Risk-Sensitive Safety Analysis Using Conditional Value-at-Risk

Margaret P. Chapman, *Member, IEEE*, Riccardo Bonalli, Kevin M. Smith, Insoon Yang, *Member, IEEE*, Marco Pavone, *Member, IEEE*, and Claire J. Tomlin, *Fellow, IEEE*

***Abstract*—This article develops a safety analysis method for stochastic systems that is sensitive to the possibility and severity of rare harmful outcomes. We define *risk-sensitive safe sets* as sublevel sets of the solution to a nonstandard optimal control problem, where a random maximum cost is assessed via Conditional Value-at-Risk (CVaR). The objective function represents the maximum extent of constraint violation of the state trajectory, averaged over a given percentage of worst cases. This problem is well-motivated but difficult to solve tractably because the temporal decomposition for CVaR is history-dependent. Our primary theoretical contribution is to derive computationally tractable underapproximations to risk-sensitive safe sets. Our method provides a novel, theoretically guaranteed, parameter-dependent upper bound to the CVaR of a maximum cost without the need to augment the state space. For a fixed parameter value, the solution to only one Markov decision process problem is required to obtain the underapproximations for any family of risk-sensitivity levels. In addition, we propose a second definition for risk-sensitive safe sets and provide a tractable method for their estimation without using a parameter-dependent upper bound. The second definition is expressed in terms of a new coherent risk functional, which is inspired by CVaR. We demonstrate our primary theoretical contribution via numerical examples.**

***Index Terms*—Conditional Value-at-Risk (CVaR), Markov decision processes (MDPs), safety analysis, stochastic optimal control.**

I. Introduction

Control-theoretic formal verification methods for dynamical systems typically fall in the robust domain [20]–[24] or in the stochastic domain [25]–[28]. Robust methods for formal verification assume that uncertain disturbances lack probabilistic descriptions, live in bounded sets, and exhibit adversarial behavior. These assumptions are appropriate if probabilistic information about disturbances is not available, and if the conservative policy or safety specification that results from a pessimistic world view is useful in practice. However, when one considers formal verification as a design tool for safety-critical systems in the digital world today, it is reasonable to assume that simulation tools or sensor data are available to estimate probabilistic descriptions for disturbances. Moreover, it is reasonable to consider the following world view: disturbances need not be adversarial, but rare harmful outcomes are still possible.

Control-theoretic stochastic formal verification methods do assume that disturbances are probabilistic and can be non-adversarial [25], [26] or adversarial [27], [28] in nature. These methods compute the probability of safety or performance by using expected indicator cost functions. The expectation, however, is not designed to quantify the features in the tails of a distribution, and the probability of a harmful outcome need not indicate its severity. Thus, formal verification methods at the intersection of the robust and stochastic domains are emerging. A method for distributionally robust safety analysis has been proposed [29], and methods that use risk measures to assess

Manuscript received 8 July 2021; revised 11 October 2021; accepted 14 November 2021. Date of publication 26 November 2021; date of current version 5 December 2022. This work was supported in part by the National Science Foundation (NSF) under Grant NSF PIRE UNIV59732, in part by the National Cancer Institute (NCI) Cancer Systems Biology Consortium "Measuring, Modeling, and Controlling Heterogeneity" through Oregon Health and Science University under NCI Grant 1U54CA209988-01A1, and in part by the Computational Hydraulics International University Grant Program for complementary use of PCSWMM Professional software. The work of M. P. Chapman was supported in part by the NSF Graduate Research Fellowship Program. The work of K. M. Smith was supported in part by the US National Science Foundation under Grant NSF-NRT 2021874. The work of I. Yang was supported in part by the National Research Foundation of Korea funded by MSIT under Grant 2020R1C1C1009766. This paper was presented in part at the 2019 American Control Conference, Institute of Electrical and Electronics Engineers Inc., Philadelphia, PA, July 2019 [DOI: 10.23919/ACC.2019.8815169]. Recommended by Associate Editor R. Jain. *(Corresponding author: Margaret P. Chapman.)*

Margaret P. Chapman is with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto ON M5S 3G4, Canada (e-mail: mchapman@ece.utoronto.ca).

Riccardo Bonalli is with the Laboratory of Signals and Systems (L2S), Université Paris-Saclay, Centre National de la Recherche Scientifique (CNRS), CentraleSupélec, France (e-mail: riccardo.bonalli@centralesupelec.fr).

Kevin M. Smith is with the Department of Civil and Environmental Engineering, Tufts University, Medford, MA 02155 USA, and also with OptiRTC, Inc., Boston, MA 02116 USA (e-mail: kms2227@columbia.edu).

Insoon Yang is with the Department of Electrical and Computer Engineering, Automation and Systems Research Institute, Seoul National University, Seoul, South Korea (e-mail: insoonyang@snu.ac.kr).

Marco Pavone is with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305 USA (e-mail: pavone@stanford.edu).

Claire J. Tomlin is with the Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, CA 94720 USA (e-mail: tomlin@eecs.berkeley.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2021.3131149.

Digital Object Identifier 10.1109/TAC.2021.3131149

harmful tail costs, e.g., [30] and our prior work [38], have been introduced.[1]

While the notion of risk-sensitive formal verification is recent, it is related to the notion of risk-sensitive Markov decision processes (MDPs), which dates back to the early 1970s. In 1972, Howard and Matheson studied risk-sensitive MDPs on finite state spaces, where the cost is evaluated in terms of exponential utility [1]. This idea was transferred to linear control systems by Jacobson in 1973 [2] and was further developed in later decades. For example, see the seminal works by Whittle [4], [35] and di Masi and Stettner [3]. The exponential utility of a nonnegative random cost Y $\mathcal{J}_\theta(Y) := \frac{-2}{\theta}\log(E(e^{\frac{-\theta}{2}Y}))$ assesses the risk of Y in terms of the moments of Y and is parametrized by a nonzero scalar θ. Under appropriate conditions, $\mathcal{J}_\theta(Y)$ tends to $E(Y)$ as $\theta \to 0$ and $\mathcal{J}_\theta(Y) \approx E(Y) - \frac{\theta}{4}\text{Variance}(Y)$ if $|\theta|$ is sufficiently small [4]. The risk-averse setting corresponds to $\theta < 0$. However, if θ is too negative, the controller can suffer from a phenomenon called "neurotic breakdown" in the linear-quadratic-Gaussian setting [4].

Hence, the notion of risk-sensitive MDPs has been generalized beyond exponential utility. Kreps used the expectation of a utility function as a risk-sensitive performance criterion for MDPs [7]. Ruszczyński defined a risk-sensitive performance criterion for MDPs in terms of a composition of risk measures [47]. State-space augmentation has been used to optimize the cumulative cost of an MDP, where the cost is assessed via CVaR [16] or a certainty equivalent risk measure [17]. The former problem is called a CVaR-MDP. Convex analytic methods have been used to solve MDPs with expected utility or CVaR criteria via state-space augmentation and infinite-dimensional linear programming [34]. A temporal decomposition for CVaR [40], [41] has been used to propose a dynamic programming (DP) algorithm on an augmented state space to solve a CVaR-MDP problem approximately [31]. Analysis at the intersection of mean field games, linear systems, and risk measures with connections to CVaR is provided by [32].

Ruszczyński's approach [47] and MDPs that assess cumulative costs via expectation or exponential utility are *time-consistent* problems. That is, these problems satisfy Bellman's principle of optimality on the original state space.[2] However, a CVaR-MDP is time-inconsistent. Several solution concepts for time-inconsistent problems have been proposed. For example, a game-theoretic solution concept is studied in [8], which considers the problem as a game against one's future self. Another popular approach is to focus on *precommitment strategies* that cannot be revised at later stages. Optimal or nearly optimal precommitment strategies can be obtained using the structure of CVaR; see [16], [34], and [39], for example. Although an optimal precommitment strategy is globally optimal only at the initial stage, maintaining suitable empirical performance at later stages is possible, particularly when the time horizon is not too long [18]. In mean-CVaR asset allocation problems, optimal precommitment strategies are shown to be effective even with long time horizons [19].

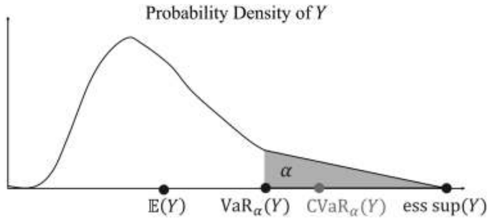

Fig. 1. CVaR quantifies the upper tail of a cost distribution. For an absolutely continuous, bounded random variable Y representing a cost and $\alpha \in (0, 1]$, we illustrate the expected cost in the $\alpha \cdot 100\%$ worst cases, which is $\text{CVaR}_\alpha(Y)$ in this setting. The area of the shaded region is α. The expectation of Y, the Value-at-Risk of Y at level α (the lowest cost in the $\alpha \cdot 100\%$ worst cases), and the essential supremum of Y are also shown.

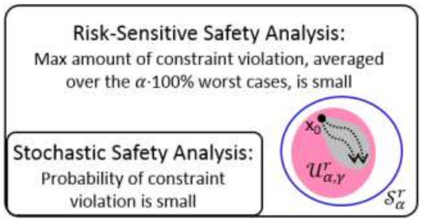
Risk-Sensitive Safety Analysis:
Max amount of constraint violation, averaged over the α·100% worst cases, is small
Stochastic Safety Analysis:
Probability of constraint violation is small
x_0
$\mathcal{U}^r_{\alpha,\gamma}$
$\mathcal{S}^r_\alpha$

Fig. 2. We develop a safety analysis method that generalizes stochastic safety analysis by assessing the severity of random harmful outcomes. We define the risk-sensitive safe set $\mathcal{S}^r_\alpha$ in terms of CVaR and derive an underapproximation $\mathcal{U}^r_{\alpha,\gamma}$ that is computationally tractable. $\mathcal{S}^r_\alpha$ represents the set of initial states from which the maximum extent of constraint violation of the state trajectory, averaged over the $\alpha \cdot 100\%$ worst cases, can be reduced to a threshold r.

A line of research that falls between risk-sensitive MDPs and standard risk-neutral MDPs is risk-constrained MDPs [30], [33], [34], [42]. Here, the goal is to minimize an expected cumulative cost subject to a risk constraint that limits the extent of a cost. The authors in [33], [42], and [30], for example, express this constraint in terms of CVaR.

The additional effort required to solve time-inconsistent problems, including CVaR-MDPs, may be justified for safety-critical applications. A strong theoretical basis for using CVaR to assess harmful tail costs has been in development since the early 2000s, e.g., see [36] and the references therein. Informally, CVaR represents the expected cost in the $\alpha \cdot 100\%$ worst cases, where $\alpha \in (0, 1]$ (Fig. 1). CVaR quantifies the more harmful tail of a distribution, and managing this tail is paramount in safety-critical applications.

This article proposes a method to assess how well a stochastic system can remain within a desired operating region with respect to a range of worst-case perspectives. We call this method *risk-sensitive safety analysis* (Fig. 2). Its foundation is a nonstandard optimal control problem that evaluates a random maximum cost via CVaR. The objective function represents the maximum extent of constraint violation of the state trajectory, averaged over the $\alpha \cdot 100\%$ worst cases, where $\alpha \in (0, 1]$ is a *risk-sensitivity level*. This problem is difficult to solve tractably because the temporal decomposition for CVaR is history-dependent [40],

[1] A *risk measure* (risk functional) is a map from a set of random variables to the extended real line. Exponential utility, Value-at-Risk, CVaR, and Mean-Deviation are examples [37]. The terms risk measure and risk functional are interchangeable.

[2] Different meanings for time consistency have been proposed; e.g., see [5], [47], [6]. We refer to the meaning for time consistency from [6].

[41]. We define *risk-sensitive safe sets* as sublevel sets of the solution to this nonstandard problem. These sets are powerful tools for safety analysis. Indeed, they assess system behavior on a spectrum of worst cases, while being sensitive to the possibility and severity of rare harmful outcomes.

Our primary theoretical contribution is to derive computationally tractable underapproximations to risk-sensitive safe sets. We derive these underapproximations by proving the following: For any control policy and any initial state, the CVaR of a maximum cost is upper bounded by a scaled logarithm of an expected cumulative cost, where the stage cost has a specific analytical form. For this proof, we use various properties of CVaR and the log-sum-exponential approximation to the maximum. The latter approximation depends on a parameter $\gamma \in \mathbb{R}$. For a fixed γ, the solution to one MDP problem is required to obtain the underapproximations for any family of risk-sensitivity levels. We provide practical insights on how to choose such a parameter in the experimental section.

Our method provides a novel, theoretically guaranteed upper bound to the CVaR of a *maximum* cost for the purpose of *safety analysis* without the need to augment the state space. (Augmenting the state space may be less tractable in some settings, e.g., when the range of the augmented state is large.) In contrast, existing methods aim to compute the CVaR of a cumulative cost via state-space augmentation. By taking different approaches to augment the state space, the papers [16] and [34] minimize the CVaR of a cumulative cost, and the paper [31] minimizes the CVaR of a cumulative cost approximately. These related works are focused on controller synthesis but are not focused on safety analysis.

Our secondary theoretical contribution is to propose a second definition for risk-sensitive safe sets and provide a tractable method for their estimation without using a parameter-dependent upper bound. The second definition is expressed in terms of a new risk functional, which is inspired by CVaR and has certain desirable properties. In particular, this risk functional admits an upper bound that can be computed via DP (on the original state space and without an additional parameter that requires tuning). This result forges a new path to estimate risk-sensitive safety criteria with desirable computational attributes.

Organization: We present notation and background on CVaR in Section II. Our primary and secondary theoretical contributions are provided in Sections III and IV, respectively. We develop computational examples of a temperature system and a stormwater system to demonstrate our primary theoretical contribution in Section V. Finally, Section VI presents conclusions and directions for future work.

II. Background on Conditional Value-at-Risk

We use the following notation. If S is a metrizable space, $\mathcal{B}(S)$ is the Borel sigma algebra on S. If $(\Omega, \mathcal{F}, \mu)$ is a probability space and $1 \leq p \leq \infty$, $L^p(\Omega, \mathcal{F}, \mu)$ is the associated L^p space, and $|| \cdot ||_p$ is the associated norm. Typically, we use upper-case letters to denote random variables or sets, whereas lower-case letters denote deterministic quantities, including parameters. Exceptions are the length of a time horizon is expressed in terms of $T \in \mathbb{N}$ and $E(\cdot)$ denotes expectation.

Next, we present a standard definition for CVaR and facts about CVaR that are relevant to this work.[3] Let Y be a random variable with finite first moment, representing a cost, defined on a probability space $(\Omega, \mathcal{F}, \mu)$. That is, let $Y \in L^1(\Omega, \mathcal{F}, \mu)$, where smaller realizations of Y are preferred. The *Conditional Value-at-Risk* of $Y \in L^1(\Omega, \mathcal{F}, \mu)$ at the *risk-sensitivity level* $\alpha \in (0, 1]$ is defined by

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

where $E(\cdot)$ is the expectation with respect to (w.r.t.) μ. We note the following consequences of Definition (1):

1) $\text{CVaR}_1(Y) = E(Y)$.
2) If $0 < \alpha_1 \leq \alpha_2 \leq 1$, then $\text{CVaR}_{\alpha_1}(Y) \geq \text{CVaR}_{\alpha_2}(Y)$ and $\text{CVaR}_{\alpha_i}(Y) \in \mathbb{R}$ for $i = 1, 2$.

