A Risk-Sensitive Finite-Time Reachability Problem for Safety of Stochastic Dynamic Systems

Margaret P. Chapman[1], Jonathan P. Lacotte[2], Donggun Lee[3], Kevin Smith[4], Victoria Cheng[5], Jaime Fernandez-Fisac[1], Aviv Tamar[1], Susmit Jha[6], Claire J. Tomlin[1]

Abstract*— A classic reachability problem for safety of dynamic systems is to compute the set of initial states from which the state trajectory is guaranteed to stay inside a given constraint set over some time horizon. In this paper, we leverage existing theory of reachability analysis and risk measures to formulate a *risk-sensitive* reachability problem for safety of stochastic dynamic systems under non-adversarial disturbances over a finite time horizon. We provide two key contributions to the reachability literature. First, our formulation quantifies the distance between the boundary of the constraint set and the state trajectory for a stochastic dynamic system. In the literature, Hamilton-Jacobi (HJ) reachability methods quantify this distance for non-deterministic systems subject to adversarial disturbances, while stochastic reachability methods reduce the distance to a binary random variable in order to quantify the probability of safety. Second, our formulation accounts for rare high-consequence events by posing the optimal control problem in terms of a risk measure, called *Conditional Value-at-Risk* (CVaR). HJ reachability assumes that high-consequence events occur always, which may yield overly conservative solutions in practice, whereas stochastic reachability does not explicitly account for rare high-consequence events, since the optimal control problem is posed in terms of the expectation operator. We define a *risk-sensitive safe set* as the set of initial states from which the risk of extreme constraint violations can be made small via an appropriate control policy, where risk is quantified using CVaR. We show that certain risk-sensitive safe sets enjoy probabilistic safety guarantees. We provide a dynamic programming algorithm to compute under-approximations of risk-sensitive safe sets and prove the correctness of the algorithm for finite probability spaces. Our proof is a novel contribution, as it does not require the assumption of strong duality, which was required in a previous paper. Finally, we demonstrate the utility of risk-sensitive reachability analysis on a numerical example.

I. Introduction

Reachability analysis is a formal verification method based on optimal control theory that is used to prove safety or performance properties of dynamic systems [1]. A classic reachability problem for safety is to compute the set of initial states from which the state trajectory is guaranteed to stay inside a given constraint set over some time horizon. This problem was first considered for discrete-time dynamic systems by Bertsekas and Rhodes under the assumption that disturbances are uncertain but belong to known sets [2], [3], [4]. In this context, the problem is solved using a minimax formulation, in which disturbances behave adversarially and safety is described as a binary notion based on set membership [2], [3], [4].[1]

In practice, minimax formulations can yield overly conservative solutions, particularly because disturbances are not often adversarial. Most storms do not cause major floods, and most vehicles are not involved in pursuit-evader games. If there are enough observations of the system, one can estimate a probability distribution for the disturbance, and then assess safety properties of the system in a more realistic context.[2] For stochastic discrete-time dynamic systems, Abate et al. developed an algorithm that computes the set of initial states from which the probability of safety of the state trajectory can be made large by an appropriate control policy [6].[3] Summers and Lygeros extended the algorithm of Abate et al. to quantify the probability of safety and performance of the state trajectory, by specifying that the state trajectory should also reach a target set [7].

Both the stochastic reachability methods [6], [7] and the minimax reachability methods [2], [3], [4] for discrete-time dynamic systems describe safety as a binary notion based on set membership. In Abate et al., for example, the probability of safety to be optimized is the expectation of the product (or maximum) of indicator functions, where each indicator encodes the event that the state at a particular time point is inside a given set [6]. The stochastic reachability methods [6], [7] do not generalize to quantify the random distance between the state trajectory and the boundary of the constraint set, since they use indicator functions to convert probabilities to expectations to be optimized.

In contrast, Hamilton-Jacobi (HJ) reachability methods quantify the deterministic analogue of this distance for continuous-time systems subject to adversarial disturbances (e.g., see [1], [8], [9], [10]). Quantifying the distance between the state trajectory and the boundary of the constraint set in a non-binary fashion may be important in applications where the boundary is not known exactly, or where mild constraint violations are inevitable, but extreme constraint violations must be avoided.

[1]M.C., J.F., A.T., and C.T. are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, USA. chapmanm@berkeley.edu

[2]J.L. is with the Department of Aeronautics and Astronautics, Stanford University, USA.

[3]D.L. is with the Department of Mechanical Engineering, University of California, Berkeley, USA.

[4]K.S. is with OptiRTC, Inc. and the Department of Civil and Environmental Engineering, Tufts University, USA.

[5]V.C. is with the Department of Civil and Environmental Engineering, University of California, Berkeley, USA.

[6]S.J. is with SRI International, Menlo Park, California, USA.

[1]in ref. [4], see Sec. 3.6.2, "Control within a Target Tube"

[2]Ref. [5] presents methods for estimating probability distributions.

[3]Safety of the state trajectory is the event that the state trajectory stays in the constraint set over a finite time horizon.

It is imperative that reachability methods for safety take into account the possibility that rare events can occur with potentially damaging consequences. Reachability methods that assume adversarial disturbances (e.g., [1], [3]) suppose that harmful events can always occur, which may yield solutions with limited practical utility, especially in applications with large uncertainty sets. Stochastic reachability methods [6], [7] do not explicitly account for rare high-consequence events because the optimal control problem is expressed as an expectation.

