2018 Annual American Control Conference (ACC)
June 27–29, 2018. Wisconsin Center, Milwaukee, USA

Safety-Aware Optimal Control of Stochastic Systems Using Conditional Value-at-Risk

Samantha Samuelson Insoon Yang

Abstract— **In this paper, we consider a multi-objective control problem for stochastic systems that seeks to minimize a cost of interest while ensuring safety. We introduce a novel measure of safety risk using the conditional value-at-risk and a set distance to formulate a safety risk-constrained optimal control problem. Our reformulation method using an extremal representation of the safety risk measure provides a computationally tractable dynamic programming solution. A useful byproduct of the proposed solution is the notion of a *risk-constrained safe set*, which is a new stochastic safety verification tool. We also establish useful connections between the risk-constrained safe sets and the popular probabilistic safe sets. The tradeoff between the risk tolerance and the mean performance of our controller is examined through an inventory control problem.**

I. Introduction

Control and verification of safety-critical systems have been an important problem in many domains such as air traffic control, autonomous vehicles, robotics, energy systems, and food supply chains. A part of such systems can often be modeled as a stochastic system due to the environmental and/or model uncertainty. To verify that a stochastic system is evolving within a safe range of operation with a prespecified probability and to construct an associated safety-preserving controller, several reachability-based tools have been developed using Markov chain approximations [1], Hamilton-Jacobi-Isaacs reachability [2], barrier certificates [3], dynamic programming for probabilistic safe sets [4], and infinite-dimensional linear programming [5], among others. However, ensuring safety may not be the only objective in practice: it is also desirable to minimize a cost function of interest by employing an optimal controller. Even with the aforementioned verification tools, multi-objective stochastic optimal control of safety-critical systems is challenging. One safety-oriented suboptimal approach is to use a safe control action whenever the system ventures near the boundary of a probabilistic safe set; otherwise, an optimal control action is used [6]. A lexicographic optimal control approach has been proposed in [7] to guarantee that the probability for a system being safe is close to the maximum possible safety probability. On the other hand, [8] uses linear temporal logic as a constraint to enforce safety with probability 1 in an optimal control problem.

Departing from such probabilistic and temporal logic-based methods, this paper proposes a risk-based approach to solving safety-aware optimal control of stochastic systems. Our method employs the conditional value-at-risk (CVaR) and a set distance to measure the risk of a system being unsafe. By solving an optimal control problem with associated safety risk constraints, we can design a control strategy that minimizes a cost function of interest while limiting the risk of unsafety. Unlike the probabilistic methods [6], [7], our proposed method does not require us to separately solve a verification problem to compute probabilistic safe sets or safe control policies. In other words, the risk-based approach merges the verification and optimal control procedures into a single step. A useful byproduct of our risk-constrained optimal control method is a novel verification tool, called the *risk-constrained safe set*. Such a set contains all initial states that can be driven to satisfy all safety risk constraints by an admissible control policy.

The contributions of this work can be summarized as follows. First, we introduce a novel measure of safety risk by using CVaR and the distance between the system state and a desired set A for safety. This safety risk measure represents the conditional expectation of the distance between the state and A within the $(1-\alpha)$ worst-case quantile of an associated safety loss distribution, where $\alpha \in (0,1)$. Our method enjoys an important advantage of CVaR over the value-at-risk (VaR) or the chance constraints in that CVaR takes into account the possibility of tail events in which safety losses exceed VaR while VaR is incapable of distinguishing situations beyond VaR [9]. Second, we develop a computationally tractable dynamic programming solution through two reformulation procedures. The first step reformulates the CVaR constraint in an associated Bellman equation into a tractable expectation constraint. The second step removes the minimization problem for computing a set distance from the constraint in the case of finitely supported disturbance distributions. As a result, the reformulated Bellman equation can be solved by existing convex optimization algorithms when the system dynamics are affine and the cost function is convex. Third, we establish interesting connections between the proposed risk-constrained safe sets and the popular probabilistic safe sets. In addition, we propose a simple method to compute the risk-constrained safe sets from the Bellman equation. The tradeoff between the mean performance and risk tolerance of our controller is also studied through an example of stochastic inventory control.