Definition (1) is not the most intuitive, so we present an alternative definition that explains the names Conditional Value-at-Risk and Average Value-at-Risk. The alternative definition is written in terms of the Value-at-Risk of $Y \in L^1(\Omega, \mathcal{F}, \mu)$ at level $\alpha \in (0, 1)$, which is given by

$$\text{VaR}_\alpha(Y) := \inf \{ y \in \mathbb{R} : \mu(\{Y \leq y\}) \geq 1 - \alpha \} \tag{2}$$

where $\mu(\{Y \leq y\})$ is the probability of the event $\{Y \leq y\} := \{\omega \in \Omega : Y(\omega) \leq y\} \in \mathcal{F}$. In other words, $\text{VaR}_\alpha(Y)$ is the generalized inverse cumulative distribution function of Y at level $1 - \alpha$, or equivalently, the left-side $(1 - \alpha)$-quantile of the distribution of Y [10]. The CVaR of $Y \in L^1(\Omega, \mathcal{F}, \mu)$ at level $\alpha \in (0, 1)$ is equivalent to an average of the Value-at-Risk [37, Th. 6.2]:

$$\text{CVaR}_\alpha(Y) = \frac{1}{\alpha} \int_{1-\alpha}^{1} \text{VaR}_{1-p}(Y) \; \mathrm{d}p. \tag{3}$$

The above equation explains the commonly used name *Average Value-at-Risk*. Now, to explain the name *Conditional Value-at-Risk*, suppose that the cumulative distribution function $F_Y(y) := \mu(\{Y \leq y\})$ is continuous at $y = \text{VaR}_\alpha(Y)$. Continue to assume that $Y \in L^1(\Omega, \mathcal{F}, \mu)$ and $\alpha \in (0, 1)$. Then, $\text{CVaR}_\alpha(Y)$ is a *conditional* expectation that is expressed in terms of the *Value-at-Risk* [37, Th. 6.2]:

$$\text{CVaR}_\alpha(Y) = E(Y \,|\, Y \geq \text{VaR}_\alpha(Y)). \tag{4}$$

Equation (4) means that $\text{CVaR}_\alpha(Y)$ represents the expected value of Y in the $\alpha \cdot 100\%$ worst cases.

CVaR is a commonly cited example of a *coherent* risk functional [10], [37]. Coherent risk functionals are a class of risk functionals, first proposed by Artzner *et al.* [11], that satisfy four properties, which are particularly meaningful in applications where sensitivity to risk is critical. We present these properties in the context of CVaR at level $\alpha \in (0, 1]$, where $Y_i \in L^1(\Omega, \mathcal{F}, \mu)$ below.

1) *Monotonicity:* If $Y_1(\omega) \leq Y_2(\omega)$ for almost every (a.e.) $\omega \in \Omega$, then $\text{CVaR}_\alpha(Y_1) \leq \text{CVaR}_\alpha(Y_2)$. That is, a random cost that is larger than another almost everywhere incurs a larger risk.
2) *Subadditivity:* $\text{CVaR}_\alpha(Y_1 + Y_2) \leq \text{CVaR}_\alpha(Y_1) + \text{CVaR}_\alpha(Y_2)$. If Y_i is the (random) stage cost of a

[3] Additional names for CVaR include Average Value-at-Risk, Expected Shortfall, and Expected Tail Loss. We present the definition for CVaR that is used by Shapiro and colleagues, e.g., [9], [10], [37].

control system at time i, then the risk of the cumulative cost over a finite horizon is at most the sum of the risks of the stage costs.

3) *Translation equivariance:* If $a \in \mathbb{R}$, then $\text{CVaR}_\alpha(Y_1 + a) = \text{CVaR}_\alpha(Y_1) + a$.
4) *Positive homogeneity:* If $0 \leq \lambda < \infty$, then $\text{CVaR}_\alpha(\lambda Y_1) = \lambda \text{CVaR}_\alpha(Y_1)$.

The last two properties ensure that shifting or scaling a random variable provides an analogous transformation to the risk of the random variable. In particular, the expectation operator satisfies the four properties above and thus is a coherent risk functional. We use some of these properties in our proofs. We also use the fact that a real-valued coherent risk functional can be represented in terms of a supremum over a family of expectations.[4] This representation takes the following form for CVaR at level $\alpha \in (0, 1]$ [10]: for any $Y \in L^1(\Omega, \mathcal{F}, \mu)$,

$$\text{CVaR}_\alpha(Y) = \sup_{Q \in \mathcal{Q}_\alpha} \int_\Omega Y \, \mathrm{d}Q = \sup_{\xi \in \mathcal{A}_\alpha} \int_\Omega Y \xi \, \mathrm{d}\mu \tag{5a}$$

where the definitions of $\mathcal{Q}_\alpha$ and $\mathcal{A}_\alpha$ follow. $Q \in \mathcal{Q}_\alpha$ if and only if Q is a probability measure that is absolutely continuous with respect to μ, i.e., of the form $Q(B) = \int_B \xi \mathrm{d}\mu$, where $B \in \mathcal{F}$ and $\xi \in \mathcal{A}_\alpha$. $\mathcal{A}_\alpha$ is a set of densities defined by

$$\mathcal{A}_\alpha := \left\{ \xi \in L^\infty(\Omega, \mathcal{F}, \mu) : 0 \leq \xi \leq \frac{1}{\alpha} \text{ a.e.}, \int_\Omega \xi \, \mathrm{d}\mu = 1 \right\}. \tag{5b}$$

III. CVaR-Based Risk-Sensitive Safety Analysis

We use the CVaR functional to pose a safety analysis problem. We consider a stochastic system evolving on a discrete, finite-time horizon and start with the standard set-up for this setting. Let S and A be Borel spaces, representing the set of states and the set of controls of the system, respectively. Define the sample space $\Omega := (S \times A)^T \times S$, where $\omega := (x_0, u_0, \ldots, x_{T-1}, u_{T-1}, x_T) \in \Omega$ is a finite sequence of states and controls that may be realized on a time horizon of length $T + 1$ and $T \in \mathbb{N}$ is given. The random state $X_t : \Omega \to S$ and the random control $U_t : \Omega \to A$ are projections. That is, for any $\omega \in \Omega$ of the form above, define $X_t(\omega) := x_t$ and $U_t(\omega) := u_t$, where the coordinates of ω have casual dependencies, to be described. The initial state X_0 is fixed arbitrarily at $x \in S$. The system's evolution is affected by W-valued random disturbances $(D_0, D_1, \ldots, D_{T-1})$ with a common distribution P_D, where W is a Borel space. D_t is independent of the states, controls, and D_s for any $s \neq t$. The distribution of X_{t+1} conditioned on $(X_t, U_t) = (x_t, u_t) \in S \times A$ is defined as follows: for any $B \in \mathcal{B}(S)$,

$$Q(B|x_t, u_t) := P_D\left(\{d_t \in W : f(x_t, u_t, d_t) \in B\}\right) \tag{6}$$

where $f : S \times A \times W \to S$ is a Borel-measurable map that models the system dynamics. We use the typical class of random, history-dependent policies Π. Each $\pi \in \Pi$ takes the form $\pi = (\pi_0, \pi_1, \ldots, \pi_{T-1})$, where each π_t is a Borel-measurable stochastic kernel on A given $H^t := (S \times A)^t \times S$.

The above set-up is standard in discrete-time stochastic control. One reason is that, given $x \in S$ and $\pi \in \Pi$, the set-up allows the construction of a unique probability measure P_x^π that characterizes the system's evolution, provided that the system is initialized at x and uses the policy π (Ionescu-Tulcea Theorem). The measure P_x^π permits the prediction of the system's performance over time under uncertainty. Random costs incurred by the system are defined on $(\Omega, \mathcal{B}(\Omega), P_x^\pi)$, a probability space parametrized by x and π. The notation $E_x^\pi(\cdot)$ is the expectation operator with respect to P_x^π.

A. On Evaluating a Random Cost via CVaR

We use $(\Omega, \mathcal{B}(\Omega), P_x^\pi)$ to define a random cost for the system and to evaluate this cost via CVaR. Suppose that there is a *constraint set* $K \in \mathcal{B}(S)$, where the state trajectory $(X_0, X_1, \ldots, X_T)$ of the system should remain inside. It may be impossible for the system to remain inside K always due to random disturbances in the environment. Let $g_K : S \to \mathbb{R}$ be a bounded Borel-measurable function that represents a notion of distance between a state realization and the boundary of K. Specifically, $g_K(x_t)$ is the *extent of constraint violation* of x_t, a realization of the random state X_t. More specifically, if x_t is outside of K and far from the boundary of K, then $g_K(x_t)$ has a large positive value. However, if x_t is inside of K, then $g_K(x_t)$ may be either of the following:

1) zero, if one does not favor certain trajectories inside of K;
2) a more negative value when x_t is more deeply inside of K, if one favors trajectories that remain deeply inside of K.

Using g_K, we define a random $\mathbb{R}$-valued cost that quantifies the *maximum extent of constraint violation of the state trajectory:* for any $\omega = (x_0, u_0, \ldots, x_{T-1}, u_{T-1}, x_T) \in \Omega$,

$$G(\omega) := \max_{t=0,1,\ldots,T} g_K(X_t(\omega)) = \max_{t=0,1,\ldots,T} g_K(x_t). \tag{7}$$

In other words, G quantifies how well the random state trajectory satisfies the safety criterion to remain inside of K. Hence, G quantifies the safety of the random state trajectory, which is defined with respect to the constraint set K via the function g_K. A deterministic (and continuous-time) version of (7) is used in Hamilton-Jacobi reachability analysis, a robust safety analysis method for (nonstochastic) uncertain systems, which has been established over the past 15 years; e.g., see [21], [43], and [24], and the references therein. A standard choice for g_K is a clipped signed distance function with respect to K [43, p. 8]. In our numerical example of a thermostatically controlled load, we use $g_K(x_t) = \max(x_t - 21, 20 - x_t)$ to quantify how far a state realization x_t can be inside or outside of $K = [20, 21]$ °C (Section V-A).

It holds that $G \in L^\infty := L^\infty(\Omega, \mathcal{B}(\Omega), P_x^\pi)$ and $G \in L^1 := L^1(\Omega, \mathcal{B}(\Omega), P_x^\pi)$. The function g_K composed with X_t is an element of L^∞ because $g_K : S \to \mathbb{R}$ is bounded and Borel measurable and $X_t : \Omega \to S$ is Borel measurable. Thus, G is a point-wise maximum of finitely many functions in L^∞. Therefore, G inherits the measurability properties of these functions

[4] The family of expectations has specific properties that are out of the scope of this article. The representation was developed over several years, e.g., see [10], [11], [13], [14].

and is essentially bounded. Since $(\Omega, \mathcal{B}(\Omega), P_x^\pi)$ is a probability space, L^∞ is a subset of L^1, and it follows that $G \in L^1$ as well.

Now, we express the CVaR of G. The CVaR of $G \in L^1(\Omega, \mathcal{B}(\Omega), P_x^\pi)$ at level $\alpha \in (0, 1]$ is given by

$$\text{CVaR}_{\alpha,x}^\pi(G) := \inf_{s\in\mathbb{R}} \left(s + \tfrac{1}{\alpha} E_x^\pi(\max(G - s, 0))\right). \tag{8a}$$

By using (5), it holds that

$$\text{CVaR}_{\alpha,x}^\pi(G) = \sup_{\xi\in\mathcal{A}_{\alpha,x}^\pi} \int_\Omega G\xi \, \mathrm{d}P_x^\pi \tag{8b}$$

where $\mathcal{A}_{\alpha,x}^\pi$ is a set of densities defined by

$$\mathcal{A}_{\alpha,x}^\pi := \left\{\xi \in L^\infty(\Omega, \mathcal{B}(\Omega), P_x^\pi) : 0 \leq \xi \leq \tfrac{1}{\alpha} \text{ a.e.}, E_x^\pi(\xi) = 1\right\}. \tag{8c}$$

We use (7) and (8) to define risk-sensitive safe sets next.

B. Risk-Sensitive Safe Sets

Definition 1 (Risk-Sensitive Safe Sets): Let $\alpha \in (0, 1]$ and $r \in \mathbb{R}$ be given. The (α, r)-*risk-sensitive safe set for a given policy* $\pi \in \Pi$ is defined by

$$\mathcal{S}_\alpha^{r,\pi} := \left\{x \in S : \text{CVaR}_{\alpha,x}^\pi \left(\max_{t=0,1,\ldots,T} g_K(X_t)\right) \leq r\right\}. \tag{9}$$

The (α, r)-*risk-sensitive safe set* is defined by

$$\mathcal{S}_\alpha^r := \left\{x \in S : \inf_{\pi\in\Pi} \text{CVaR}_{\alpha,x}^\pi \left(\max_{t=0,1,\ldots,T} g_K(X_t)\right) \leq r\right\}. \tag{10}$$

We denote the infimum in (10) by $W_\alpha^*(x)$. Risk-sensitive safe sets are well-motivated. These sets represent the sets of initial states from which the maximum extent of constraint violation of the state trajectory, averaged over the $\alpha \cdot 100\%$ worst cases, can be made sufficiently small. The maximum extent of constraint violation of the state trajectory is the real-valued random variable $G := \max_{t=0,1,\ldots,T} g_K(X_t)$. We allow g_K to be negative so that decision-makers can encode preferences for trajectories remaining deeper inside of K over trajectories near the boundary of K, if desired. In our numerical example of a thermostatically controlled load, we allow g_K to take on both negative and nonnegative values to express a preference for trajectories that remain closer to 20.5 °C (Section V-A). In our numerical example of a stormwater system, however, we choose a nonnegative g_K to utilize all capacity in the water storage tanks without penalty (Section V-B).

Using CVaR to define risk-sensitive safe sets is well-justified from a decision-theoretic point of view because CVaR is a coherent risk measure. That is, CVaR satisfies the axioms of monotonicity, subadditivity, positive homogeneity, and translation equivariance. Section II provides intuitive interpretations for these axioms. Besides having an axiomatic justification, CVaR has the useful interpretation of quantifying the upper tail of a distribution. Indeed, CVaR provides a *quantitative characterization of risk aversion* by representing the expected cost in the $\alpha \cdot 100\%$ worst cases, where $\alpha \in (0, 1]$ is selected by the decision-maker. This interpretation is exact if continuous random variables in L^1 are evaluated.

Risk-sensitive safe sets generalize probabilistic safe sets [25] by quantifying the maximal *extent* of constraint violation at a given risk-sensitivity level rather than the probability of constraint violation. Risk-sensitive safe sets quantify *how much* constraint violation occurs on average in the $\alpha \cdot 100\%$ worst cases, whereas probabilistic safe sets [25] quantify whether or not constraint violation occurs with some probability. Indeed, let $\epsilon \in [0, 1]$ be a maximum tolerable probability of constraint violation. Choose $\alpha = 1$, $r = \epsilon$, and $g_K = I_{\bar{K}}$, where $I_{\bar{K}}(x) = 1$ if $x \notin K$ and $I_{\bar{K}}(x) = 0$ if $x \in K$. Then, the $(1, \epsilon)$-risk-sensitive safe set is

$$\mathcal{S}_1^\epsilon = \left\{x \in S : \inf_{\pi\in\Pi} E_x^\pi \left(\max_{t=0,1,\ldots,T} I_{\bar{K}}(X_t)\right) \leq \epsilon\right\} \tag{11}$$

which is the *maximal probabilistic safe set* at the ϵ-safety level [25] for the system of Section III. (The paper [25] considers discrete-time stochastic hybrid systems that evolve under Markov policies.)

Risk-sensitive safe sets indicate higher degrees of safety as α decreases and r decreases. We state this fact formally next.

Lemma 1: Suppose that $1 \geq \alpha_1 \geq \alpha_2 > 0$ and $r_1 \geq r_2$. Then, $\mathcal{S}_{\alpha_2}^{r_2} \subseteq \mathcal{S}_{\alpha_1}^{r_1}$. If $\pi \in \Pi$, then $\mathcal{S}_{\alpha_2}^{r_2,\pi} \subseteq \mathcal{S}_{\alpha_1}^{r_1,\pi}$.

Proof: Let $x \in S$ and $\pi \in \Pi$. Since $1 \geq \alpha_1 \geq \alpha_2 > 0$ and $G \in L^1(\Omega, \mathcal{B}(\Omega), P_x^\pi)$, $\text{CVaR}_{\alpha_2,x}^\pi(G) \geq \text{CVaR}_{\alpha_1,x}^\pi(G)$. Since $G = \max_{t=0,1,\ldots,T} g_K(X_t)$ and g_K is bounded, there exists a $b \in \mathbb{R}$ such that $G(\omega) \geq b$ for almost every $\omega \in \Omega$. Since CVaR is monotonic and $b \in \mathbb{R}$, $\text{CVaR}_{\alpha_1,x}^\pi(G) \geq b$. Take the infimum over $\pi \in \Pi$ to obtain $\inf_{\pi\in\Pi} \text{CVaR}_{\alpha_2,x}^\pi(G) \geq \inf_{\pi\in\Pi} \text{CVaR}_{\alpha_1,x}^\pi(G) \geq b$, which holds for any $x \in S$. Now, suppose $x \in \mathcal{S}_{\alpha_2}^{r_2}$. Then, $r_2 \geq \inf_{\pi\in\Pi} \text{CVaR}_{\alpha_2,x}^\pi(G) \geq \inf_{\pi\in\Pi} \text{CVaR}_{\alpha_1,x}^\pi(G)$. Since $r_1 \geq r_2$, we have $r_1 \geq \inf_{\pi\in\Pi} \text{CVaR}_{\alpha_1,x}^\pi(G)$, which shows that $x \in \mathcal{S}_{\alpha_1}^{r_1}$. The proof for the last statement is similar. ■

The risk-sensitive safe set $\mathcal{S}_\alpha^r$ specifies that the CVaR_α of the worst constraint violation of the state trajectory must be below a given threshold. In contrast, the safe set in [30] specifies that for each t the CVaR_α of the constraint violation of the state at time t must be below a given threshold. Hence, $\mathcal{S}_\alpha^r$ assesses the risk of the entire trajectory, whereas the safe set in [30] is concerned with the risk of each state in the trajectory separately. A specification that assesses the risk of the entire trajectory may be preferable in certain applications because this approach treats the trajectory as a unified entity representing the behavior of a control system.