In contrast, we leverage existing results on *risk measures* to formulate an optimal control problem that explicitly encodes a realistic viewpoint on the possibility of rare high-consequence events: harmful events are likely to occur at some point, but they are unlikely to occur always. A *risk measure* is a function that maps a random variable, Z, representing loss into the real line, according to the risk associated with Z (see [11], Sec. 6.3; see [12], Sec. 2.2). Risk-sensitive optimization is being studied in applied mathematics [13], reinforcement learning [14], [15], [16], and optimal control [17].[4] Risk-sensitive formulations have the potential to inform practical decision-making that also protects against damaging outcomes [18], where the level of conservatism can be modified as needed.

We use a particular risk measure, called *Conditional Value-at-Risk* (CVaR), in this paper. If Z is a random cost with finite expectation, then the Conditional Value-of-Risk of Z at confidence level $\alpha \in (0,1)$ is,

$$\text{CVaR}_\alpha[Z] = \min_{t\in\mathbb{R}} \left\{ t + \frac{1}{\alpha}\mathbb{E}\big[\max\{Z-t,0\}\big]\right\}; \quad (1)$$

see [11], Equation 6.22.[5] Note that $\text{CVaR}_\alpha[Z]$ increases from $\mathbb{E}[Z]$ to $\sup Z$, as α decreases from 1 to 0.[6] Further, there is a well-established relationship between CVaR and chance constraints that we use to obtain probabilistic safety guarantees. Chow et al. provides tractable methods to compute the CVaR of a cumulative cost incurred by a Markov Decision Process [15] that we also leverage. CVaR has additional desirable properties that are of particular interest to researchers in financial risk management and are summarized in ref. [18].

The key contributions of this paper follow. We formulate a risk-sensitive reachability problem for safety of stochastic dynamic systems under non-adversarial disturbances over a finite time horizon. In particular, our formulation quantifies the non-binary distance between the boundary of the constraint set and the state trajectory for a stochastic dynamic system. This is an extension of stochastic reachability methods (e.g., [6], [7]), which reduce this distance to a binary random variable. Further, in contrast to stochastic reachability methods, our formulation explicitly accounts for rare high-consequence events by posing the optimal control problem in terms of Conditional Value-at-Risk instead of expectation. This is the first use of risk measures in the reachability literature to our knowledge. In Sec. II, we define the notion of a *risk-sensitive safe set* and formalize the problem statement. Sec. **??** summarizes properties of risk-sensitive safe sets, including their relation to probabilistic safety. Sec. IV provides a dynamic programming algorithm to compute under-approximations of risk-sensitive safe sets. In Sec. V, we provide a numerical example in the context of the design of stormwater infrastructure. Sec. VI provides steps for future work.

II. Problem Statement

We consider a stochastic discrete-time dynamic system over a finite time horizon,[7]

$$x_{k+1} = f(x_k, u_k, w_k), \quad k = 0, 1, \dots, N-1, \quad (2)$$

such that $x_k \in \mathbb{R}^n$ is the state of the system at time k, $u_k \in U$ is the control at time k, and $w_k \in D$ is the random disturbance at time k. U and D are finite sets of real-valued vectors. The dynamics function, $f : \mathbb{R}^n \times U \times D \to \mathbb{R}^n$, is bounded and Lipschitz continuous. The probability that the disturbance equals $d_j \in D$ at time k is, $\mathbb{P}[w_k = d_j] = p_j$, where $0 \leq p_j \leq 1$ and $\sum_{j=1}^W p_j = 1$.[8] The only source of randomness in the system is the disturbance. The controls are not random. The initial condition, x_0, is not random. The states, $(x_1, \dots, x_N)$, are random because they depend on the random disturbance. The collection of *admissible control policies* is,

$$\Pi := \big\{(\mu_0, \mu_1, \dots, \mu_{N-1}), \text{ where } \mu_k : \mathbb{R}^n \to U\big\}. \quad (3)$$

We are given a constraint set, $\mathcal{K} \subset \mathbb{R}^n$, and the safety criterion that the state of the system should stay inside $\mathcal{K}$ over time. For example, if the system is a pond, then x_k may be the water level of the pond at time k, and $\mathcal{K} := [0, 5\text{ft})$ indicates that the pond overflows if the water level exceeds 5ft. We quantify the extent of constraint violation/satisfaction using a surface function that characterizes the constraint set. Let $g : \mathbb{R}^n \to \mathbb{R}$ satisfy,

$$x \in \mathcal{K} \iff g(x) < 0, \quad (4)$$

where we adopt the convention provided by [9] in Equation 2.3. For example, we may choose $g(x) := x - 5$ to characterize $\mathcal{K} := [0, 5\text{ft})$ on the state space, $\mathbb{R}_+ := [0, \infty)$.

We define a *risk-sensitive safe set* as a set of initial states from which risk of extreme constraint violation over time can be made small using an admissible control policy (3), where risk is quantified by *Conditional Value-at-Risk* (1). Formally, the risk-sensitive safe set at the confidence level, $\alpha \in (0,1)$, and the risk level, $r \in \mathbb{R}$, is defined as,

$$\mathcal{S}_\alpha^r := \{x \in \mathbb{R}^n \mid W_0^*(x, \alpha) < r\}, \quad (5a)$$

[4] In risk-sensitive optimization, the risk of a cost is minimized, where risk is quantified using a risk measure. Conversely, in stochastic optimization, we usually minimize the expected value of a cost.

[5] Conditional Value-at-Risk is also called *Average Value-at-Risk*, which is abbreviated as AV@R in [11].