The remainder of this paper is organized as follows. Section II introduces the safety risk measure and an associated safety-aware optimal control problem. Its dynamic programming solution is developed in Section III. The con-

This work is supported in part by NSF under ECCS-1708906 and CNS-1657100. S. Samuelson is with the Electrical Engineering Department, University of Southern California, Los Angeles, CA 90089, USA. I. Yang is with the Electrical and Computer Engineering Department, Seoul National University, Seoul, Korea. {sasamuel, insoonya}@usc.edu

nections between risk-constrained safe sets and probabilistic safe sets are discussed in Section IV. The performance and risk aversion of the designed controller are demonstrated in Section V through an application to inventory control. The mathematical proofs omitted in this paper are contained in an extended version [10].

II. Set Distance-Based Safety Risk

Consider the following discrete-time stochastic system:

$$x_{t+1} = f(x_t, u_t, w_t), \tag{1}$$

where $x_t \in \mathbb{R}^n$ is the system state. The control input u_t is assumed to lie in a convex set $\mathbb{U} \subseteq \mathbb{R}^m$. The stochastic disturbance $w_t \in \mathbb{W} \subseteq \mathbb{R}^l$ is defined on a standard filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. Note that with the filtration $\{\mathcal{F}_t\}_{t\geq 0}$, w_t is $\mathcal{F}_t$-measurable and thus so is x_{t+1}. We say that the system is *safe* at stage t if x_t lies in a desired set A for safety, where A is a compact Borel set in $\mathbb{R}^n$. The set A represents a safe range of operation in the state space. Such a setting has been extensively used in the literature of stochastic reachability analysis (e.g., [4]).

A. Safety Specification Using Conditional Value-at-Risk

For stochastic systems of the form (1), we can measure the *loss of safety* or the *degree of unsafety* at stage t as the distance between x_t and the set A. The distance between a point $\boldsymbol{x} \in \mathbb{R}^n$ and a set $A \subseteq \mathbb{R}^n$ is defined as follows:

$$\mathrm{dist}(\boldsymbol{x}, A) := \inf_{y\in A} \|\boldsymbol{x} - y\|. \tag{2}$$

If the system is safe at stage t, i.e., $x_t \in A$, then the loss of safety $\mathrm{dist}(x_t, A) = 0$. On the other hand, when the system is unsafe, i.e., $x_t \notin A$, the loss of safety increases as x_t moves farther from the desired set A. Note that $\boldsymbol{x} \mapsto \mathrm{dist}(\boldsymbol{x}, A)$ is convex due to the triangular inequality when the set A is convex. In addition, there exists a minimizer if the set A is compact. We assume that our desired set A for safety is convex and compact.

To quantify safety risk, we adopt Conditional Value-at-Risk (CVaR) among several risk measures that are *coherent* in the sense of Artzner *et al.* [11]. CVaR measures the expected value conditioned on being within a user-specified percentile ($(1-\alpha)\times 100\%$) of the worst-case loss scenario. CVaR of a random loss X is defined as[1] $\mathrm{CVaR}_\alpha(X) := \mathbb{E}[X \mid X \geq \mathrm{VaR}_\alpha(X)]$ with $\alpha \in (0,1)$, where the value-at-risk (VaR) of X (with the cumulative distribution function F_X) is given by $\mathrm{VaR}_\alpha(X) := \inf\{x \in \mathbb{R} \mid F_X(x) \geq \alpha\}$. The following *extremal* representation of CVaR is particularly useful in optimization of CVaR [12], [9]:

$$\mathrm{CVaR}_\alpha(X) = \min_{z\in\mathbb{R}} \mathbb{E}\left[z + \frac{(X-z)^+}{1-\alpha}\right]. \tag{3}$$

Suppose that the minimization problem above has a unique optimal solution. Then, it corresponds to VaR at probability α, and CVaR is equal to VaR plus the expected safety losses exceeding VaR divided by the probability, $1-\alpha$, of these losses occurring.