C. Underapproximation Method

Risk-sensitive safe sets are well-motivated but difficult to compute due to the presence of the CVaR and the maximum. Before presenting our approach to estimate risk-sensitive safe sets, we describe related methods in further detail.

Several methods in the literature apply state-space augmentation techniques to estimate the risk of a random cost incurred by an MDP.[5] Bäuerle and Ott use dynamic programming (DP) to minimize the CVaR of a sum of stage costs by defining an augmented state space [16]. The range of the second state is $[0, \text{ess sup} \sum_{t=0}^T C_t]$, where C_t is the stage cost at time t

[5] An approach that does not require state-space augmentation is to evaluate a cumulative cost via a composition of risk functionals [47]. We take inspiration from this idea in Section IV.

[16, Remark 5.1]. This state-space augmentation approach has been extended to optimize certainty equivalent risk functionals for MDPs [17]. A certainty equivalent approximates the sum of the expectation and a function of the variance under particular conditions [17], and more generally, characterizes risk aversion in terms of functions of moments. However, CVaR provides a quantitative characterization of risk aversion by penalizing a random cost in a given fraction of the worst cases.

Chow *et al.* proposed a DP algorithm to minimize approximately the CVaR of a cumulative cost via state-space augmentation, where the additional state ranges from 0 to 1 [31]. This approach is expected to be more tractable than the approach in [16]; compare the ranges of the additional states. However, it is not known if the algorithm in [31] provides an upper bound or a lower bound to the solution to a CVaR-MDP problem. The algorithm in [31] is based on a CVaR Decomposition Theorem [40, Th. 6] [41, Th. 21, Lemma 22], which requires knowledge of the history of a stochastic process. How to remove the history dependence and apply the Decomposition Theorem to derive the algorithm in [31] is still an open research question.

The algorithms invented by [16] and [31] aim to minimize the CVaR of a *cumulative* cost subject to the dynamics of an MDP. The algorithm proposed by [40] aims to minimize the CVaR of a more general cost (not necessarily a sum) but is history-dependent, which limits its computational tractability. The proof of the DP algorithm in [40] requires an exchange between an essential supremum and an expectation, whose validity in multistage settings for MDPs with Borel state and control spaces is not known.

Here, we propose a method to provide tractable, theoretically guaranteed underapproximations to risk-sensitive safe sets, which we define via CVaR. We focus on CVaR due to its *quantitative characterization of risk aversion* and since we aim to assess the degree of safety of a control system in terms of rarer, higher-consequence outcomes. In contrast, a certainty equivalent assesses risk in terms of functions of variance and other moments. In particular, variance does not distinguish between rarer, higher-consequence outcomes in the upper tail and rarer, lower-consequence outcomes in the lower tail. Unlike the methods in [16], [17], [31], our method does not use state-space augmentation because this technique typically reduces computational tractability. For this reason, we do not augment the state space with the running maximum over each time period $Z_t := \max_{i=0,1,\ldots,t} g_K(X_i)$. The range of Z_t may be large since the bounds of g_K may be large. Instead of using state-space augmentation to handle the CVaR and the maximum, we use a scaled expectation to upper bound the CVaR and a log-sum-exponential function to upper bound the maximum, $G := \max_{t=0,1,\ldots,T} g_K(X_t)$. Our first main result is below.

Theorem 1 (Upper Bound for CVaR of G): For any $\pi \in \Pi$, $x \in S$, $\alpha \in (0,1]$, and $\gamma \geq 1$, it holds that

$$W_\alpha(x,\pi) := \text{CVaR}^\pi_{\alpha,x}(G) \leq \frac{1}{\gamma}\log\left(\frac{1}{\alpha}E^\pi_x\left(\sum_{t=0}^{T} e^{\gamma g_K(X_t)}\right)\right). \tag{12}$$

The quantity $W_\alpha(x,\pi)$ represents the maximum extent of constraint violation of the state trajectory, averaged over the $\alpha \cdot 100\%$ worst cases, when the system uses the policy π and starts from the state x. The right-hand side of (12) can be estimated more readily than $W_\alpha(x,\pi)$ for small α and provides a conservative approximation to $W_\alpha(x,\pi)$. If α is small, more samples of G are required to estimate $W_\alpha(x,\pi) = \text{CVaR}^\pi_{\alpha,x}(G)$ since small α corresponds to rarer larger realizations of G. (We are more interested in using small α for safety-critical applications.) Theorem 1 is powerful because it can be used to estimate the performance of *any* control policy $\pi \in \Pi$ with respect to $W_\alpha(x,\pi)$. Policies may be designed for different objectives, e.g., efficiency in power or fuel consumption, robustness to bounded adversarial disturbances, robustness to bounded nonlinearities, etc. It may be beneficial to estimate their performance with respect to a risk-sensitive safety criterion, such as $W_\alpha(x,\pi)$, efficiently. The proof of Theorem 1 requires two lemmas.

Lemma 2 (CVaR-Expectation Inequality): Let $(\Omega,\mathcal{F},\mu)$ be a probability space, $Y \in L^1(\Omega,\mathcal{F},\mu)$ such that $Y \geq 0$ a.e. w.r.t. μ, and $\alpha \in (0,1]$. Then, $\text{CVaR}_\alpha(Y) \leq \frac{1}{\alpha}E(Y)$.

A version of the inequality is stated without proof in [10]. We provide a short proof below.

Proof: Start from the CVaR definition (1), and select $s = 0$. Then, $\text{CVaR}_\alpha(Y) \leq \frac{1}{\alpha}E(\max(Y,0))$. Since $Y \geq 0$ a.e., $\max(Y,0) = Y$ a.e., so $\text{CVaR}_\alpha(Y) \leq \frac{1}{\alpha}E(Y)$. ■

Lemma 2 provides an upper bound for CVaR in terms of the expectation and the risk-sensitivity level α when nonnegative random variables are evaluated. In addition to Lemma 2, the proof of Theorem 1 requires the following result, which relates the CVaR of the logarithm to the logarithm of the CVaR.

Lemma 3 (CVaR-Log Inequality): Let $\alpha \in (0,1]$ and $Y \in L^\infty(\Omega,\mathcal{F},\mu)$. Suppose that there are real numbers $\bar{b} \geq \underline{b} > 0$ such that $\bar{b} \geq Y(\omega) \geq \underline{b}$ for every $\omega \in \Omega$. Then, $\text{CVaR}_\alpha(\log(Y)) \leq \log(\text{CVaR}_\alpha(Y))$.

Proof: Let $\alpha \in (0,1]$ and $\xi \in \mathcal{A}_\alpha$ (5b). Define $\mu_\xi(B) := \int_B \xi \mathrm{d}\mu$, where $B \in \mathcal{F}$. $(\Omega,\mathcal{F},\mu_\xi)$ is a probability space, and $\int_\Omega Y \mathrm{d}\mu_\xi$ is finite. View Y as a random variable on $(\Omega,\mathcal{F},\mu_\xi)$. It holds that $Y(\omega) \in (0,\infty)$ for all $\omega \in \Omega$, and $-\log$ is a convex function from $(0,\infty)$ to $\mathbb{R}$. Thus, by Jensen's Inequality, $\int_\Omega -\log(Y)\, \mathrm{d}\mu_\xi \geq -\log(\int_\Omega Y\, \mathrm{d}\mu_\xi)$. Moreover, since Y is nonnegative and bounded everywhere, ξ is nonnegative and bounded a.e., and by using the definition of μ_ξ, it follows that

$$\log\left(\int_\Omega Y\xi\, \mathrm{d}\mu\right) \geq \int_\Omega \log(Y)\xi\, \mathrm{d}\mu. \tag{13}$$

Since $\xi \in \mathcal{A}_\alpha$ is arbitrary in the analysis above, the inequality (13) holds for all $\xi \in \mathcal{A}_\alpha$. In addition, we have $\text{CVaR}_\alpha(Y) = \sup_{\xi\in\mathcal{A}_\alpha}\int_\Omega Y\xi \mathrm{d}\mu$ by (5), $\text{CVaR}_\alpha(Y) \in \mathbb{R}$ because $Y \in L^1(\Omega,\mathcal{F},\mu)$, and $\int_\Omega Y \mathrm{d}\mu_\xi = \int_\Omega Y\xi \mathrm{d}\mu \geq \underline{b} > 0$ for all $\xi \in \mathcal{A}_\alpha$. Thus, $\log(\text{CVaR}_\alpha(Y)) = \log(\sup_{\xi\in\mathcal{A}_\alpha}\int_\Omega Y\xi \mathrm{d}\mu) \in \mathbb{R}$. Since the natural logarithm is increasing,

$$\log\left(\text{CVaR}_\alpha(Y)\right) \geq \log\left(\int_\Omega Y\xi\, \mathrm{d}\mu\right) \quad \forall \xi \in \mathcal{A}_\alpha. \tag{14}$$

By (13) and (14), it holds that $\log(\text{CVaR}_\alpha(Y)) \geq \int_\Omega \log(Y)\xi \mathrm{d}\mu$ for all $\xi \in \mathcal{A}_\alpha$. Since the supremum is the least upper bound, we conclude that $\log(\text{CVaR}_\alpha(Y)) \geq \sup_{\xi\in\mathcal{A}_\alpha}\int_\Omega \log(Y)\xi \mathrm{d}\mu = \text{CVaR}_\alpha(\log(Y))$. ■

We use Lemmas 2 and 3 to prove Theorem 1.

Proof of Theorem 1: Note the log-sum-exp approximation for the maximum [44, Sec. 3.1.5, p. 72]: If $y \in \mathbb{R}^p$ and $\gamma \geq 1$, then

$$\max_{i=1,\ldots,p} y_i \overset{(a)}{\leq} \tfrac{1}{\gamma}\log\left(\textstyle\sum_{i=1}^{p} e^{\gamma y_i}\right) \leq \max_{i=1,\ldots,p} y_i + \tfrac{\log(p)}{\gamma}. \quad (15)$$

Let $\pi \in \Pi$, $x \in S$, $\alpha \in (0,1]$, and $\gamma \geq 1$. Recall that $G = \max_{t=0,1,\ldots,T} g_K(X_t) \in L^\infty(\Omega, \mathcal{B}(\Omega), P_x^\pi)$, where we have presented Ω and P_x^π at the start of Section III. Since g_K is $\mathbb{R}$-valued,

$$Y(\omega) := \textstyle\sum_{t=0}^{T} e^{\gamma g_K(X_t(\omega))} > 0 \quad \forall \omega \in \Omega. \quad (16)$$

Since g_K is bounded and Y is a sum of finitely many exponential functions of g_K, there exist real numbers $\bar{b} \geq \underline{b} > 0$ such that $\bar{b} \geq Y(\omega) \geq \underline{b}$ for every $\omega \in \Omega$. It follows that $Y \in L^\infty(\Omega, \mathcal{B}(\Omega), P_x^\pi)$ satisfies the assumptions of Lemma 3, and thus,

$$\text{CVaR}_{\alpha,x}^{\pi}(\log(Y)) \leq \log(\text{CVaR}_{\alpha,x}^{\pi}(Y)). \quad (17)$$

By the inequality (a) in (15) and by the definitions of G and Y, the inequality $G \leq \frac{1}{\gamma}\log(\sum_{t=0}^{T} e^{\gamma g_K(X_t)}) = \frac{1}{\gamma}\log(Y)$ holds a.e. w.r.t. P_x^π. Since CVaR is monotonic and positively homogeneous, and since $\frac{1}{\gamma} > 0$,

$$\text{CVaR}_{\alpha,x}^{\pi}(G) \leq \text{CVaR}_{\alpha,x}^{\pi}\left(\tfrac{1}{\gamma}\log(Y)\right) = \tfrac{1}{\gamma}\text{CVaR}_{\alpha,x}^{\pi}(\log(Y)). \quad (18)$$

We use (17) and (18) to find that

$$\text{CVaR}_{\alpha,x}^{\pi}(G) \leq \tfrac{1}{\gamma}\log(\text{CVaR}_{\alpha,x}^{\pi}(Y)). \quad (19)$$

Note that $\text{CVaR}_{\alpha,x}^{\pi}(Y) \in \mathbb{R}$ such that $\text{CVaR}_{\alpha,x}^{\pi}(Y) > 0$. Indeed, $Y \in L^\infty(\Omega, \mathcal{B}(\Omega), P_x^\pi)$ and so is also an element of $L^1(\Omega, \mathcal{B}(\Omega), P_x^\pi)$, hence $\text{CVaR}_{\alpha,x}^{\pi}(Y) \in \mathbb{R}$. Y is bounded everywhere, and in particular, from below by a real number $\underline{b} > 0$. Therefore, $\text{CVaR}_{\alpha,x}^{\pi}(Y) \geq \underline{b} > 0$. Consequently, $\log(\text{CVaR}_{\alpha,x}^{\pi}(Y)) \in \mathbb{R}$. In addition, the assumptions of Lemma 2 are satisfied, and therefore,

$$\text{CVaR}_{\alpha,x}^{\pi}(Y) \leq \tfrac{1}{\alpha}E_x^\pi(Y). \quad (20)$$

Use (19), (20), and $\log$ being increasing to derive that $\text{CVaR}_{\alpha,x}^{\pi}(G) \leq \frac{1}{\gamma}\log(\text{CVaR}_{\alpha,x}^{\pi}(Y)) \leq \frac{1}{\gamma}\log(\frac{1}{\alpha}E_x^\pi(Y))$. ■

We use the conclusion of Theorem 1 to define particular subsets of the state space. First, we call these sets approximations, and then, we prove that they are underapproximations to risk-sensitive safe sets in Theorem 2.

Definition 2 (Approximations to Risk-Sensitive Safe Sets): Let $\alpha \in (0,1]$, $r \in \mathbb{R}$, and $\gamma \geq 1$ be given. The (α, r, γ)-*approximation set for a given policy* $\pi \in \Pi$ is defined by

$$\mathcal{U}_{\alpha,\gamma}^{r,\pi} := \left\{x \in S : \tfrac{1}{\alpha}E_x^\pi\left(\textstyle\sum_{t=0}^{T} e^{\gamma g_K(X_t)}\right) \leq e^{\gamma r}\right\}. \quad (21)$$

The (α, r, γ)-*approximation set* is defined by

$$\mathcal{U}_{\alpha,\gamma}^{r} := \left\{x \in S : \inf_{\pi\in\Pi} \tfrac{1}{\alpha}E_x^\pi\left(\textstyle\sum_{t=0}^{T} e^{\gamma g_K(X_t)}\right) \leq e^{\gamma r}\right\}. \quad (22)$$

We denote the infimum in (22) by

$$J_{\alpha,\gamma}^{*}(x) := \inf_{\pi\in\Pi} J_{\alpha,\gamma}(x,\pi) := \inf_{\pi\in\Pi} \tfrac{1}{\alpha}E_x^\pi\left(\textstyle\sum_{t=0}^{T} e^{\gamma g_K(X_t)}\right) \quad (23)$$

where Π is the set of randomized history-dependent policies, which also includes deterministic Markov policies. Estimating $J_{\alpha,\gamma}^{*}$ is the critical step for estimating the sets $\mathcal{U}_{\alpha,\gamma}^{r}$. The problem of estimating $J_{\alpha,\gamma}^{*}$ is an MDP problem. Thus, $J_{\alpha,\gamma}^{*}$ and a deterministic Markov policy $\pi_\gamma \in \Pi$ such that $J_{\alpha,\gamma}(x, \pi_\gamma) = J_{\alpha,\gamma}^{*}(x)$ for all $x \in S$ can be computed via DP, in principle, if a measurable selection condition holds.[6] Therefore, for a fixed $\gamma \geq 1$, an algorithm to estimate $\{J_{\alpha,\gamma}^{*} : \alpha \in \Lambda\}$, where $\Lambda \subseteq (0,1]$ is a family of risk-sensitivity levels, exists and is tractable. The next theorem shows that the sets in Definition 2 are underapproximations to risk-sensitive safe sets (Definition 1).