[6] Technically, $\text{CVaR}_\alpha[Z] \to \text{ess}\sup Z$ as $\alpha \to 0$, where $\text{ess}\sup Z$ is the *essential supremum* of Z. Informally, essential supremum is a supremum for random variables.

[7] The system model is a special case of the model given by [4] in Sec. 1.2.

[8] We also assume that w_k is independent of x_k, u_k, and disturbances at any other times.

where

$$W_0^*(x,\alpha) := \inf_{\pi\in\Pi} \text{CVaR}_\alpha\left[Z_x^\pi\right], \quad Z_x^\pi := \max\left\{g(x_k) \mid k = 0,\dots,N\right\}, \tag{5b}$$

such that the state trajectory $(x_0, x_1, ..., x_N)$ evolves according to the dynamics model (2) with the initial state, $x_0 := x$, under the admissible policy, $\pi \in \Pi$. Note that g characterizes the constraint set, $\mathcal{K}$, according to (4). The goal of this paper is to compute a family of risk-sensitive safe sets at different levels of confidence, $\alpha \in (0,1)$, and risk, $r \in \mathbb{R}$.

III. Rationale

Computing risk-sensitive safe sets is a well-motivated problem for several reasons. This problem is more general than the stochastic reachability problem that is addressed by Abate et al. [6]. Abate et al. solves for the *maximal probabilistic safe set* at any safety level, $\epsilon \in [0,1]$,

$$S^*(\epsilon) = \{x \mid \inf_{\pi\in\Pi} \mathbb{E}[Q_x^\pi] \le \epsilon\}, \tag{6a}$$

where

$$Q_x^\pi := \max\left\{\mathbf{1}_{\bar{\mathcal{K}}}(x_k) \mid k = 0,\dots,N\right\}, \tag{6b}$$

such that the state trajectory $(x_0, x_1, ..., x_N)$ evolves according to a hybrid dynamics model with the initial state, $x_0 := x$, under the admissible policy, $\pi \in \Pi$, and

$$\mathbf{1}_{\bar{\mathcal{K}}}(x) := \begin{cases} 1 \text{ if } x \notin \mathcal{K} \\ 0 \text{ if } x \in \mathcal{K} \end{cases}; \tag{6c}$$

see [6], Equations 11 and 13. If we choose $\alpha := 1$, $g(x) := \mathbf{1}_{\bar{\mathcal{K}}}(x) - \frac{1}{2}$, and $r := \epsilon - \frac{1}{2}$ in (5), then we can compute (6). [Do you agree? What about the non-strict inequality in $S^*(\epsilon)$? What is the difference between the set of Markov policies $\mathcal{M}_m$ in Abate, and our Π? Abate did his formulation for hybrid systems, but we are not doing this explicitly]

Further, risk-sensitive safe sets have two desirable mathematical properties. The first property is that $\mathcal{S}_\alpha^r$ shrinks as the risk level, r, or the confidence level, α, decrease. Since $\mathcal{S}_\alpha^r$ is an r-sublevel set and CVaR_α increases as α decreases, one can show that,

$$\mathcal{S}_{\alpha_2}^{r_2} \subseteq \mathcal{S}_{\alpha_1}^{r_2} \subseteq \mathcal{S}_{\alpha_1}^{r_1}, \text{ and} \quad \mathcal{S}_{\alpha_2}^{r_2} \subseteq \mathcal{S}_{\alpha_2}^{r_1} \subseteq \mathcal{S}_{\alpha_1}^{r_1} \tag{7}$$

hold for any $r_1 \ge r_2$ and $1 > \alpha_1 \ge \alpha_2 > 0$. In other words, as the allowable level of risk of constraint violation (r) decreases, or as the fraction of damaging outcomes that are not fully addressed (α) decreases, $\mathcal{S}_\alpha^r$ encodes a higher degree of safety.

The second property is that the risk-sensitive safe sets at risk level, $r := 0$, enjoy probabilistic safety guarantees.

Lemma 1: If $x \in \mathcal{S}_\alpha^0$, then the probability that the state trajectory initialized at x exits the constraint set can be made strictly less than α by an admissible control policy.

Proof: The proof follows from the fact,[9] $\text{CVaR}_\alpha[Z_x^\pi] < 0 \implies \mathbb{P}[Z_x^\pi \ge 0] < \alpha$. Further, the event, $Z_x^\pi \ge 0$, is equivalent to the event that there exists a state, x_k, of the associated trajectory that exits the constraint set, since $g(x) \ge 0 \iff x \notin \mathcal{K}$.[10] ■

Remark 1: Lemma 1 indicates that $\mathcal{S}_\alpha^0$ is a subset of Abate et al.'s *maximal probabilistic safe set* at the safety level, α; see [6], Equations 9 and 11.

IV. Computational Method

The computation of risk-sensitive safe sets is challenging due to the presence of the maximum of costs (as opposed to a summation of costs) and the use of Conditional Value-at-Risk (as opposed to an expectation). In this paper, we provide under-approximations for risk-sensitive safe sets, and an algorithm to compute these under-approximations. Define $\mathcal{U}_\alpha^r$ for the confidence level, $\alpha \in (0,1)$, and the risk level, $r \in \mathbb{R}$,

$$\mathcal{U}_\alpha^r := \{x \in \mathbb{R}^n \mid V_0^*(x,\alpha) < \beta e^{m\cdot r}\}, \tag{8a}$$

where

$$V_0^*(x,\alpha) := \inf_{\pi\in\Pi} \text{CVaR}_\alpha\left[Y_x^\pi\right], \quad Y_x^\pi := \textstyle\sum_{k=0}^N \beta e^{m\cdot g(x_k)}, \tag{8b}$$

such that the state trajectory $(x_0, x_1, ..., x_N)$ satisfies (2) with the initial state, $x_0 := x$, under the policy, $\pi \in \Pi$; $\beta > 0$ and $m > 0$ are constants.