Using CVaR, we can quantify the risk of the system unsafety at stage $t+1$ given the information collected up to stage $t-1$ as

$$\mathrm{CVaR}_\alpha[\mathrm{dist}(x_{t+1}, A) \mid \mathcal{F}_{t-1}] := \min_{Z\in\mathcal{L}_2(\Omega,\mathcal{F}_{t-1},\mathbb{P})} Z + \mathbb{E}\left[\frac{(\mathrm{dist}(x_{t+1}, A) - Z)^+}{1-\alpha} \mid \mathcal{F}_{t-1}\right], \tag{4}$$

which is a random variable adapted to $\mathcal{F}_{t-1}$. The safety risk (4) measures the conditional expectation of the distance between x_{t+1} and the desired set A within the $(1-\alpha)$ worst-case quantile of the safety loss distribution. Note that we use the conditional version of CVaR (conditioned on $\mathcal{F}_{t-1}$) [13], [14], which is not only practical but also essential to formulate an optimal control problem in a *time-consistent* way as explained in Section III-A.

B. Safety-Aware Stochastic Optimal Control

Our goals in designing a controller are twofold: while controlling the system (1), we want (i) to limit the safety risk (4) by a predefined threshold δ and (ii) to minimize a cost function of interest. These objectives can be achieved by solving the following risk-constrained stochastic optimal control problem:

$$\begin{aligned} \min_{\pi\in\Pi} \quad & \mathbb{E}^\pi\left[\sum_{t=0}^{T-1} r(x_t, u_t) + q(x_T)\right] \\ \text{s.t.} \quad & \mathrm{CVaR}_\alpha^\pi[\mathrm{dist}(x_{t+1}, A) \mid \mathcal{F}_{t-1}] \leq \delta, t \in \mathcal{T}, \end{aligned} \tag{5}$$

where $r : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $q : \mathbb{R}^n \to \mathbb{R}$ are a stage-wise and terminal cost function of interest, respectively, and $\mathcal{T} := \{0, 1, \cdots, T-1\}$. Here, the set Π of admissible control strategies is given by $\Pi := \{\pi := (\pi_0, \cdots, \pi_{T-1}) \mid \pi_t(\mathbb{U}|h_t) = 1 \ \forall h_t \in H_t\}$, where H_t is the set of *histories* up to stage t whose element is of the form $h_t := (x_0, w_0, \cdots, x_{t-1}, w_{t-1}, x_t)$ and π_t is a stochastic kernel from H_t to $\mathbb{U}$. In addition, $\mathbb{E}^\pi$ and CVaR_α^π represent the expectation and CVaR taken with respect to the probability measure induced by a control strategy π.

The risk tolerance parameter δ is nonnegative to be consistent with the nonnegativity of $\mathrm{dist}(x_{t+1}, A)$. When $\delta = 0$, x_{t+1} must lie in the set A with probability 1, and thus the risk constraint becomes a hard (deterministic) constraint. The threshold δ is a user-specified design parameter and has a practical meaning: δ represents the maximum allowable expected deviation of the state from the set A conditioned on being in the $(1-\alpha)$ worst-case quantile. The effect of δ on the minimal cost depends on problem instances and is studied through an example in Section V.

III. Dynamic Programming and Convexity

A. Time-Consistency and Bellman Equation

Suppose for a moment that we employ different safety risk constraints of the form $\mathrm{CVaR}_\alpha[\mathrm{dist}(x_{s+1}, A) \mid \mathcal{F}_0] \leq \delta$ $\forall s \in \mathcal{T}$. In words, we guarantee the risk constraints assuming that all of them are viewed at stage 0 with no information.

[1]This definition is valid when the probability distribution of X has no atom. For the definition of CVaR in general cases, refer to [9].

