of Toronto. L. Lessard is partially supported by NSF awards 1710892 and 1750162.

References

[1] T. Başar and P. Bernhard, *H^∞-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*, 2nd ed. Birkhauser, 1995.

[2] K. Zhang, B. Hu, and T. Basar, "On the stability and convergence of robust adversarial reinforcement learning: A case study on linear quadratic systems," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[3] J. Löfberg, *Minimax approaches to robust model predictive control*. Linköping University Electronic Press, 2003, vol. 812.

[4] K. Zhang, B. Hu, and T. Basar, "Policy optimization for $\mathcal{H}_2$ linear control with $\mathcal{H}_\infty$ robustness guarantee: Implicit regularization and global convergence," in *Learning for Dynamics and Control*. PMLR, 2020, pp. 179–190.

[5] R. A. Howard and J. E. Matheson, "Risk-sensitive markov decision processes," *Management Science*, vol. 18, no. 7, pp. 356–369, 1972.

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

[7] P. Whittle, "Risk-sensitive linear/quadratic/gaussian control," *Advances in Applied Probability*, pp. 764–777, 1981.

[8] ——, *Risk-sensitive Optimal Control*. Wiley, 1990, vol. 2.

[9] G. B. Di Masi and L. Stettner, "Risk-sensitive control of discrete-time markov processes with infinite horizon," *SIAM Journal on Control and Optimization*, vol. 38, no. 1, pp. 61–78, 1999.

[10] N. Bäuerle and U. Rieder, "More risk-sensitive markov decision processes," *Mathematics of Operations Research*, vol. 39, no. 1, pp. 105–120, 2014.

[11] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *Journal of Banking & Finance*, vol. 26, no. 7, pp. 1443–1471, 2002.

[12] J. Kisiala, "Conditional value-at-risk: Theory and applications," Master's thesis, School of Mathematics, University of Edinburgh, Aug. 2015.

[13] S. Samuelson and I. Yang, "Safety-aware optimal control of stochastic systems using conditional value-at-risk," in *American Control Conference*, 2018, pp. 6285–6290.

[14] M. P. Chapman, J. Lacotte, A. Tamar, D. Lee, K. M. Smith, V. Cheng, J. F. Fisac, S. Jha, M. Pavone, and C. J. Tomlin, "A risk-sensitive finite-time reachability approach for safety of stochastic dynamic systems," in *American Control Conference*, 2019, pp. 2958–2963.

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

[16] N. Bäuerle and J. Ott, "Markov decision processes with Average-Value-at-Risk criteria," *Mathematical Methods of Operations Research*, vol. 74, no. 3, pp. 361–379, 2011.

[17] O. Hernández-Lerma and J. B. Lasserre, *Discrete-time Markov control processes: Basic optimality criteria*. Springer, 1996, vol. 30.

[18] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Mathematical Programming*, vol. 137, pp. 167–198, 2013.

[19] P. Gahinet and P. Apkarian, "A linear matrix inequality approach to H∞ control," *International Journal of Robust and Nonlinear Control*, vol. 4, no. 4, pp. 421–448, 1994.