Lemma 2: $\mathcal{U}_\alpha^r$ is a subset of the risk-sensitive set, $\mathcal{S}_\alpha^r$. Also, the gap between $\mathcal{U}_\alpha^r$ and $\mathcal{S}_\alpha^r$ can be made smaller by increasing m.

Proof: The proof relies on two facts. The first fact is,

$$\max\{x_1,\dots,x_p\} \le \log(e^{x_1} + \cdots + e^{x_p}) \le \max\{x_1,\dots,x_p\} + \log p, \tag{9}$$

for any $x \in \mathbb{R}^p$; see [19], Sec. 3.1.5 Examples. Using this fact, one can show the following,

$$\max\{y_1,\dots,y_p\} \le \frac{1}{m}\log(e^{my_1} + \cdots + e^{my_p}) \le \max\{y_1,\dots,y_p\} + \frac{\log p}{m}, \tag{10a}$$

for any $y \in \mathbb{R}^p$, $m > 0$. So, as $m \to \infty$,

$$\frac{1}{m}\log(e^{my_1} + \cdots + e^{my_p}) \to \max\{y_1,\dots,y_p\}. \tag{10b}$$

The second fact is that Conditional Value-at-Risk is a *coherent risk measure*, so it satisfies useful properties. In particular, CVaR is positively homogeneous,

$$\text{CVaR}_\alpha[\lambda Z] = \lambda \text{CVaR}_\alpha[Z],$$

for any $\lambda \ge 0$ and random variable Z, and monotonic,

$$\text{CVaR}_\alpha[Y] \le \text{CVaR}_\alpha[Z],$$

for any random variables, $Y \le Z$; see [12], Sec. 2.2. Further, CVaR can be expressed as the supremum expectation over a particular set of probability density functions; see [11],

[9] The constraint, $\text{CVaR}_\alpha[Z] \le 0$, gives a conservative approximation of the chance constraint, $\mathbb{P}[Z > 0] \le \alpha$, for any random variable Z with finite expectation (see [11], Sec. 6.2.4). We do not show $\text{CVaR}_\alpha[Z] < 0 \implies \mathbb{P}[Z \ge 0] < \alpha$ for brevity.

[10] "Associated trajectory" refers to the trajectory that is initialized at x and evolves under the policy, $\pi \in \Pi$, according to the dynamics model (2).

Equations 6.40 and 6.70. Using this property and the fact, $\mathbb{E}[\log(Z)] \leq \log(\mathbb{E}[Z])$, one can show,

$$\text{CVaR}_\alpha[\log(Z)] \leq \log\left(\text{CVaR}_\alpha[Z]\right), \tag{11}$$

for any random variable, Z, with finite expectation. By monotonicity, positive homogeneity, (10), and (11),

$$\begin{aligned}\text{CVaR}_\alpha\big[Z_x^\pi\big] &\leq \frac{1}{m}\text{CVaR}_\alpha\big[\log\left(Y_x^\pi/\beta\right)\big] \\ &\leq \frac{1}{m}\log\left(\text{CVaR}_\alpha\big[Y_x^\pi/\beta\big]\right).\end{aligned} \tag{12}$$

If $x \in \mathcal{U}_\alpha^r$, then $\exists \pi \in \Pi$ such that,[11]

$$\begin{aligned}&\text{CVaR}_\alpha\big[Y_x^\pi/\beta\big] < e^{m\cdot r} \iff \frac{1}{m}\log\left(\text{CVaR}_\alpha\big[Y_x^\pi/\beta\big]\right) < r \\ &\implies \text{CVaR}_\alpha\big[Z_x^\pi\big] < r,\end{aligned}$$

where the last line holds by (12). So, $x \in \mathcal{S}_\alpha^r$. ■

Remark 2: β is included in (8) to help counter numerical issues that may arise if m is chosen very large.

Next, we provide an algorithm to compute the under-approximation, $\mathcal{U}_\alpha^r$, at different levels of confidence and risk, that is based on insights by Chow et al. [15].

A. Algorithm

The value function, $V_0^*(x,\alpha)$, defined in (8), is the smallest risk at the confidence level, α, of the cumulative scaled constraint violation of the state trajectory that satisfies (2) with $x_0 := x$. This value function can be computed by considering an observable *augmented state*, (x,α), that consists of the original state and the confidence level, and the augmented set of control policies,

$$\begin{aligned}&\bar{\Pi}_k := \{(\mu_k,\mu_{k+1},\ldots,\mu_{N-1}), \mu_i : \mathbb{R}^n \times (0,1) \to U\}, \\ &k = 0,\ldots,N-1.\end{aligned} \tag{13}$$

Our algorithm computes an analogue of $V_0^*(x,\alpha)$ that is optimized over the augmented policy space,

$$J_0^*(x,\alpha) := \inf_{\pi\in\bar{\Pi}_0} \text{CVaR}_\alpha\Big[\textstyle\sum_{k=0}^N c(x_k)\Big], \tag{14}$$

where $c : \mathbb{R}^n \to \mathbb{R}$ is a stage cost.[12] The algorithm is based on a prior result that provides the confidence level dynamics.