Theorem 2 (Underapproximations to Risk-Sensitive Safe Sets): Let $\alpha \in (0,1]$, $r \in \mathbb{R}$, and $\gamma \geq 1$. For any policy $\pi \in \Pi$, it holds that

$$\mathcal{U}_{\alpha,\gamma}^{r,\pi} \subseteq \mathcal{S}_{\alpha}^{r,\pi} \quad (24)$$

where $\mathcal{U}_{\alpha,\gamma}^{r,\pi}$ is defined by (21) and $\mathcal{S}_{\alpha}^{r,\pi}$ is defined by (9). Moreover, the (α, r, γ)-approximation set is a subset of the (α, r)-risk-sensitive safe set, i.e.,

$$\mathcal{U}_{\alpha,\gamma}^{r} \subseteq \mathcal{S}_{\alpha}^{r} \quad (25)$$

where $\mathcal{U}_{\alpha,\gamma}^{r}$ is defined by (22) and $\mathcal{S}_{\alpha}^{r}$ is defined by (10).

Proof: Equation (24) follows from Theorem 1. Let $\alpha \in (0,1]$, $r \in \mathbb{R}$, $\gamma \geq 1$, and $\pi \in \Pi$ be given. Let $x \in \mathcal{U}_{\alpha,\gamma}^{r,\pi}$. Then,

$$\tfrac{1}{\alpha}E_x^\pi\left(\textstyle\sum_{t=0}^{T} e^{\gamma g_K(X_t)}\right) \leq e^{\gamma r} \quad (26)$$

where the left-hand side is bounded below by a positive real number since $Y := \sum_{t=0}^{T} e^{\gamma g_K(X_t)}$ is as well. It follows that $\log(\frac{1}{\alpha}E_x^\pi(\sum_{t=0}^{T} e^{\gamma g_K(X_t)}))$ is finite. Since the natural logarithm is increasing and $\gamma \geq 1$, we have

$$\tfrac{1}{\gamma}\log\left(\tfrac{1}{\alpha}E_x^\pi\left(\textstyle\sum_{t=0}^{T} e^{\gamma g_K(X_t)}\right)\right) \leq r. \quad (27)$$

By Theorem 1, it holds that

$$\text{CVaR}_{\alpha,x}^{\pi}(G) \leq \tfrac{1}{\gamma}\log\left(\tfrac{1}{\alpha}E_x^\pi\left(\textstyle\sum_{t=0}^{T} e^{\gamma g_K(X_t)}\right)\right). \quad (28)$$

Combine (27) and (28) to find that $\text{CVaR}_{\alpha,x}^{\pi}(G) \leq r$, which shows that $x \in \mathcal{S}_{\alpha}^{r,\pi}$ and proves (24). Now, to prove (25), let $x \in \mathcal{U}_{\alpha,\gamma}^{r}$, which implies that

$$\inf_{\pi\in\Pi} \tfrac{1}{\alpha}E_x^\pi\left(\textstyle\sum_{t=0}^{T} e^{\gamma g_K(X_t)}\right) \leq e^{\gamma r}. \quad (29)$$

Let $\epsilon > 0$ be given. Since the left-hand side of (29) is finite, there is a $\pi^\epsilon \in \Pi$ such that

$$\begin{aligned}\tfrac{1}{\alpha}E_x^{\pi^\epsilon}\left(\textstyle\sum_{t=0}^{T} e^{\gamma g_K(X_t)}\right) &\leq \epsilon + \inf_{\pi\in\Pi} \tfrac{1}{\alpha}E_x^\pi\left(\textstyle\sum_{t=0}^{T} e^{\gamma g_K(X_t)}\right) \\ &\leq \epsilon + e^{\gamma r}\end{aligned} \quad (30)$$

where the second line holds by (29). Note that the quantity $\log(\frac{1}{\alpha}E_x^{\pi^\epsilon}(\sum_{t=0}^{T} e^{\gamma g_K(X_t)}))$ is finite. Take the logarithm of (30) and then divide by $\gamma \geq 1$ to obtain

$$\tfrac{1}{\gamma}\log\left(\tfrac{1}{\alpha}E_x^{\pi^\epsilon}\left(\textstyle\sum_{t=0}^{T} e^{\gamma g_K(X_t)}\right)\right) \leq \tfrac{1}{\gamma}\log\left(\epsilon + e^{\gamma r}\right). \quad (31)$$

[6] Measurable selection conditions, e.g., see [45, Ch. 3.3] or [15], are commonly invoked to guarantee the existence of a policy that optimizes or nearly optimizes an expected cumulative cost subject to an MDP.

By Theorem 1, it holds that

$$\text{CVaR}^{\pi^\epsilon}_{\alpha,x}(G) \leq \tfrac{1}{\gamma}\log\left(\tfrac{1}{\alpha}E^{\pi^\epsilon}_x\left(\textstyle\sum_{t=0}^{T} e^{\gamma g_K(X_t)}\right)\right). \tag{32}$$

Therefore, $\text{CVaR}^{\pi^\epsilon}_{\alpha,x}(G) \leq \frac{1}{\gamma}\log(\epsilon + e^{\gamma r})$. Since $\pi^\epsilon \in \Pi$, it follows that

$$W^*_\alpha(x) := \inf_{\pi\in\Pi} \text{CVaR}^{\pi}_{\alpha,x}(G) \leq \text{CVaR}^{\pi^\epsilon}_{\alpha,x}(G). \tag{33}$$

Consequently, we have

$$W^*_\alpha(x) \leq \tfrac{1}{\gamma}\log\left(\epsilon + e^{\gamma r}\right). \tag{34}$$

This analysis holds for any $\epsilon > 0$. Let $\epsilon \to 0$, and use the continuity of the logarithm to obtain

$$W^*_\alpha(x) \leq \lim_{\epsilon\to 0}\tfrac{1}{\gamma}\log\left(\epsilon + e^{\gamma r}\right) = \tfrac{1}{\gamma}\log\left(\lim_{\epsilon\to 0}\epsilon + e^{\gamma r}\right) = r. \tag{35}$$

Since $W^*_\alpha(x) \leq r$, we conclude that $x \in \mathcal{S}^r_\alpha$. Since any $x \in \mathcal{U}^r_{\alpha,\gamma}$ is also an element of $\mathcal{S}^r_\alpha$, it holds that $\mathcal{U}^r_{\alpha,\gamma} \subseteq \mathcal{S}^r_\alpha$. ■

Since we have shown that $\mathcal{U}^{r,\pi}_{\alpha,\gamma}$ and $\mathcal{U}^{r}_{\alpha,\gamma}$ are subsets of the risk-sensitive safe sets, $\mathcal{S}^{r,\pi}_\alpha$ and $\mathcal{S}^r_\alpha$, respectively, we now refer to $\mathcal{U}^{r,\pi}_{\alpha,\gamma}$ and $\mathcal{U}^{r}_{\alpha,\gamma}$ as *underapproximations*.

Remark 1 (Assessment of Approximation Errors): Three approximations are required for the proof above. First, we use a soft-maximum, under which we have

$$0 \leq \tfrac{1}{\gamma}\text{CVaR}^{\pi}_{\alpha,x}(\log(Y)) - \text{CVaR}^{\pi}_{\alpha,x}(G) \leq \tfrac{\log(T+1)}{\gamma} \tag{36}$$

where $Y = \sum_{t=0}^{T} e^{\gamma g_K(X_t)}$, and there are positive constants $\underline{b}$ and $\bar{b}$ (which depend on T, γ, and the bounds of g_K) such that $Y \in [\underline{b}, \bar{b}]$ everywhere. The inequality (36) implies an improved approximation with larger values of γ or smaller values of T. However, since it is not feasible to optimize $\frac{1}{\gamma}\text{CVaR}^{\pi}_{\alpha,x}(\log(Y))$ directly, our next step is to leverage the CVaR-log inequality provided by Lemma 3. The associated error is given by

$$\eta^{\pi,\gamma}_{\alpha,x} := \tfrac{1}{\gamma}\log(\text{CVaR}^{\pi}_{\alpha,x}(Y)) - \tfrac{1}{\gamma}\text{CVaR}^{\pi}_{\alpha,x}(\log(Y)) \geq 0. \tag{37}$$

Since the range of Y is $[\underline{b}, \bar{b}]$, it follows that $\eta^{\pi,\gamma}_{\alpha,x} \leq \frac{1}{\gamma}\log(\bar{b}/\underline{b})$. Therefore, we anticipate a smaller error $\eta^{\pi,\gamma}_{\alpha,x}$ when Y has a smaller range, which occurs when T is smaller, for example.

The last approximation is $\log(\text{CVaR}^{\pi}_{\alpha,x}(Y)) \leq \log(\frac{1}{\alpha}E^\pi_x(Y))$, which of course is poor as $\alpha \to 0$. However, for a fixed $\alpha \in (0,1)$, we anticipate that this approximation performs well when P^π_x has a fat (upper) tail, which we state formally in the following lemma.

Lemma 4 (Tightness of $\log(\text{CVaR}_\alpha(Y)) \leq \log(\frac{1}{\alpha}E(Y))$): Assume the conditions of Lemma 3, and let $\alpha \in (0,1)$. Suppose that for some finite $m > 0$, it holds that

$$0 < m\int_0^{1-\alpha}\text{VaR}_{1-p}(Y)\,\mathrm{d}p \leq \int_{1-\alpha}^{1}\text{VaR}_{1-p}(Y)\,\mathrm{d}p. \tag{38}$$

Then, $0 \leq \log(\frac{1}{\alpha}E(Y)) - \log(\text{CVaR}_\alpha(Y)) \leq \log(\frac{1}{m}+1)$.

Remark 2 (Fat tail condition (38)): The second inequality in (38) means that the cumulative VaR in the upper α-fraction of the distribution of Y, $\int_{1-\alpha}^{1}\text{VaR}_{1-p}(Y)\mathrm{d}p$, is at least m times greater than the cumulative VaR in the lower $(1-\alpha)$-fraction of the distribution of Y, $\int_0^{1-\alpha}\text{VaR}_{1-p}(Y)\mathrm{d}p$. The maximum value of m that satisfies (38) is $\hat{m} = \frac{\int_{1-\alpha}^{1}\text{VaR}_{1-p}(Y)\mathrm{d}p}{\int_0^{1-\alpha}\text{VaR}_{1-p}(Y)\mathrm{d}p}$, which gives a measure of tail "fatness." For example, if the distribution of Y is a standard log-normal with parameters $\mu = 0$ and $\sigma = 1$, and if $\alpha = 0.05$, then numerical integration yields $\hat{m} \approx \frac{0.42}{1.2} \approx 0.35$. If σ is increased to 2 under the same conditions, then $\hat{m} \approx \frac{4.7}{2.7} \approx 1.7$.

Next, we prove Lemma 4.

Proof of Lemma 4: The representation of CVaR in (3) and the inequality (38) imply that

$$\frac{1}{\alpha}\int_0^{1-\alpha}\text{VaR}_{1-p}(Y)\mathrm{d}p \leq \frac{\text{CVaR}_\alpha(Y)}{m}. \tag{39}$$

The expectation and the VaR are related by $E(Y) = \text{CVaR}_1(Y) = \int_0^1 \text{VaR}_{1-p}(Y)\mathrm{d}p$. It follows that $\frac{1}{\alpha}E(Y) \leq (\frac{1}{m}+1)\text{CVaR}_\alpha(Y)$. From this and Lemma 2, we have

$$\text{CVaR}_\alpha(Y) \leq \tfrac{1}{\alpha}E(Y) \leq \left(\tfrac{1}{m}+1\right)\text{CVaR}_\alpha(Y). \tag{40}$$

Then, take the logarithm of (40) and subtract $\log(\text{CVaR}_\alpha(Y)) \in \mathbb{R}$ to complete the derivation. ■

From Theorem 2, we obtain tractable underapproximations to risk-sensitive safe sets. In practice, one selects $\gamma \geq 1$ manually and then estimates $J^*_{\alpha,\gamma}$ (23) for a family of risk-sensitivity levels. For a fixed γ, only *one* MDP problem on the original state space needs to be solved for any family of risk-sensitivity levels because $J^*_{\alpha,\gamma}$ is a standard MDP problem scaled by α. In Section V, which presents numerical examples, we take one approach to choose a suitable value of γ manually by visual inspection. Before proceeding to the numerical examples, we present one additional theoretical contribution.

IV. Toward a Parameter-Independent Safety Analysis Framework

Previously, we have defined risk-sensitive safe sets in terms of the CVaR of a maximum random cost. However, this risk-sensitive safety criterion is difficult to optimize exactly without using state-space augmentation, which motivated us to derive a parameter-dependent upper bound. One may wonder whether there is another coherent risk functional (ideally related to CVaR) that admits an upper bound, which can be computed via DP on the original state space without an additional parameter that requires tuning. The answer is indeed positive, as presented below.

Definition 3 (Proposed Risk Functional): Let $\alpha \in (0,1]$, $x \in S$, $\pi \in \Pi$, and $Y \in L^\infty(\Omega, \mathcal{B}(\Omega), P^\pi_x)$ be given. Let $\mathcal{D}_\alpha$ be a set of tuples of densities. Each tuple $\zeta \in \mathcal{D}_\alpha$ takes the form $\zeta = (\xi_0, \xi_1, \ldots, \xi_{T-1})$, where the properties of the densities follow. For each t, $\xi_t(\cdot|\cdot,\cdot) : S \times S \times A \to \mathbb{R}$ is Borel measurable, and for every $(x,u) \in S \times A$, it holds that $\xi_t(\cdot|x,u) \in \mathcal{R}_\alpha(x,u)$. Here, $\mathcal{R}_\alpha(x,u)$ is the set of Borel-measurable functions of the form $\nu : S \to \mathbb{R}$ such that $\nu \in [0, \alpha^{-1/T}]$ a.e. w.r.t. $Q(\cdot|x,u)$ and $\int_S \nu\,\mathrm{d}Q(\cdot|x,u) = 1$. We define $\rho^\pi_{\alpha,x}(Y)$ by

$$\rho^\pi_{\alpha,x}(Y) := \sup_{(\xi_0,\xi_1,\ldots,\xi_{T-1})\in\mathcal{D}_\alpha}\int_\Omega Y\prod_{t=0}^{T-1}\xi_t(x_{t+1}|x_t,u_t)\,\mathrm{d}P^\pi_x. \tag{41}$$

Remark 3 (Interpretation for $\mathcal{R}_\alpha(x,u)$): $\mathcal{R}_\alpha(x,u)$ is related to the set of densities in the CVaR representation given by (5). If the probability space is $(S, \mathcal{B}(S), Q(\cdot|x,u))$, then $\mathcal{A}_{\alpha'} = \mathcal{R}_\alpha(x,u)$, where $\alpha' = \alpha^{1/T}$.

Remark 4 (Interpretation for $\rho^\pi_{\alpha,x}$): Although we do not yet have an exact interpretation for $\rho^\pi_{\alpha,x}$, we provide a preliminary interpretation here. The quantity $\rho^\pi_{\alpha,x}(Y)$ is a distributionally robust expectation of Y, such that an uncertainty ξ_t perturbs the system's nominal transition law Q at each time t. ξ_t may depend on the current time, state, and control. Moreover, $\rho^\pi_{\alpha,x}(Y)$ strikes a balance between the expectation and CVaR, as formalized below.

Lemma 5 (Coherence of $\rho^\pi_{\alpha,x}$, relation to CVaR): The risk functional $\rho^\pi_{\alpha,x} : L^\infty(\Omega, \mathcal{B}(\Omega), P^\pi_x) \rightarrow \mathbb{R}$ is coherent. In addition, for any $Y \in L^\infty(\Omega, \mathcal{B}(\Omega), P^\pi_x)$, the inequality $E^\pi_x(Y) \leq \rho^\pi_{\alpha,x}(Y) \leq \text{CVaR}^\pi_{\alpha,x}(Y)$ holds.