Lemma 3: Lemma 22 of [20] implies the following CVaR-decomposition for the system (2) at time k, under a given policy, $\pi_k := (\mu_k, \pi_{k+1}) \in \bar{\Pi}_k$,

$$\begin{aligned}&\text{CVaR}_\alpha\Big[Z\Big|x_k,\pi_k\Big] \\ &= \max_{R\in\mathcal{R}(\alpha,\mathbb{P})} \mathbb{E}\Big[R\cdot\text{CVaR}_{\alpha R}\big[Z\big|x_{k+1},\pi_{k+1}\big]\Big|(x_k,\alpha),\mu_k\Big],\end{aligned} \tag{15a}$$

where

$$\mathcal{R}(\alpha,\mathbb{P}) := \left\{R : D \to \left[0,\alpha^{-1}\right], \sum_{j=1}^W R(d_j)\mathbb{P}[w_k = d_j] = 1\right\} \tag{15b}$$

is a set of discrete random variables, and $Z := \sum_{i=k+1}^N c(x_i)$ is the random cumulative cost starting at time $k+1$.

Remark 3: The proof of Lemma 3, which we do not provide due to lack of space, relies on the measure-theoretic definition of conditional expectation, and the fact that the system (2) is Markov. please provide more details here

Remark 4: $\text{CVaR}_\alpha\big[\sum_{i=k+1}^N c(x_i)\big|x_k,\pi_k\big]$ is the risk of the cumulative cost starting at time $k+1$ of the trajectory that starts at time k, is initialized at the state, $x_k \in \mathbb{R}^n$, and evolves under the policy, $\pi_k \in \bar{\Pi}_k$.

Remark 5: For the system (2),

$$\begin{aligned}&\mathbb{E}\Big[R\cdot\text{CVaR}_{\alpha R}\big[Z\big|x_{k+1},\pi_{k+1}\big]\Big|(x_k,\alpha),\mu_k\Big] \\ &= \textstyle\sum_{j=1}^W v_j\cdot\text{CVaR}_{\alpha v_j}\big[Z\big|x_{k+1}^j,\pi_{k+1}\big]\cdot\mathbb{P}[w_k = d_j], \text{ where} \\ &x_{k+1}^j = f(x_k,\mu_k(x_k,\alpha),d_j).\end{aligned}$$

In words, x_{k+1}^j is the j^{th} sample of the state at time $k+1$, given the control at time k, $\mu_k(x_k,\alpha) \in U$, and the disturbance at time k, $w_k = d_j$. Further, $v_j = R(d_j)$ is the j^{th} sample of the random variable, $R \in \mathcal{R}(\alpha,\mathbb{P})$.

Remark 6: Chow et al. [15] interpreted Lemma 22 of [20] for Markov Decision Processes. Lemma 3 restates the result for stochastic dynamic systems.

Next, we use Lemma 3 to provide a dynamic programming value-iteration that computes (14).

Theorem 1: Define the functions, $J_{N-1}, \ldots, J_0$, recursively as follows, $\forall (x_k,y_k) \in \mathbb{R}^n \times (0,1)$,

$$\begin{aligned}&J_k(x_k,y_k) := \\ &\min_{u_k\in U}\left\{c(x_k) + \max_{R\in\mathcal{R}(y_k,\mathbb{P})} \mathbb{E}\Big[RJ_{k+1}(x_{k+1},y_kR)\Big|(x_k,y_k),u_k\Big]\right\}, \\ &k = N-1,\ldots,0,\end{aligned} \tag{16}$$

where $J_N(x_k,y_k) := c(x_k)$, x_{k+1} satisfies (2), and $\mathcal{R}(y_k,\mathbb{P})$ is defined in (15). Then, $J_0^*(x,\alpha)$, as defined in (14), is given by $J_0(x,\alpha)$, the value of the function at the last step of the recursion evaluated at $(x,\alpha) \in \mathbb{R}^n \times (0,1)$.

Remark 7: The functions, $J_{N-1}, \ldots, J_0$, are well-defined and finite because D is a finite set (see [4], Sec. 1.5).

Remark 8: Chow et al. proposed the recursion (16) and applied it to the infinite-time discounted problem [15]. The part of their proof that addresses the finite time case assumes that the min/max can be exchanged(see [15], Theorem 4), which our proof does not assume.

Our proof of Theorem 1 is provided in the Appendix. The idea is to use a sub-optimal value function as machinery to demonstrate that each J_k, as defined recursively in (16), is sufficiently close to the optimal cost-to-go of the sub-problem that starts at time k,

$$J_k^*(x_k,y_k) := \min_{\pi_k\in\bar{\Pi}_k} \text{CVaR}_{y_k}\Big[\textstyle\sum_{i=k}^N c(x_i)\Big|(x_k,y_k),\pi_k\Big], \tag{17}$$

via induction. This technique is also used to prove the suitability of the classic finite-time dynamic programming algorithm, where the value function is the expected cumulative cost (see [4], Sec. 1.5). We recommend the reader to review this proof before proceeding to ours.

[11] The minimum is attained because the random variables in this paper have finite expectation.

[12] We set $c(x) := \beta e^{mg(x)}$ in this paper.

TABLE I

Sample moment	Value
Mean	12.16 ft^3/s
Variance	3.22 ft^6/s^2
Skewness	1.68 ft^9/s^3

In this paper, we do not explicitly construct an optimal policy for practical use. This is an important step for future work and may require different arguments than those used by [15] (see Theorem 5). In particular, this policy would likely depend on the history of the states and the initial confidence level. However, the policy given by the algorithm (16) requires the availability of the current state and the current confidence level. This subtle distinction deserves careful study that is out of scope of the current paper.