Proof: The first step is to verify the properties of monotonicity, subadditivity, translation equivariance, and positive homogeneity, which we omit in the interest of space. To show that $E^\pi_x(Y) \leq \rho^\pi_{\alpha,x}(Y)$, note that $\zeta = (\xi_0, \xi_1, \dots, \xi_{T-1})$ such that ξ_t equals 1 for each t is an element of $\mathcal{D}_\alpha$. The inequality $\rho^\pi_{\alpha,x}(Y) \leq \text{CVaR}^\pi_{\alpha,x}(Y)$ follows from (8b)–(8c). ■

We use the risk functional (41) to define a safe set.

Definition 4 ($\bar{\mathcal{S}}^r_\alpha$-Risk-Sensitive Safe Set): For any $\alpha \in (0, 1]$ and $r \in \mathbb{R}$, define $\bar{\mathcal{S}}^r_\alpha := \{x \in S : \inf_{\pi \in \Pi} \rho^\pi_{\alpha,x}(Y) \leq r\}$.

Definition 4 is inspired by Definition 1, and the form of $\rho^\pi_{\alpha,x}$ (41) is inspired by the representation for CVaR in (8b)–(8c). We emphasize a key distinction. In (41), there is a function ξ_t for each t that depends on the current state and control. In (8b)–(8c), however, each function in $\mathcal{A}^\pi_{\alpha,x}$ depends on the entire history. The "separable" structure of (41) allows us to derive a DP algorithm on the original state space to upper bound $\inf_{\pi \in \Pi} \rho^\pi_{\alpha,x}(Y)$ without using a parameter that requires tuning. In this section, we make two assumptions.

Assumption 1 (Properties of Y): We consider the case when $Y := c_T(X_T) + \sum_{t=0}^{T-1} c_t(X_t, U_t)$ is cumulative. The functions $c_t : S \times A \rightarrow \mathbb{R}$ for all $t \in \{0, 1, \dots, T-1\}$ and $c_T : S \rightarrow \mathbb{R}$ are bounded and upper semi-continuous (usc).

Assumption 2 (Continuity property of Q): The transition kernel Q (6) is continuous in total variation; i.e., if $(x_n, u_n) \rightarrow (x, u)$, then $|Q(\cdot|x_n, u_n) - Q(\cdot|x, u)|(S) \rightarrow 0$.

Remark 5 (Example that satisfies Assumption 2): Suppose that P_D has a continuous nonnegative density and f in (6) has the form $f(x, u, d) = f_1(x, u) + d \cdot f_2(x, u)$, where $W = S$ is a vector space with field $\mathbb{R}$, $f_1 : S \times A \rightarrow S$ and $f_2 : S \times A \rightarrow \mathbb{R}$ are continuous, and f_2 is nonzero. Then, by Scheffé's Lemma, Assumption 2 is satisfied. We note that the continuity of f is a typical condition in stochastic control, e.g., see [15, p. 209], and requiring additional structure on the dynamics to achieve tractable algorithms is standard. For example, under some assumptions the dynamics may be decomposed into overlapping systems, to obtain conservative underapproximations to reachable sets for continuous-time, nonstochastic systems [23], [55]. A mixed monotone structure has been assumed to approximate reachable sets for discrete-time nonstochastic systems, with applications to traffic safety [56], [57]. More broadly, additive continuous noise is a realistic assumption in many domains, e.g., additive Gaussian noise in information theory and control (classical references include [4], [58]) and additive Brownian motion in continuous-time epidemiological modeling [59], [60].

Boundedness and upper semi-continuity of c_t for all t ensures that $Y \in L^\infty(\Omega, \mathcal{B}(\Omega), P^\pi_x)$ for any $x \in S$ and $\pi \in \Pi$. Also, boundedness of c_t ensures that the iterates of a DP recursion are bounded, which we use to show that a supremum over $\mathcal{R}_\alpha(x, u)$ of the form $\phi(x, u) := \sup_{\xi \in \mathcal{R}_\alpha(x,u)} \int_S J\xi \, \mathrm{d}Q(\cdot|x, u)$ is attained (Lemma 6, Appendix). This attainment and Assumption 2 together guarantee that the supremum is usc in (x, u) (Lemma 8, Appendix). The upper semi-continuity of the supremum permits the derivation of an upper bound for $\inf_{\pi \in \Pi} \rho^\pi_{\alpha,x}(Y)$ via DP.

Theorem 3 (DP to Upper Bound $\inf_{\pi \in \Pi} \rho^\pi_{\alpha,x}(Y)$): Let Assumptions 1–2 hold, and let $\alpha \in (0, 1]$ be given. Define

$$J^\alpha_T := c_T \tag{42a}$$

and for $t = T-1, \dots, 1, 0$, define

$$J^\alpha_t(x) := \inf_{u \in A} v^\alpha_t(x, u) \quad \forall x \in S \tag{42b}$$

where $v^\alpha_t := c_t + \varphi^\alpha_t$ and

$$\varphi^\alpha_t(x, u) := \sup_{\xi \in \mathcal{R}_\alpha(x,u)} \int_S J^\alpha_{t+1}\xi \, \mathrm{d}Q(\cdot|x, u) \tag{42c}$$

for all $(x, u) \in S \times A$. Then, J^α_t is usc and bounded for all $t = 0, 1, \dots, T$. For all $\epsilon > 0$, there is a deterministic Markov policy $\pi^*_\epsilon \in \Pi$ such that $\rho^{\pi^*_\epsilon}_{\alpha,x}(Y) \leq J^\alpha_0(x) + \epsilon$ for all $x \in S$. In particular, $\inf_{\pi \in \Pi} \rho^\pi_{\alpha,x}(Y) \leq J^\alpha_0(x)$ for all $x \in S$.

A proof for Theorem 3 is in the Appendix, where we include supporting results as well.

Theorem 3 is exciting for two main reasons: 1) It provides a more numerically tractable way to estimate safe sets (the upper bound does not have a parameter that requires tuning, and the algorithm does not require an augmented state space); and 2) more broadly, the result initiates new avenues for tractable solutions to risk-sensitive safety analysis problems.

V. Numerical Examples

Here, we present examples of risk-sensitive safe sets and their underapproximations as in Definition 1 for a temperature system and a stormwater system.[7] For each example, we have chosen a value of γ by exploring increasing integer values and then stopping the exploration when improvements in the estimates of $\mathcal{U}^r_{\alpha,\gamma}$ were no longer apparent.

A. Temperature System

Consider a thermostatically controlled load evolving on a finite-time horizon $t = 0, 1, \dots, T-1$ via a deterministic Markov policy $\pi = (\pi_0, \pi_1, \dots, \pi_{T-1})$,

$$X_{t+1} = aX_t + (1-a)(b - \eta\bar{r}\bar{p}\pi_t(X_t)) + D_t.$$

This model is from [29] and [48]. X_t is the $\mathbb{R}$-valued random temperature (°C) of a thermal mass at time t. $\pi_t(X_t)$ is the [0,1]-valued control at time t. The amount of power supplied to the system decreases as the value of the control increases from 0 to 1. $(D_0, D_1, \dots, D_{T-1})$ is a $\mathbb{R}$-valued, iid stochastic process that arises due to environmental uncertainties. We consider three discrete distributions for the disturbance process, where each distribution has a distinct skew (left skew, no skew, or right

[7] We used the Tufts Linux Research Cluster (Medford, MA) with MATLAB (The Mathworks, Inc.). Our code is available from https://github.com/risk-sensitive-reachability/IEEE-TAC-2021.

TABLE I
TEMPERATURE SYSTEM PARAMETERS

Symbol	Description	Value
a	time delay	$e^{\frac{-\triangle\tau}{\bar{c}\bar{r}}}$ (no units)
b	temperature shift	32 °C
$\bar{c}$	thermal capacitance	$2\ \frac{\text{kWh}}{°\text{C}}$
η	control efficiency	0.7 (no units)
K	constraint set	$[20, 21]$ °C
$\bar{p}$	range of energy transfer to/from thermal mass	14 kW
$\bar{r}$	thermal resistance	$2\ \frac{°\text{C}}{\text{kW}}$
$\triangle\tau$	duration of $[t, t+1)$	$\frac{5}{60}$ h
T	length of discrete time horizon	12 (= 1 h)
A	control space	$[0, 1]$ (no units)
S	state space	$[18, 23]$ °C

h = hours, kW = kilowatts, °C = degrees Celsius.

skew). In each distribution, the minimum disturbance value is -0.5 °C, and the maximum disturbance value is 0.5 °C. Table I provides the model parameters.

We have chosen $g_K(X_t) = \max(X_t - 21, 20 - X_t)$ to quantify the extent of constraint violation of the state X_t with respect to the constraint set $K = [20, 21]$ °C. K is a temperature range, where the state trajectory should remain inside whenever possible. For different values of γ (see next paragraph), we have implemented classical DP with linear interpolation to estimate

$$J_\gamma^*(x) := \inf_{\pi\in\Pi} J_\gamma(x,\pi) := \inf_{\pi\in\Pi} E_x^\pi \left(\textstyle\sum_{t=0}^T e^{\gamma g_K(X_t)} \right) \quad (43)$$

and a deterministic Markov policy $\pi_\gamma \in \Pi$ such that $J_\gamma^*(x) = J_\gamma(x, \pi_\gamma)$ for all $x \in S$. DP on continuous state and control spaces is implemented typically via discretization and interpolation. In particular, we have discretized the set of controls $A = [0, 1]$ and the set of states $S = [18, 23]$ °C uniformly at a resolution of 0.1. To improve the efficiency of DP, approximate DP methods are being developed, e.g., see [49], [50], and the references therein. While these methods are exciting, we leave investigations of their applicability to risk-sensitive safety analysis for future work.

We have used $\gamma \in \Gamma := \{3, 4, \ldots, 20\}$ because for all $y \in S$ and $\gamma \in \Gamma$, the stage cost $e^{\gamma g_K(y)}$ is at most $e^{20\cdot 2}$, a large number that a personal computer can handle. We have considered risk-sensitivity levels from nearly risk-neutral ($\alpha = 0.99$) to more risk-averse (α near 0). Specifically, we have chosen $\alpha \in \Lambda := \{0.99, 0.05, 0.01, 0.005, 0.001\}$. A typical risk-sensitivity level is $\alpha = 0.05$ or $\alpha = 0.01$, and we have considered smaller values of α as well. For $\gamma \in \Gamma$ and $\alpha \in \Lambda$, we have estimated $J_{\alpha,\gamma}^*$ (23) by dividing our estimate of J_γ^* (43) by α. Let $\hat{S}$ denote the state space grid. By using our estimate of π_γ, we have simulated 100,000 trajectories from each initial state $x \in \hat{S}$ to generate an empirical distribution of $G := \max_{t=0,1,\ldots,T} g_K(X_t)$. Then, for each $\alpha \in \Lambda$, we have used a consistent CVaR estimator [37, p. 300] to estimate $\text{CVaR}_{\alpha,x}^{\pi_\gamma}(G)$.

Fig. 3 provides a visual summary of the inequality that we have proved in Theorem 1:

$$\begin{aligned} &\text{CVaR}_{\alpha,x}^{\pi_\gamma}\left(\max_{t=0,1,\ldots,T} g_K(X_t) \right) \\ &\leq \tfrac{1}{\gamma} \log \left(\tfrac{1}{\alpha} E_x^{\pi_\gamma} \left(\textstyle\sum_{t=0}^T e^{\gamma g_K(X_t)} \right) \right). \end{aligned} \quad (44)$$

Each plot in Fig. 3 shows estimates of the right-hand side of (44) on the vertical axis versus estimates of the left-hand side of (44) on the horizontal axis for the five values of α in Λ. In each plot, each solid colored line consists of five points, one for each $\alpha \in \Lambda$. Points associated with smaller values of α (more risk-averse) are positioned farther away from the origin. In each plot, there are three solid colored lines, one for each distribution of the disturbance process. In each plot, $\gamma \in \Gamma$ and an initial state $x \in \hat{S}$ are fixed. We have chosen initial states inside or on the boundary of the constraint set $K = [20, 21]$ °C. Fig. 3 is consistent with the inequality that we have proved in Theorem 1 since the solid colored lines are located above the gray line of slope 1. Fig. 3 suggests that there is no unique value of γ that provides the best approximation for all initial states x, risk-sensitivity levels α, and disturbance distributions.

However, by Theorem 2, we have flexibility in choosing the value of γ. In particular, we favor the quality of the approximations for small values of α due to our focus on safety and present sets using $\gamma = 14$ as an example of a value that reflects this preference (Fig. 4).[8] Fig. 4 provides estimates of the (α, r)-risk-sensitive safe set for $\pi_\gamma \in \Pi$ (9)

$$\mathcal{S}_\alpha^{r,\pi_\gamma} := \left\{ x \in S : \text{CVaR}_{\alpha,x}^{\pi_\gamma}\left(\max_{t=0,1,\ldots,T} g_K(X_t) \right) \leq r \right\}$$

and the (α, r, γ)-underapproximation set (22)

$$\begin{aligned} \mathcal{U}_{\alpha,\gamma}^r &= \left\{ x \in S : \tfrac{1}{\alpha} E_x^{\pi_\gamma} \left(\textstyle\sum_{t=0}^T e^{\gamma g_K(X_t)} \right) \leq e^{\gamma r} \right\} \\ &= \left\{ x \in S : \tfrac{1}{\gamma} \log \left(\tfrac{1}{\alpha} E_x^{\pi_\gamma} \left(\textstyle\sum_{t=0}^T e^{\gamma g_K(X_t)} \right) \right) \leq r \right\}. \end{aligned}$$

Note that $\mathcal{U}_{\alpha,\gamma}^r = \mathcal{U}_{\alpha,\gamma}^{r,\pi_\gamma}$.[9] In Fig. 4, estimates of $\mathcal{U}_{\alpha,\gamma}^{r,\pi_\gamma}$ (solid red) and $\mathcal{S}_\alpha^{r,\pi_\gamma}$ (white circles with blue boundary) are shown for the risk-sensitivity levels $\alpha \in \Lambda$ and various $r \in \mathbb{R}$ with $\gamma = 14$. The estimates of $\mathcal{U}_{\alpha,\gamma}^{r,\pi_\gamma}$ are subsets of the estimates of $\mathcal{S}_\alpha^{r,\pi_\gamma}$, which we expect by Theorem 2. The estimates of $\mathcal{S}_\alpha^{r,\pi_\gamma}$ form an increasing sequence of subsets as α increases and r increases, which is consistent with Lemma 1.

B. Stormwater System

Next, we illustrate risk-sensitive safety analysis using a gravity-driven stormwater system with an automated valve. Consider a two-tank stormwater system evolving on a finite-time horizon $t = 0, 1, \ldots, T-1$ using a deterministic Markov policy $\pi = (\pi_0, \pi_1, \ldots, \pi_{T-1})$, $X_{t+1} = X_t + \bar{f}(X_t, \pi_t(X_t), D_t) \cdot \triangle\tau$. Let $\mathbb{R}_+^n := \{y = (y_1, \ldots, y_n)^\text{T} \in \mathbb{R}^n : y_i \geq 0\ \forall i\}$. The state X_t is the $\mathbb{R}_+^2$-valued random water elevations in the tanks at time t (ft, ft). $\pi_t(X_t)$ is the [0,1]-valued valve setting at time t (closed to open). $(D_0, D_1, \ldots, D_{T-1})$ is a $\mathbb{R}_+$-valued, iid stochastic process of surface runoff. $\triangle\tau$ is the duration of $[t, t+1)$. The function

[8] Higher-quality approximations are those in which the estimates of the underapproximations are generally closer to the estimates of the risk-sensitive safe sets when considering all three disturbance distributions. We suggest an approach to quantify the quality of the approximations in Fig. 4.

[9] Recall that $\pi_\gamma \in \Pi$ is a policy that satisfies $J_\gamma^*(x) = J_\gamma(x, \pi_\gamma)\ \forall x \in S$. That is, π_γ is an optimal policy for the MDP problem that defines $\mathcal{U}_{\alpha,\gamma}^r$.

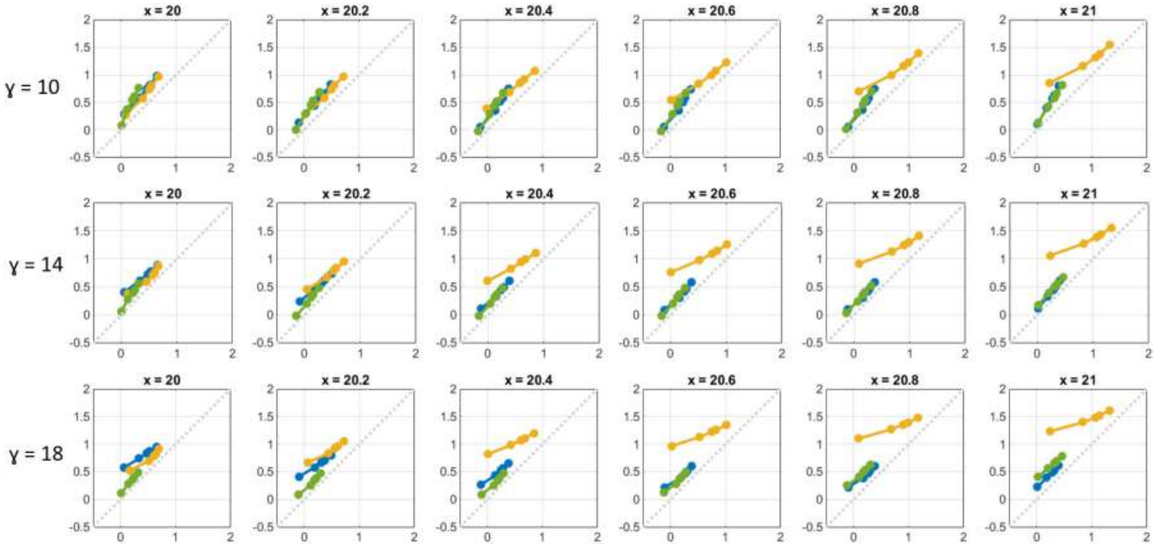

Fig. 3. Computations of the inequality that we have proved in Theorem 1 are shown for the temperature system. In each plot, the horizontal axis provides estimates of $\text{CVaR}_{\alpha,x}^{\pi_\gamma}(\max_{t=0,1,\ldots,T} g_K(X_t))$, and the vertical axis provides estimates of $\frac{1}{\gamma}\log(\frac{1}{\alpha}E_x^{\pi_\gamma}(\sum_{t=0}^{T} e^{\gamma g_K(X_t)}))$ for five different risk-sensitivity levels $\alpha \in \Lambda := \{0.99, 0.05, 0.01, 0.005, 0.001\}$. Points associated with smaller values of α (more risk-averse) are positioned farther away from the origin. For a fixed γ, π_γ is an optimal (deterministic, Markov) policy for the MDP problem (23). In each plot, there are three solid colored lines, one for each distribution of the disturbance process (green = no skew, yellow = left skew, blue = right skew). In each plot, $\gamma \in \{10, 14, 18\}$ and an initial state $x \in \{20, 20.2, \ldots, 21\}$ are fixed. The value of γ varies along the rows, and the value of x varies along the columns. A dotted gray line of slope 1 is shown for visual comparison.

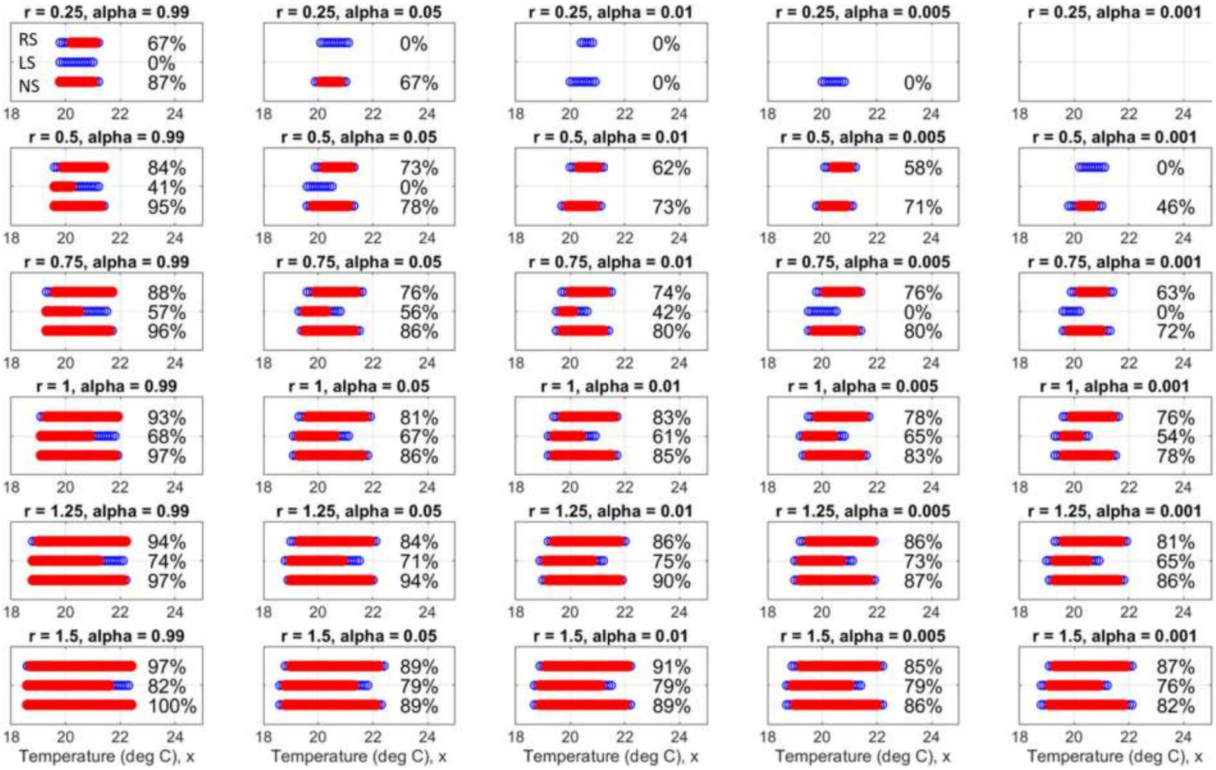

Fig. 4. For the temperature system with $\gamma = 14$, estimates of the (α, r, γ)-underapproximation set $\mathcal{U}_{\alpha,\gamma}^{r} = \mathcal{U}_{\alpha,\gamma}^{r,\pi_\gamma}$ are shown (solid red circles). Estimates of the (α, r)-risk-sensitive safe set for the control policy $\pi_\gamma \in \Pi$, $\mathcal{S}_{\alpha}^{r,\pi_\gamma}$, are shown (white circles with blue boundary). Each plot presents the estimated sets for the different disturbance distributions [top interval: right skew (RS), middle interval: left skew (LS), and bottom interval: no skew (NS); see the labels in the first plot]. Each percentage $\frac{\text{Number of states in estimate of } \mathcal{U}_{\alpha,\gamma}^{r}}{\text{Number of states in estimate of } \mathcal{S}_{\alpha}^{r,\pi_\gamma}} \cdot 100\%$ indicates the estimated quality of the underapproximation. These percentages are shown whenever the estimate of $\mathcal{S}_{\alpha}^{r,\pi_\gamma}$ is not empty. The risk-sensitivity level α varies from nearly risk-neutral ($\alpha = 0.99$, left-most column) to more risk-averse ($\alpha = 0.001$, right-most column).

TABLE II
STORMWATER SYSTEM PARAMETERS

Symbol	Description	Value
a_1	surface area of tank 1	28292 ft^2
a_2	surface area of tank 2	25965 ft^2
c_d	discharge coefficient	0.61 (no units)
$\bar{g}$	acceleration due to gravity	32.2 $\frac{\text{ft}}{\text{s}^2}$
k_1	maximum water level in tank 1	3.5 ft
k_2	maximum water level in tank 2	5 ft
$\bar{\pi}$	circle circumference-to-diameter ratio	≈ 3.14
r_{d}	radius of drain	2/3 ft
r_{v}	radius of valve	1/3 ft
$\triangle\tau$	duration of $[t, t+1)$	5 min
T	length of discrete time horizon	24 (= 2 h)
A	control space	$[0, 1]$ (no units)
S	state space	$[0, 5]$ ft $\times$ $[0, 6.5]$ ft
z_1	invert elevation of pipe from base of tank 1	1 ft
$z_{1,\text{in}}$	invert elevation of pipe from base of tank 2	2.5 ft
z_2	elevation from base of tank 2 to orifice	1 ft

ft = feet, s = seconds, min = minutes, h = hours.

TABLE III

	$\alpha = 0.99$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.001$
$r = 0.5$ ft	74.3 %	77.3 %	76.6 %	76.0 %	72.5 %
$r = 1$ ft	82.9 %	84.2 %	83.1 %	83.0 %	81.2 %
$r = 1.5$ ft	84.4 %	75.6 %	71.2 %	69.9 %	66.0 %

This table provides the percentages $\frac{\text{Number of states in estimate of } \mathcal{U}^r_{\alpha,\gamma}}{\text{Number of states in estimate of } \mathcal{S}^{r,\pi_\gamma}_\alpha} \cdot 100\%$ for the sets in Fig. 5 (stormwater system, $\gamma = 22$).

$\bar{f} : \mathbb{R}^2_+ \times [0, 1] \times \mathbb{R}_+ \to \mathbb{R}^2$ is given by

$$\bar{f}(x, u, d) := \left[\frac{d - q_{\text{valve}}(x, u)}{a_1}, \frac{d + q_{\text{valve}}(x, u) - q_{\text{drain}}(x)}{a_2} \right]^T$$

$$q_{\text{valve}}(x, u) := u \cdot \bar{\pi} r_{\text{v}}^2 \cdot \text{sgn}(h(x)) \cdot \sqrt{2\bar{g}|h(x)|}$$

$$h(x) := \max(x_1 - z_1, 0) - \max(x_2 - z_{1,\text{in}}, 0)$$

$$q_{\text{drain}}(x) := \begin{cases} c_{\text{d}} \bar{\pi} r_{\text{d}}^2 \sqrt{2\bar{g}(x_2 - z_2)} & \text{if } x_2 \geq z_2 \\ 0 & \text{otherwise.} \end{cases}$$

Model parameters are in Table II. The constraint set $K = [0, k_1] \times [0, k_2]$ specifies the maximum water elevations that the tanks can hold without surcharge. The stage cost $g_K(x) = \max(x_1 - k_1, x_2 - k_2, 0)$ is the maximum surcharged water level when the system occupies the state $x \in \mathbb{R}^2_+$.

We have identified a discrete distribution for the disturbance process with the approximate statistics, mean (12.2 cfs), variance (9.9 cfs^2), and skew (0.74), where cfs is cubic feet per second. In previous work, we obtained runoff samples by simulating a design storm in PCSWMM (Computational Hydraulics International), which extends the US Environmental Protection Agency's Stormwater Management Model [52], [53]. In this previous work, the empirical distribution had positive skew, and the mean was about 12.2 cfs [52], which are reflected in the current distribution (not shown in the interest of space).

In Fig. 5, we show estimates of risk-sensitive safe sets and their underapproximations using $\gamma = 22$ for five risk-sensitivity levels (see also Table III). The shape of the contour of $\mathcal{S}^{r,\pi_\gamma}_\alpha$ indicates a critical tradeoff between the maximum initial water elevations in the two tanks from which the system meets a desired degree of safety. The similarity in the shapes of $\mathcal{S}^{r,\pi_\gamma}_\alpha$ and $\mathcal{U}^r_{\alpha,\gamma}$ is notable, suggesting that $\mathcal{U}^r_{\alpha,\gamma}$ may be a useful tool for inferring these critical tradeoffs in networked water systems.

VI. CONCLUDING REMARKS

This article develops trajectory-wise safety specifications for control systems that quantify the severity of random harmful outcomes and thereby generalize classical stochastic safety analysis. Our primary contribution is to develop a tractable, interpretable safety analysis method with theoretical guarantees that assesses the upper tail of a cost distribution by using CVaR. It is notable that our method provides a parameter-dependent upper bound to the CVaR of a maximum cost without augmenting the state space. We have developed compelling numerical examples, which demonstrate the utility and tractability of our underapproximation approach. Moreover, we have proposed a risk-sensitive safe set definition in terms of a new coherent risk functional, inspired by CVaR, that admits a parameter-independent upper bound. We show that this upper bound can be computed via DP on the original state space by proving the regularity of a supremum over a function space for a class of transition kernels. Numerical investigations of leveraging our approximation to provide an efficient preliminary estimate to the exact CVaR is an exciting future direction. For instance, we have recently demonstrated the usefulness of efficient approximate "warm-start" computations to examine the effect of different design changes to stormwater infrastructure [61]. More broadly, combining techniques from approximate DP, stochastic rollout, and risk-sensitive safety analysis could lead to novel controller synthesis algorithms for higher-dimensional systems.

APPENDIX

Lemma 6 (Attainment of Supremum): Let $J : S \to \mathbb{R}$ be Borel measurable and bounded, and let $\alpha \in (0, 1]$. Define the function $\phi : S \times A \to \mathbb{R}$ by

$$\phi(x, u) := \sup \left\{ \int_S J\xi \, \mathrm{d}Q(\cdot|x, u) : \xi \in \mathcal{R}_\alpha(x, u) \right\}. \quad (45)$$

Then, for any $(x, u) \in S \times A$, there is a $\xi^*(\cdot|x, u) \in \mathcal{R}_\alpha(x, u)$ such that

$$\phi(x, u) = \int_S J\xi^*(\cdot|x, u) \, \mathrm{d}Q(\cdot|x, u). \quad (46)$$

Proof: Let $(x, u) \in S \times A$, and fix the probability space $(S, \mathcal{B}(S), Q(\cdot|x, u))$. Denote $L^p_{x,u} := L^p(S, \mathcal{B}(S), Q(\cdot|x, u))$ for brevity, and view $\mathcal{R}_\alpha(x, u)$ as a subset of $L^2_{x,u}$ with the weak topology. Define the functional $\psi : L^2_{x,u} \to \mathbb{R}$ by

$$\psi(\xi) := \int_S J\xi \, \mathrm{d}Q(\cdot|x, u). \quad (47)$$

It suffices to show that ψ is weakly continuous and $\mathcal{R}_\alpha(x, u)$ is weakly compact. Weak continuity follows from two well-known facts: 1) A linear functional on a normed vector space is weakly continuous if and only if it is strongly continuous [54, Prop. 2.5.3]; and 2) a linear functional on a normed vector space is strongly continuous if and only if it is bounded [12, Prop. 5.2]. By applying standard techniques, it follows that ψ is a bounded linear functional on a normed vector space, and thus, ψ is weakly continuous. As $\mathcal{R}_\alpha(x, u)$ is a bounded and weakly closed subset

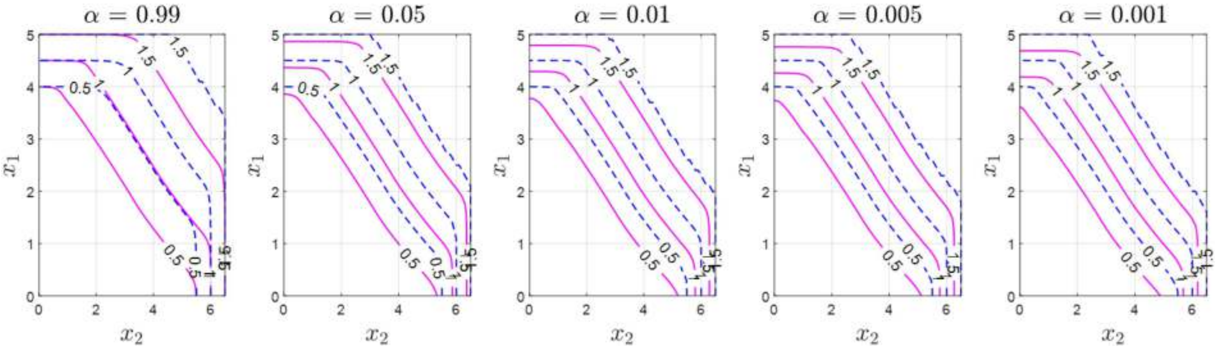

Fig. 5. For the stormwater system with $\gamma = 22$, estimates of the boundary of the (α, r, γ)-underapproximation set $\mathcal{U}^r_{\alpha,\gamma} = \mathcal{U}^{r,\pi_\gamma}_{\alpha,\gamma}$ are shown (solid pink). Estimates of the boundary of the (α, r)-risk-sensitive safe set for the control policy $\pi_\gamma \in \Pi$, $\mathcal{S}^{r,\pi_\gamma}_\alpha$, are shown (dotted blue). We present $r \in \{0.5, 1, 1.5\}$ and $\alpha \in \{0.99, 0.05, 0.01, 0.005, 0.001\}$. The percentages $\frac{\text{Number of states in estimate of } \mathcal{U}^r_{\alpha,\gamma}}{\text{Number of states in estimate of } \mathcal{S}^{r,\pi_\gamma}_\alpha} \cdot 100\%$ indicate the estimated quality of the underapproximations. We list these percentages for the plots in this figure in Table III.

of $L^2_{x,u}$, $\mathcal{R}_\alpha(x,u)$ is weakly compact by the Banach-Alaoglu Theorem [37, p. 401]. Here, we use the fact that $L^2_{x,u}$ is reflexive, and hence, the weak and weak* topologies of $L^2_{x,u}$ are the same. We provide details about weak closedness in a footnote.[10] ■

We use similar techniques to prove Lemma 7. Lemma 7 is needed to guarantee that a supremum over $\mathcal{R}_\alpha(x,u)$ (45) is usc in (x,u).