V. Numerical Example

We demonstrate the utility of computing approximate risk-sensitive safe sets on a practical example: to evaluate the design of a stormwater retention pond. Stormwater management facilities, such as retention ponds, are required to operate safely in the presence of precipitation uncertainty, but must be designed within the scope of public resources (e.g., money, land). Standard design practices assess how empty ponds respond to a given *design storm*, which is a synthetic storm based on historical rainfall. In our prior work, we proposed using reachability analysis to augment existing design practices, as it can assess system behavior from a larger number of initial conditions, but we treated the surface runoff generated by the design storm as a deterministic input [21]. Here we consider the first pond from the example in our prior work as a stochastic discrete-time dynamic system,

$$
\begin{aligned}
x_{k+1} &= x_k + \frac{\triangle t}{A}(w_k - q_p(x_k,u_k)),\ k=0,\dots,N-1,\\
q_p(x_k,u_k) &:= \begin{cases} C_d \pi r^2 u_k \sqrt{2\gamma(x-E)} & \text{if } x_k \geq E \\ 0 & \text{if } x_k < E, \end{cases}
\end{aligned} \tag{18}
$$

where $x_k \geq 0$ is the water level of the pond in feet at time k, $u_k \in \{0,1\}$ is the valve setting at time k, and $w_k \in D := \{d_1,\dots,d_{10}\}$ is the random surface runoff in feet-cubed-per-second at time k [21].[13] We estimated a finite probability distribution for w_k using the surface runoff samples that we previously generated from a time-varying design storm [21]. We averaged each sample over time and solved for a distribution that satisfied the empirical statistics of the time-averaged samples (Tables I, II). We set $\triangle t := 300$ seconds, and $N := 48$ to yield a 4-hour horizon. We set the constraint set, $\mathcal{K} := [0, 5\text{ft})$, and $g(x) := x - 5$.

[13] $\gamma = 32.2\text{ft/s}^2$ is acceleration due to gravity, $\pi \approx 3.14$ is the usual constant, $r = 1/3$ft is the outlet radius, $A = 28,292\text{ft}^2$ is the pond surface area, $C_d = 0.61$ is the discharge coefficient, and $E = 1$ft is the elevation of the outlet. Some of the parameter names are changed from [21] to avoid abuse of notation.

TABLE II

Disturbance sample	Probability
$d_1 = 8.57$ ft^3/s	$\mathbb{P}[w_k = d_1] = 0.0236$
$d_2 = 9.47$ ft^3/s	$\mathbb{P}[w_k = d_2] = 10^{-4}$
$d_3 = 10.37$ ft^3/s	$\mathbb{P}[w_k = d_3] = 10^{-4}$
$d_4 = 11.26$ ft^3/s	$\mathbb{P}[w_k = d_4] = 0.5249$
$d_5 = 12.16$ ft^3/s	$\mathbb{P}[w_k = d_5] = 0.3272$
$d_6 = 13.06$ ft^3/s	$\mathbb{P}[w_k = d_6] = 10^{-4}$
$d_7 = 13.95$ ft^3/s	$\mathbb{P}[w_k = d_7] = 10^{-4}$
$d_8 = 14.85$ ft^3/s	$\mathbb{P}[w_k = d_8] = 10^{-4}$
$d_9 = 15.75$ ft^3/s	$\mathbb{P}[w_k = d_9] = 10^{-4}$
$d_{10} = 16.65$ ft^3/s	$\mathbb{P}[w_k = d_{10}] = 0.1237$

We computed risk-sensitive safe sets, $\{\mathcal{S}_y^r\}$ (5), using a Monte Carlo procedure and the under-approximations, $\{\mathcal{U}_y^r\}$ (8), using the dynamic programming algorithm (16). The risk-sensitive sets and the under-approximations are shown for various confidence and risk levels in Fig. 1. We performed the computations on a grid of states, $x \in G_s := \{0, 0.1\text{ft}, \dots, 6.4\text{ft}, 6.5\text{ft}\}$, and confidence levels, $\alpha \in G_c := \{0.999, 0.95, 0.80, \dots, 0.20, 0.05, 0.001\}$; $G := G_s \times G_c$. Since the initial state, x_0, is non-negative and the smallest value of w_k is about 8.5ft^3/s, $x_{k+1} \geq x_k$ for all k; see (18) and Table II. If $x_{k+1} > 6.5$ft during our computations, we set $x_{k+1} := 6.5$ft to stay within the grid. All computations were done in MATLAB R2016b (The MathWorks, Inc., Natick, MA), and the code is here [22].

Dynamic programming implementation. To compute the under-approximations, $\{\mathcal{U}_y^r\}$ (8), in Fig. 1, we estimated the value function, J_0^* (14), over the grid, G, using the algorithm (16), and then extracted the βe^{mr}-sublevel sets. We used the interpolation method over the confidence levels proposed by Chow et al. [15] to approximate the expectation in (16) as a piecewise linear concave function, which we maximized by solving a linear program.[14] Further, we used multi-linear interpolation to approximate the value of $J_{k+1}(x_{k+1}, \alpha)$ at each $\alpha \in G_c$. Although the backwards recursion at step k provides the value of $J_{k+1}(x, \alpha)$ for each $(x, \alpha) \in G$, $x_k \in G_s$ does *not* imply $x_{k+1} \in G_s$, since the state space for the pond dynamics is continuous (18). We set $m := 10$ and $\beta := 10^{-3}$. The function, J_0, generated by the algorithm (16) with $c(x) := \beta e^{mg(x)}$, is provided in Fig. 2. J_0 estimates J_0^*, as defined in (14). The computation is inexact due to the interpolations over the grid.