Lemma 7 (Existence of weakly convergent subsequence): Let μ be a probability measure on $(S, \mathcal{B}(S))$, $\mathcal{G}_\alpha(\mu)$ the set of functions $\xi \in L^2_\mu := L^2(S, \mathcal{B}(S), \mu)$ such that $\xi \in [0, \alpha^{-1/T}]$ a.e. w.r.t. μ, and $(\xi_n)_{n\in\mathbb{N}} \subseteq \mathcal{G}_\alpha(\mu)$. Then, there exist $(\xi_{n_k})_{k\in\mathbb{N}} \subseteq (\xi_n)_{n\in\mathbb{N}}$ and $\xi^* \in \mathcal{G}_\alpha(\mu)$ such that $(\xi_{n_k})_{k\in\mathbb{N}}$ converges to ξ^* in the weak topology of L^2_μ.

Proof: The proof requires two facts. The first fact is [51, Th. 3.18]: *Assume that E is a reflexive Banach space, and let (x_n) be a (uniformly) bounded sequence in E. Then, there is a subsequence $(x_{n_k}) \subseteq (x_n)$ that converges in the weak topology.* L^2_μ is a reflexive Banach space, and $||\xi_n||_{L^2_\mu} \leq \alpha^{-1/T}$ for all $n \in \mathbb{N}$. Thus, there exist $(\xi_{n_k})_{k\in\mathbb{N}} \subseteq (\xi_n)_{n\in\mathbb{N}}$ and $\xi^* \in L^2_\mu$ such that $(\xi_{n_k})_{k\in\mathbb{N}}$ converges weakly to ξ^*. Moreover, it holds that $\xi^* \in \mathcal{G}_\alpha(\mu)$ using [51, Th. 3.7] (Footnote 10). Indeed, $\mathcal{G}_\alpha(\mu)$ is a convex subset of L^2_μ, and $\mathcal{G}_\alpha(\mu)$ is strongly closed in L^2_μ. Thus, $\mathcal{G}_\alpha(\mu)$ is weakly closed in L^2_μ, which implies that $\xi^* \in \mathcal{G}_\alpha(\mu)$. ■

We use Lemma 7 to prove the next supporting result.

Lemma 8 (Properties of ϕ): Let $J : S \to \mathbb{R}$ be Borel measurable and bounded and $\alpha \in (0,1]$. Under Assumption 2, ϕ (45) is usc and bounded.

Proof: Boundedness of ϕ follows from $Q(\cdot|x,u)$-a.e.-boundedness of $J\xi$ for any $\xi \in \mathcal{R}_\alpha(x,u)$. Now, ϕ is usc if and only if

$$\mathcal{C}_a := \{(x,u) \in S \times A : \phi(x,u) \geq a\}$$

is closed for every $a \in \mathbb{R}$. Let $a \in \mathbb{R}$ and $(x_n, u_n)_{n\in\mathbb{N}} \subseteq \mathcal{C}_a$ converging to $(x,u) \in S \times A$ be given, and we shall show that $(x,u) \in \mathcal{C}_a$. It suffices to show that there exist $(x_{n_k}, u_{n_k})_{k\in\mathbb{N}} \subseteq (x_n, u_n)_{n\in\mathbb{N}}$ and $(c_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$ with $c_k \to 0$ such that

$$\phi(x_{n_k}, u_{n_k}) \leq c_k + \phi(x,u) \quad \forall k \in \mathbb{N}.$$

Indeed, if so, then

$$a \leq \limsup_{k\to\infty} \phi(x_{n_k}, u_{n_k}) \leq \limsup_{k\to\infty} c_k + \phi(x,u) = \phi(x,u).$$

Denote $z_n := (x_n, u_n)$ and $z := (x,u)$ for brevity. By Lemma 6, for every $n \in \mathbb{N}$,

$$\exists \xi_n := \xi^*(\cdot|z_n) \in \mathcal{R}_\alpha(z_n) \text{ s.t. } \phi(z_n) = \int_S J\xi_n \, \mathrm{d}Q(\cdot|z_n).$$

Since $\xi_n \in [0, \alpha^{-1/T}]$ a.e. w.r.t. $Q(\cdot|z_n)$,

$$\exists B(z_n) \in \mathcal{B}(S) \text{ s.t. } \xi_n(y) \in [0, \alpha^{-1/T}] \quad \forall y \in B(z_n)$$

where $Q(S \setminus B(z_n)|z_n) = 0$. Define

$$\tilde{\xi}_n := I_{B(z_n)}\xi_n.$$

It follows that $\tilde{\xi}_n \in \mathcal{R}_\alpha(z_n)$ with $\tilde{\xi}_n \in [0, \alpha^{-1/T}]$ everywhere. Also, it holds that $(\tilde{\xi}_n)_{n\in\mathbb{N}} \subseteq \mathcal{G}_\alpha(Q(\cdot|z))$, where

$$\mathcal{G}_\alpha(Q(\cdot|z)) := \left\{\xi \in L^2_z : \xi \in [0, \alpha^{-1/T}] \text{ a.e. w.r.t. } Q(\cdot|z)\right\}$$

and $L^2_z := L^2(S, \mathcal{B}(S), Q(\cdot|z))$. By Lemma 7, there exist $(\tilde{\xi}_{n_k})_{k\in\mathbb{N}} \subseteq (\tilde{\xi}_n)_{n\in\mathbb{N}}$ and $\xi^\dagger \in \mathcal{G}_\alpha(Q(\cdot|z))$ such that $(\tilde{\xi}_{n_k})_{k\in\mathbb{N}}$ converges to $\xi^\dagger$ in the weak topology of L^2_z. It holds that $\xi^\dagger \in \mathcal{R}_\alpha(z)$, and we explain why $\int_S \xi^\dagger \mathrm{d}Q(\cdot|z) = 1$ next. For any $k \in \mathbb{N}$, it holds that $|\tilde{\xi}_{n_k}| \leq \alpha^{-1/T}$ everywhere, and it follows that

$$\left|\int_S \xi^\dagger \mathrm{d}Q(\cdot|z) - 1\right| \leq \text{Term1}(k) + \text{Term2}(k)$$

[10] For weak closedness, recall the fact [51, Th. 3.7]: *Let E be a Banach space, and let C be a convex subset of E. Then, C is closed in the weak topology if and only if it is closed in the strong topology.* Since $\mathcal{R}_\alpha(x,u) \subseteq L^2_{x,u}$ is convex, to show that $\mathcal{R}_\alpha(x,u)$ is weakly closed, it suffices to show that $\mathcal{R}_\alpha(x,u)$ is strongly closed. Strong closedness of $\mathcal{R}_\alpha(x,u)$ follows from 1) strong convergence implying weak convergence and 2) strong convergence implying the existence of a subsequence that converges a.e. to the same limit function [46, Ths. 2.5.1 & 2.5.3]. Let $(\xi_n)_{n\in\mathbb{N}} \subseteq \mathcal{R}_\alpha(x,u)$ converge strongly to $\xi^* \in L^2_{x,u}$. The first fact ensures that $\int_S \xi^* \mathrm{d}Q(\cdot|x,u) = 1$, and the second fact ensures that $0 \leq \xi^* \leq \alpha^{-1/T}$ a.e., and thus, $\xi^* \in \mathcal{R}_\alpha(x,u)$.

where

$$\text{Term1}(k) := \left| \int_S \xi^\dagger \mathrm{d}Q(\cdot|z) - \int_S \tilde{\xi}_{n_k} \mathrm{d}Q(\cdot|z) \right|$$

$$\text{Term2}(k) := \alpha^{-1/T} \big|Q(\cdot|z) - Q(\cdot|z_{n_k})\big|(S).$$

The quantity $|Q(\cdot|z) - Q(\cdot|z_{n_k})|(S)$ is the total variation of the signed measure $Q(\cdot|z) - Q(\cdot|z_{n_k})$ evaluated at the set S. By the weak convergence of $(\tilde{\xi}_{n_k})_{k\in\mathbb{N}}$ to $\xi^\dagger$ in L^2_z, $\text{Term1}(k) \to 0$ as $k \to \infty$. By Assumption 2, $\text{Term2}(k) \to 0$ as $k \to \infty$.

Now, for every $k \in \mathbb{N}$, we use the triangle inequality and everywhere boundedness of $J\tilde{\xi}_{n_k}$ to find that

$$\phi(z_{n_k}) - \phi(z) \leq \text{Term3}(k) + \text{Term4}(k)$$

where

$$\text{Term3}(k) := \frac{b}{\alpha^{1/T}} \big|Q(\cdot|z_{n_k}) - Q(\cdot|z)\big|(S),$$

$b \in \mathbb{R}$ satisfies $|J(y)| \leq b$ for all $y \in S$, and

$$\text{Term4}(k) := \left| \int_S J\tilde{\xi}_{n_k} \mathrm{d}Q(\cdot|z) - \int_S J\xi^\dagger \mathrm{d}Q(\cdot|z) \right|.$$

By the weak convergence of $(\tilde{\xi}_{n_k})_{k\in\mathbb{N}}$ to $\xi^\dagger$ in L^2_z, $\text{Term4}(k) \to 0$ as $k \to \infty$, and by Assumption 2, $\text{Term3}(k) \to 0$ as $k \to \infty$. We choose

$$c_k := \text{Term3}(k) + \text{Term4}(k) \quad \forall k \in \mathbb{N}$$

and it follows that ϕ is usc. ■

We use the upper semi-continuity of ϕ to prove Theorem 3.

Proof of Theorem 3: Proceed by induction. $J^\alpha_T = c_T$ is usc and bounded. Now, assume that for some $t = T-1, \ldots, 1, 0$, J^α_{t+1} is usc and bounded. Then, J^α_{t+1} is Borel measurable and bounded, which implies that φ^α_t is usc and bounded by Lemma 8. Since $v^\alpha_t = c_t + \varphi^\alpha_t$ is a sum of usc and bounded functions, v^α_t is usc and bounded. By [15, Prop. 7.34], we conclude that J^α_t is usc and bounded, and for every $\epsilon > 0$, there is a Borel-measurable function $\kappa^{\alpha,\epsilon}_t : S \to A$ such that $J^\alpha_t(x) \leq v^\alpha_t(x, \kappa^{\alpha,\epsilon}_t(x)) \leq J^\alpha_t(x) + \epsilon$ for all $x \in S$.

A DP argument completes the proof, which we outline below.[11] Let Π' be the set of randomized Markov policies. For $t = 0, 1, \ldots, T$, define the random cost-to-go by

$$Y_t := \begin{cases} c_T(X_T) + \sum_{i=t}^{T-1} c_i(X_i, U_i) & \text{if } t < T \\ c_T(X_T) & \text{if } t = T \end{cases}$$

and note that $Y = Y_0$. For any $\pi \in \Pi'$ and $\zeta \in \mathcal{D}_\alpha$, we denote the (π, ζ)-conditional expectation of Y_t given X_t by $W^{\pi,\zeta}_t(x_t) := E^{\pi,\zeta}(Y_t | X_t = x_t)$, where $x_t \in S$. For any $\pi = (\pi_0, \pi_1, \ldots, \pi_{T-1}) \in \Pi'$ and $\zeta = (\xi_0, \xi_1, \ldots, \xi_{T-1}) \in \mathcal{D}_\alpha$, the following recursion ("law of iterated expectations") holds: for $t = 0, 1, \ldots, T-1$ and $x \in S$,

$$W^{\pi,\zeta}_t(x) = \int_A \Big(c_t(x,u) + \psi^{\pi,\zeta}_t(x,u) \Big) \; \pi_t(\mathrm{d}u|x) \tag{48a}$$

where $\psi^{\pi,\zeta}_t$ is defined by

$$\psi^{\pi,\zeta}_t(x,u) := \int_S W^{\pi,\zeta}_{t+1}(y) \; \xi_t(y|x,u) \; Q(\mathrm{d}y|x,u) \tag{48b}$$

with $\xi_t(\cdot|x,u) \in \mathcal{R}_\alpha(x,u)$ for each $(x,u) \in S \times A$. For any policy $\pi \in \Pi'$, we have

$$\rho^\pi_{\alpha,x}(Y) = \sup_{\zeta \in \mathcal{D}_\alpha} W^{\pi,\zeta}_0(x) \quad \forall x \in S. \tag{49}$$

Let $\epsilon > 0$ be given. Then, for each $t = 0, 1, \ldots, T-1$, there exists a Borel-measurable function $\mu^{\alpha,\epsilon}_t : S \to A$ such that

$$J^\alpha_t(x) \leq v^\alpha_t(x, \mu^{\alpha,\epsilon}_t(x)) \leq J^\alpha_t(x) + \frac{\epsilon}{T} \quad \forall x \in S. \tag{50}$$

Define $\pi^*_\epsilon := (\mu^{\alpha,\epsilon}_0, \ldots, \mu^{\alpha,\epsilon}_{T-1}) \in \Pi'$, which is a deterministic Markov policy, and thus, is an element of Π (the class of randomized history-dependent policies) as well. Hence,

$$\inf_{\pi \in \Pi} \rho^\pi_{\alpha,x}(Y) \leq \rho^{\pi^*_\epsilon}_{\alpha,x}(Y) \overset{(49)}{=} \sup_{\zeta \in \mathcal{D}_\alpha} W^{\pi^*_\epsilon,\zeta}_0(x) \quad \forall x \in S.$$

It suffices to prove that

$$W^{\pi^*_\epsilon,\zeta}_t(x) \leq J^\alpha_t(x) + \frac{(T-t)\epsilon}{T} \tag{51}$$

for all $x \in S$, $\zeta \in \mathcal{D}_\alpha$, and $t \in \{0, 1, \ldots, T\}$. Indeed, by setting $t = 0$ in (51) and taking the supremum over $\mathcal{D}_\alpha$, we would derive $\rho^{\pi^*_\epsilon}_{\alpha,x}(Y) \leq J^\alpha_0(x) + \epsilon \; \forall x \in S$. Since $\epsilon > 0$ is arbitrary, the desired statement would be shown. The sufficient condition (51) holds by an inductive argument.[12] ■

Acknowledgment

The authors extend their sincerest gratitude to the four anonymous reviewers whose feedback was invaluable in improving the rigor and presentation of this work. The authors gratefully acknowledge many discussions with Jonathan Lacotte, Yuxi Han, Zhiyan Ding, and Michael Lim during the development of an earlier version of this article. The authors are also thankful for the advice provided by Serdar Yüksel, Nicole Bäuerle, Ludovic Tangpi, Michael Fauß, Alois Pichler, and Raymond Kwong.

[11] We write that a relation with a conditional expectation holds everywhere for simplicity, following [45, Th. 3.2.1].

[12] The base case holds because $W^{\pi^*_\epsilon,\zeta}_T = c_T = J^\alpha_T$ for all $\zeta \in \mathcal{D}_\alpha$. Now, assume that for some $t \in \{0, 1, \ldots, T-1\}$, it holds that $W^{\pi^*_\epsilon,\zeta}_{t+1}(x) \leq J^\alpha_{t+1}(x) + \frac{(T-t-1)\epsilon}{T}$ for all $x \in S$ and $\zeta \in \mathcal{D}_\alpha$. Let $x \in S$ and $\zeta = (\xi_0, \xi_1, \ldots, \xi_{T-1}) \in \mathcal{D}_\alpha$ be given. Since π^*_ϵ is a deterministic Markov policy, we have

$$W^{\pi^*_\epsilon,\zeta}_t(x) \overset{(48)}{=} c_t(x, \mu^{\alpha,\epsilon}_t(x)) + \psi^{\pi^*_\epsilon,\zeta}_t(x, \mu^{\alpha,\epsilon}_t(x)).$$

By the induction hypothesis and $\xi_t(\cdot|x, \mu^{\alpha,\epsilon}_t(x)) \in \mathcal{R}_\alpha(x, \mu^{\alpha,\epsilon}_t(x))$ from the definition of $\mathcal{D}_\alpha$, it follows that

$$\psi^{\pi^*_\epsilon,\zeta}_t(x, \mu^{\alpha,\epsilon}_t(x)) \leq \varphi^\alpha_t(x, \mu^{\alpha,\epsilon}_t(x)) + \frac{(T-t-1)\epsilon}{T}.$$

Since $v^\alpha_t = c_t + \varphi^\alpha_t$, we derive $W^{\pi^*_\epsilon,\zeta}_t(x) \leq v^\alpha_t(x, \mu^{\alpha,\epsilon}_t(x)) + \frac{(T-t-1)\epsilon}{T}$. Then, we complete the induction using the second inequality in (50), namely $v^\alpha_t(x, \mu^{\alpha,\epsilon}_t(x)) \leq J^\alpha_t(x) + \frac{\epsilon}{T}$.

References

[1] R. A. Howard and J. E. Matheson, "Risk-sensitive Markov decision processes," *Manage. Sci.*, vol. 18, no. 7, pp. 356–369, 1972.

[2] D. H. Jacobson, "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games," *IEEE Trans. Autom. Control*, vol. AC-18, no. 2, pp. 124–131, Apr. 1973.

[3] G. B. di Masi and L. Stettner, "Risk-sensitive control of discrete-time Markov processes with infinite horizon," *SIAM J. Control Optim.*, vol. 38, no. 1, pp. 61–78, 1999.

[4] P. Whittle, "Risk-sensitive linear/quadratic/Gaussian control," *Adv. Appl. Probability*, vol. 13, no. 4, pp. 764–777, 1981.

[5] A. Shapiro, "On a time consistency concept in risk averse multistage stochastic programming," *Operations Res. Lett.*, vol. 37, no. 3, pp. 143–147, 2009.

[6] K. Boda and J. A. Filar, "Time consistent dynamic risk measures," *Math. Methods Oper. Res.*, vol. 63, no. 1, pp. 169–186, 2006.

[7] D. Kreps, "Decision problems with expected utility critera, I: Upper and lower convergent utility," *Math. Oper. Res.*, vol. 2, no. 1, pp. 45–53, 1977.

[8] T. Bjork and A. Murgoci, "A theory of Markovian time-inconsistent stochastic control in discrete time," *Finance Stochastics*, vol. 18, pp. 545–592, 2014.

[9] L. Xin and A. Shapiro, "Bounds for nested law invariant coherent risk measures," *Oper. Res. Lett.*, vol. 40, no. 6, pp. 431–435, 2012.

[10] A. Shapiro, "Minimax and risk averse multistage stochastic programming," *Eur. J. Oper. Res.*, vol. 219, no. 3, pp. 719–726, 2012.

[11] P. Artzner *et al.*, "Coherent measures of risk," *Math. Finance*, vol. 9, no. 3, pp. 203–228, 1999.

[12] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. New York, NY, USA: Wiley, 1999.

[13] H. Föllmer and A. Schied, *Stochastic Finance: An Introduction in Discrete Time*, 2nd ed. Berlin, New York, NY: de Gruyter Studies in Mathematics, 2004.

[14] A. Ruszczyński and A. Shapiro, "Optimization of convex risk functions," *Math. Oper. Res.*, vol. 31, no. 3, pp. 433–452, 2006.

[15] D. P. Bertsekas and S. Shreve, *Stochastic Optimal Control: The Discrete-Time Case*. Belmont, MA, USA: Athena Scientific, 1996.

[16] N. Bäuerle and J. Ott, "Markov decision processes with average-value-at-risk criteria," *Math. Methods Oper. Res.*, vol. 74, no. 3, pp. 361–379, 2011.

[17] N. Bäuerle and U. Rieder, "More risk-sensitive Markov decision processes," *Math. Oper. Res.*, vol. 39, no. 1, pp. 105–120, 2014.

[18] B. Rudloff, A. Street, and D. M. Valladao, "Time consistency and risk averse dynamic decision models: Definition, interpretation and practical consequences," *Eur. J. Oper. Res.*, vol. 234, no. 3, pp. 743–750, 2014.

[19] P. A. Forsyth, "Multiperiod mean conditional value at risk asset allocation: Is it advantageous to be time consistent?" *SIAM J. Financial Math.*, vol. 11, no. 2, pp. 358–384, 2020.

[20] D. P. Bertsekas and I. B. Rhodes, "On the minimax reachability of target sets and target tubes," *Automatica*, vol. 7, no. 2, pp. 233–247, 1971.

[21] I. M. Mitchell, A. M. Bayen, and C. J. Tomlin, "A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games," *IEEE Trans. Autom. Control*, vol. 50, no. 7, pp. 947–957, Jul. 2005.

[22] K. Margellos and J. Lygeros, "Hamilton-Jacobi formulation for reach-avoid differential games," *IEEE Trans. Autom. Control*, vol. 56, no. 8, pp. 1849–1861, Aug. 2011.

[23] M. Chen, S. L. Herbert, M. S. Vashishtha, S. Bansal, and C. J. Tomlin, "Decomposition of reachable sets and tubes for a class of nonlinear systems," *IEEE Trans. Autom. Control*, vol. 63, no. 11, pp. 3675–3688, Nov. 2018.

[24] M. Chen and C. J. Tomlin, "Hamilton-Jacobi reachability: Some recent theoretical advances and applications in unmanned airspace management," *Annu. Rev. Control, Robot., Auton. Syst.*, vol. 1, no. 1, pp. 333–358, 2018.

[25] A. Abate, M. Prandini, J. Lygeros, and S. Sastry, "Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems," *Automatica*, vol. 44, no. 11, pp. 2724–2734, 2008.

[26] S. Summers and J. Lygeros, "Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem," *Automatica*, vol. 46, no. 12, pp. 1951–1961, 2010.

[27] M. Kamgarpour, J. Ding, S. Summers, A. Abate, J. Lygeros, and C. Tomlin, "Discrete time stochastic hybrid dynamical games: Verification and controller synthesis," in *Proc. IEEE Conf. Decis. Control*, 2011, pp. 6122–6127.

[28] J. Ding, M. Kamgarpour, S. Summers, A. Abate, J. Lygeros, and C. Tomlin, "A stochastic games framework for verification and control of discrete time stochastic hybrid systems," *Automatica*, vol. 49, pp. 2665–2674, 2013.

[29] I. Yang, "A dynamic game approach to distributionally robust safety specifications for stochastic systems," *Automatica*, vol. 94, pp. 94–101, 2018.

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

[31] Y. Chow, A. Tamar, S. Mannor, and M. Pavone, "Risk-sensitive and robust decision-making: A CVaR optimization approach," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1522–1530.

[32] J. F. Bonnans, P. Lavigne, and L. Pfeiffer, "Discrete-time mean field games with risk-averse agents," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 27, no. 44, pp. 1–27, 2021.

[33] V. Borkar and R. Jain, "Risk-constrained Markov decision processes," *IEEE Trans. Autom. Control*, vol. 59, no. 9, pp. 2574–2579, Sep. 2014.

[34] W. B. Haskell and R. Jain, "A convex analytic approach to risk-aware Markov decision processes," *SIAM J. Control Optim.*, vol. 53, no. 3, pp. 1569–1598, 2015.

[35] P. Whittle, *Risk-Sensitive Optimal Control*, Hoboken, NJ, USA: Wiley, 1990.

[36] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *J. Bank. Finance*, vol. 26, no. 7, pp. 1443–1471, 2002.

[37] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia, PA, USA: SIAM, 2009.

[38] M. P. Chapman *et al.*, "A risk-sensitive finite-time reachability approach for safety of stochastic dynamic systems," in *Proc. Amer. Control Conf.*, 2019, pp. 2958–2963.

[39] C. W. Miller and I. Yang, "Optimal control of conditional value-at-risk in continuous time," *SIAM J. Control Optim.*, vol. 55, no. 2, pp. 856–884, 2017.

[40] G. C. Pflug and A. Pichler, "Time-inconsistent multistage stochastic programs: Martingale bounds," *Eur. J. Oper. Res.*, vol. 249, no. 1, pp. 155–163, 2016.

[41] G. C. Pflug and A. Pichler, "Time-consistent decisions and temporal decomposition of coherent risk functionals," *Math. Oper. Res.*, vol. 41, no. 2, pp. 682–699, 2016.

[42] B. P. G. Van Parys, D. Kuhn, P. J. Goulart, and M. Morari, "Distributionally robust control of constrained stochastic systems," *IEEE Trans. Autom. Control*, vol. 61, no. 2, pp. 430–442, Feb. 2016.

[43] A. Akametalu, "A learning-based approach to safety for uncertain robotic systems," Ph.D. dissertation, Electrical Engineering and Computer Sciences, Univ. California, Berkeley, Berkeley, CA, USA, 2018.

[44] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[45] O. Hernández-Lerma and J. B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. New York, NY, USA: Springer, 1996.

[46] R. B. Ash, *Real Analysis and Probability*. New York, NY, USA: Academic Press, 1972.

[47] A. Ruszczyński, "Risk-averse dynamic programming for Markov decision processes," *Math. Program.*, vol. 125, no. 2, pp. 235–261, 2010.

[48] R. E. Mortensen and K. P. Haggerty, "A stochastic computer model for heating and cooling loads," *IEEE Trans. Power Syst.*, vol. 3, no. 3, pp. 1213–1219, Aug. 1988.

[49] P. M. Esfahani *et al.*, "From infinite to finite programs: Explicit error bounds with applications to approximate dynamic programming," *SIAM J. Optim.*, vol. 28, no. 3, pp. 1968–1998, 2018.

[50] D. P. Bertsekas, *Reinforcement Learning and Optimal Control*. Belmont, MA, USA: Athena Scientific, 2019.

[51] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. New York, NY, USA: Springer, 2010.

[52] M. P. Chapman, K. M. Smith, V. Cheng, D. L. Freyberg, and C. Tomlin, "Reachability analysis as a design tool for stormwater systems," in *Proc. IEEE Conf. Technol. Sustain.*, 2018, pp. 1–8.

[53] L. A. Rossman, *Storm Water Management Model User's Manual, Version 5.0*. National Risk Management Research Laboratory, OH, USA: Office of Research and Development, US Environmental Protection Agency, Cincinnati, 2010.

[54] R. E. Megginson, *An Introduction to Banach Space Theory*. New York, NY, USA: Springer, 1998.

[55] M. Chen, S. Herbert, and C. J. Tomlin, "Exact and efficient Hamilton-Jacobi guaranteed safety analysis via system decomposition," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2017, pp. 87–92.

[56] S. Coogan and M. Arcak, “Efficient finite abstraction of mixed monotone systems,” in *Proc. 18th Int. Conf. Hybrid Systems: Comput. Control*, 2015, pp. 58–67.

[57] S. Coogan, M. Arcak, and C. Belta, “Formal methods for control of traffic flow: Automated control synthesis from finite-state transition models,” *IEEE Control Syst. Mag.*, vol. 37, no. 2, pp. 109–128, Apr. 2017.

[58] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, “Kalman filtering with intermittent observations,” *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.

[59] M. Lefebvre, “Optimally ending an epidemic,” *Optimization*, vol. 67, no. 3, pp. 399–407, 2018.

[60] M. Lefebvre, “Computer virus propagation modelled as a stochastic differential game,” *Atti della Accademia Peloritana dei Pericolanti-Classe di Scienze Fisiche, Matematiche e Naturali*, vol. 98, no. 1, pp. 1–8, 2020.

[61] M. P. Chapman, M. Fauß, and K. M. Smith, “On optimizing the conditional value-at-risk of a maximum cost for risk-averse safety analysis,” 2021, *arXiv:2106.00776*.

Margaret Chapman (Member, IEEE) received her B.S. and M.S. degrees in mechanical engineering from Stanford University, Stanford, CA, USA, in 2012 and 2014, respectively, and her Ph.D. degree in electrical engineering and computer sciences from the University of California Berkeley, Berkeley, CA, USA, in 2020.

She is currently an Assistant Professor with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, where she joined in July 2020. Her research interests include risk-averse and stochastic control, with emphasis on safety and applications to healthcare and sustainable cities.

Dr. Chapman is a recipient of the Leon O. Chua Award for achievement in nonlinear science from UC Berkeley (2021), the US National Science Foundation Graduate Research Fellowship (2014), the Berkeley Fellowship for Graduate Study (2014), and the Stanford University Terman Engineering Scholastic Award (2012).

Riccardo Bonalli received his M.Sc. degree in mathematical engineering from Politecnico di Milano, Italy, in 2014, and his Ph.D. degree in applied mathematics from Sorbonne Universite, France, in 2018, in collaboration with ONERA–The French Aerospace Lab, France.

He was a Postdoctoral Researcher with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA, USA. He is currently a Tenured Researcher with the Laboratory of Signals and Systems (L2S), Université Paris-Saclay, Centre National de la Recherche Scientifique (CNRS), CentraleSupélec, France. His research interests include theoretical and numerical robust optimal control with techniques from differential geometry, statistical analysis, and machine learning and applications in aerospace systems and robotics.

Dr. Bonalli is the recipient of the ONERA DTIS Best Ph.D. Student Award 2018.

Kevin Smith received his B.A. degree in environmental studies from Oberlin College, Oberlin, OH, USA, and his B.S. degree in earth and environmental engineering from Columbia University, New York, NY, USA. He is working toward his Ph.D. degree in environmental and water resources engineering with Tufts University, Medford, MA, USA.

He is currently a Product Developer with OptiRTC, Inc., Boston, MA, USA, where he is responsible for developing flexible real-time systems for the continuous monitoring and adaptive control of stormwater infrastructure. His research interests include the opportunities and risks associated with semiautonomous civil infrastructure, especially when considered as a technology for mediating environmental conflicts.

Mr. Smith is a recipient of the US National Science Foundation Integrative Graduate Education and Research Traineeship (IGERT) on Water and Diplomacy.

Insoon Yang (Member, IEEE) received his Ph.D. degree in electrical engineering and computer sciences from the University of California Berkeley, Berkeley, CA, USA, in 2015.

He is currently an Associate Professor of ECE with Seoul National University, Seoul, South Korea. From 2016 to 2018, he was an Assistant Professor of ECE with the University of Southern California, Los Angeles, CA, USA. From 2015 to 2016, he was a Postdoctoral Associate with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA. His research interests include stochastic control, optimization, and reinforcement learning.

Dr. Yang is a recipient of the 2015 Eli Jury Award and a finalist for the Best Student Paper Award at the 55th IEEE Conference on Decision and Control 2016.

Marco Pavone (Member, IEEE) received his Ph.D. degree in aeronautics and astronautics from the Massachusetts Institute of Technology, Cambridge, MA, USA in 2010, where he was affiliated with the Laboratory for Information and Decision Systems.

He is currently an Associate Professor of Aeronautics and Astronautics at Stanford University, Stanford, CA, USA, where he is currently the Director of the Autonomous Systems Laboratory and Co-Director of the Center for Automotive Research. He is currently on a partial leave of absence at NVIDIA serving as Director of Autonomous Vehicle Research. His research interests include the development of methodologies for the analysis, design, and control of autonomous systems, with an emphasis on self-driving cars, autonomous aerospace vehicles, and future mobility systems.

Dr. Pavone is a recipient of the Presidential Early Career Award for Scientists and Engineers (PECASE), an Office of Naval Research Young Investigator Award, a National Science Foundation Early Career (CAREER) Award, a NASA Early Career Faculty Award, and an Early-Career Spotlight Award from the Robotics Science and Systems Foundation.

Claire Tomlin (Fellow, IEEE) received her Ph.D. degree in electrical engineering and computer sciences from the University of California Berkeley, Berkeley, CA, USA in 1998.

She is the Charles A. Desoer Professor of Engineering with the Department of Electrical Engineering and Computer Sciences (EECS), University of California Berkeley, Berkeley, CA, USA. From 1998 to 2007, she was an Assistant, Associate, and Full Professor with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA, USA, and in 2005, she joined UC Berkeley. Her research interests include control theory and hybrid systems, with applications to air traffic management, UAV systems, energy, robotics, and systems biology.

Dr. Tomlin is a MacArthur Foundation Fellow (2006), and in 2017, she was awarded the IEEE Transportation Technologies Award. In 2019, she was elected to the National Academy of Engineering and the American Academy of Arts and Sciences.