Monte Carlo implementation. To compute the risk-sensitive safe sets, $\{\mathcal{S}_y^r\}$ (5), in Fig. 1, we estimated the value function, W_0^* (5), over the grid, G, using a Monte Carlo procedure, and then extracted the r-sublevel sets. The computational burden was manageable since an optimal control policy was known *a priori*. Since $x_{k+1} \geq x_k$ for all k, and the only way to leave the constraint set is if $x_k \geq 5$ft, the optimal control policy is to keep the valve open over time, regardless of the current state or confidence level. For each $(x, \alpha) \in G$, we sampled 100,000 trajectories starting from

[14] The linear programs were solved using MOSEK (Copenhagen, Denmark) with CVX [23].

$x_0 := x$, subject to keeping the valve open over time. For each trajectory sample i, we computed the cost sample, $z_i := \max\{g(x_k^i)\}$, and estimated the Conditional Value-at-Risk of the 100,000 cost samples at the confidence level, α. We used the CVaR estimator, $\hat{\text{CVaR}}_\alpha[Z] := \frac{1}{\alpha M}\sum_{i=1}^{M} z_i \mathbf{1}_{\{z_i \geq \hat{Q}_\alpha\}}$, where $\hat{Q}_\alpha$ is the $(1-\alpha)$-quantile of the empirical distribution of the samples, $\{z_i\}_{i=1}^M$; see [11], Sec. 6.5.1. Since the estimator is designed for continuous distributions, which was not valid for every grid point, we added zero-mean Gaussian noise with a small standard deviation ($\sigma := 10^{-12}$) to each cost sample prior to computing the CVaR. The Monte Carlo estimate of W_0^*, as defined in (5), is provided in Fig. 3. 100,000 samples per grid point appears to be sufficient. We also used the Monte Carlo procedure to estimate J_0^*, as defined in (14), with 100,000 samples per grid point and small additive zero-mean Gaussian noise ($\sigma := 10^{-7}$). The results from the Monte Carlo, see Fig. 4, and the dynamic programming algorithm, see Fig. 2, are comparable in most regions of the grid. The Monte Carlo procedure, however, does not capture the higher costs at the smallest level of confidence, $\alpha := 0.001$, that are evident using the dynamic programming algorithm.

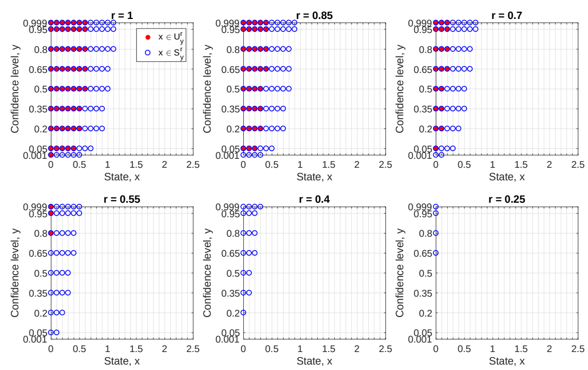

Fig. 1. Risk-sensitive safe sets, $\mathcal{S}_y^r$ (5), and their under-approximations, $\mathcal{U}_y^r$ (8) are shown for various levels of confidence and risk.

VI. Conclusion

the conditional value-at-risk is taken with respect to the probability distribution of $(w_0, \dots, w_{T-1})$.

VII. Conclusion

-inform the cost-effective design of infrastructure that must withstand rare extreme storms, -possible other applications: to reduce overly conservative error bounds that arise in safe dynamic motion planning (e.g., [8]), and to increase the amount of time that an autonomous vehicle can operate safely while simultaneously optimizing for performance.

Acknowledgment

We thank Sumeet Singh, Mo Chen, and Murat Arcak for discussions. M.C. is supported in part by a NSF Graduate Research Fellowship. This work is supported in part by NSF CPS 1740079.

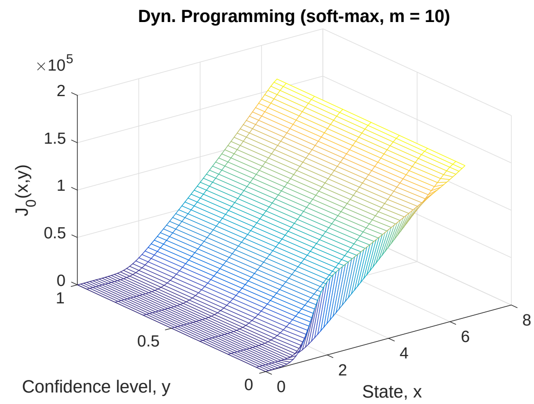

Fig. 2. $J_0(x, \alpha)$ versus $(x, \alpha) \in G$ generated by the dynamic programming algorithm (16) for the pond system is shown. J_0 estimates J_0^*, as defined in (14), where $c(x) := \beta e^{mg(x)}$, $\beta := 10^{-3}$, $m := 10$, $\mathcal{K} := [0, 5\text{ft})$, and $g(x) := x - 5$.

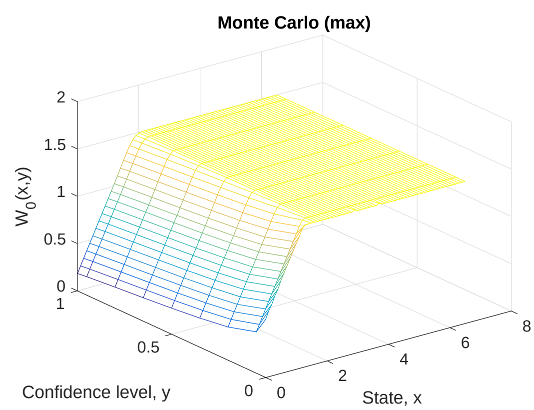

Fig. 3. The Monte Carlo estimate of $W_0^*(x, \alpha)$, as defined in (5), versus $(x, \alpha) \in G$ is shown for the pond system. 100,000 samples were generated per grid point, $g(x) := x - 5$, and $\mathcal{K} := [0, 5\text{ft})$. The maximum is 1.5ft because the system state was prevented from exceeding 6.5ft.

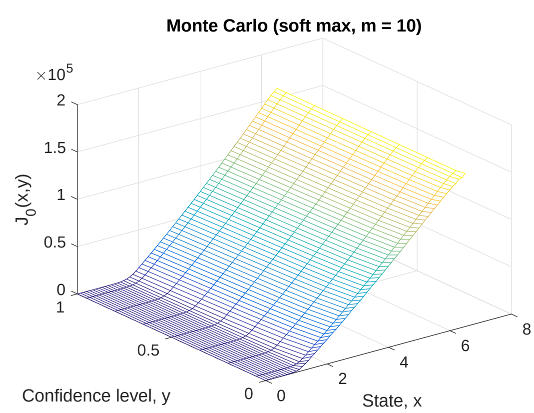

Fig. 4. The Monte Carlo estimate of $J_0^*(x, \alpha)$, as defined in (14), versus $(x, \alpha) \in G$ is shown for the pond system. $c(x) := \beta e^{mg(x)}$, $\beta := 10^{-3}$, $m := 10$, $\mathcal{K} := [0, 5\text{ft})$, and $g(x) := x - 5$. 100,000 samples were generated per grid point. See also Fig. 2.

APPENDIX

Here we provide the proof of Theorem 1. please fill in

REFERENCES

[1] S. Bansal, M. Chen, S. Herbert, and C. J. Tomlin, "Hamilton-Jacobi Reachability: A Brief Overview and Recent Advances," *arXiv preprint arXiv:1709.07523*, 2017.

[2] D. P. Bertsekas, "Control of Uncertain Systems with a Set-Membership Description of the Uncertainty," Ph.D. dissertation, Massachusetts Institute of Technology, 1971.

[3] 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.

[4] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 4th ed. Athena Scientific, 2017, vol. 1.

[5] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, 1998.

[6] 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.

[7] 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.

[8] S. L. Herbert, M. Chen, S. Han, S. Bansal, J. F. Fisac, and C. J. Tomlin, "FaSTrack: A Modular Framework for Fast and Guaranteed Safe Motion Planning," in *Decision and Control (CDC), 2017 IEEE 56th Annual Conference on*. IEEE, 2017, pp. 1517–1522.

[9] A. Akametalu, "A learning-based approach to safety for uncertain robotic systems," Ph.D. dissertation, EECS Department, University of California, Berkeley, May 2018. [Online]. Available: http://www2.eecs.berkeley.edu/Pubs/TechRpts/2018/EECS-2018-41.html

[10] I. M. Mitchell and J. A. Templeton, "A Toolbox of Hamilton-Jacobi Solvers for Analysis of Nondeterministic Continuous and Hybrid Systems," in *International Workshop on Hybrid Systems: Computation and Control*. Springer, 2005, pp. 480–494.

[11] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*. Society for Industrial and Applied Mathematics, Mathematical Programming Society, 2009.

[12] J. Kisiala, "Conditional Value-at-Risk: Theory and Applications," Master's thesis, The School of Mathematics, The University of Edinburgh, August 2015.

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

[14] T. Osogami, "Robustness and Risk-Sensitivity in Markov Decision Processes," in *Advances in Neural Information Processing Systems*, 2012, pp. 233–241.

[15] Y. Chow, A. Tamar, S. Mannor, and M. Pavone, "Risk-Sensitive and Robust Decision-Making: a CVaR Optimization Approach," in *Advances in Neural Information Processing Systems*, 2015, pp. 1522–1530.

[16] L. J. Ratliff and E. Mazumdar, "Risk-sensitive inverse reinforcement learning via gradient methods," *arXiv preprint arXiv:1703.09842*, 2017.

[17] Y.-L. Chow and M. Pavone, "A Framework for Time-consistent, Risk-Averse Model Predictive Control: Theory and Algorithms," in *American Control Conference*. IEEE, 2014, pp. 4204–4211.

[18] G. Serraino and S. Uryasev, "Conditional Value-at-Risk (CVaR)," in *Encyclopedia of Operations Research and Management Science*. Springer, 2013, pp. 258–266.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

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

[21] M. P. Chapman, K. M. Smith, V. Cheng, D. Freyberg, and C. J. Tomlin, "Reachability Analysis as a Design Tool for Stormwater Systems," in *under review for IEEE Conference on Technologies for Sustainability*.

[22] M. P. Chapman, "Risk Sensitive Reachability Project, Stormwater Example," https://github.com/chapmanmp/Risk_Sensitive_Reachability_Project/tree/stormwater_example/MATLAB_Code, 2018.

[23] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab Software for Disciplined Convex Programming," 2008.