The tower rule (or the law of total expectation) does not hold for CVaR. Thus, $\mathrm{CVaR}_\alpha[\mathrm{CVaR}_\alpha[\mathrm{dist}(x_{t+1}, A) \mid \mathcal{F}_{t-1}] \mid \mathcal{F}_0] \neq \mathrm{CVaR}_\alpha[\mathrm{dist}(x_{t+1}, A) \mid \mathcal{F}_0]$, which implies that the risk constraint may be violated when evaluated at stage t with information collected up to stage $t-1$. Therefore, this problem formulation is *time-inconsistent*, meaning that an optimal control strategy constructed before or at stage 0 is no longer optimal when viewed at later stages [15]. Dynamic programming is not directly applicable to such a time-inconsistent problem as we cannot break the problem into sub-problems whose optimal solutions can be used to solve the original problem. There are two main paths to resolve the issue of time-inconsistency. The first is to focus on optimal pre-commitment strategies that are optimal viewed only at stage 0, and cannot be revised at later stages. Several techniques have been developed to compute an optimal pre-commitment strategy for optimal control of CVaR [16], [17], [18], [19], [20], [21]. The second strategy is to employ time-consistent dynamic risk measures that guarantee the time-consistency of the control problem [15], [14], [22], [23]. This approach requires special care when interpreting the practical meaning of such risk measures as they are usually defined as a composition of multiple conditional risk mappings.

Our problem formulation is closely related to the second approach: our conditional version of CVaR (4) ensures the time-consistency of the optimal control problem (5). By solving (5), a control strategy is designed offline before stage 0 to satisfy the risk constraint $\mathrm{CVaR}_\alpha[\mathrm{dist}(x_{t+1}, A) \mid \mathcal{F}_{t-1}] \leq \delta$ at stage t using information gathered up to stage $t-1$. Thus, the designed control strategy ensures the risk constraint when viewed and evaluated at stage t. To check the applicability of dynamic programming, we decompose (5) into multiple sub-problems whose optimal solutions can be used to design an optimal strategy for (5). We define the value function associated with (5) as follows:

$$\begin{aligned} v_t(\boldsymbol{x}) := \inf_{\pi \in \Pi} \mathbb{E}^\pi \left[\sum_{s=t}^{T-1} r(x_s, u_s) + q(x_T) \mid x_t = \boldsymbol{x} \right] \\ \text{s.t } \mathrm{CVaR}_\alpha^\pi[\mathrm{dist}(x_{s+1}, A) \mid \mathcal{F}_{s-1}] \leq \delta, s \in \mathcal{T}_t, \end{aligned} \tag{6}$$

which represents the minimum expected cost-to-go given the safety risk constraints are satisfied for all stages from t to $T-1$, where $\mathcal{T}_t := \{t, t+1, \cdots, T-1\}$. We now use backward induction to confirm that

$$\begin{aligned} v_t(\boldsymbol{x}) = \inf_{\boldsymbol{u} \in \mathbb{U}} \ & \mathbb{E}[r(\boldsymbol{x}, \boldsymbol{u}) + v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, w_t))] \\ \text{s.t. } & \mathrm{CVaR}_\alpha[\mathrm{dist}(x_{t+1}, A) \mid \mathcal{F}_{t-1}] \leq \delta \end{aligned} \tag{7}$$

because the risk constraints for $s \in \mathcal{T}_{t+1}$ are considered in the optimization problem (6) for v_{t+1}. However, the Bellman equation (7) involves a triple-level minimization problem with (i) the outer minimization problem over $\boldsymbol{u}$, (ii) the middle-level minimization problem (4) for CVaR and (iii) the inner minimization problem (2) for the distance function. We can significantly simplify the Bellman equation by reformulating the CVaR constraints as expectation constraints.

Theorem 1 (Bellman equation I)**.** *The value function defined in* (6) *satisfies the following Bellman equation:*

$$\begin{aligned} v_t(\boldsymbol{x}) = \inf_{\boldsymbol{u} \in \mathbb{U}, z \in \mathbb{R}} \ & r(\boldsymbol{x}, \boldsymbol{u}) + \mathbb{E}[v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, w_t))] \\ \text{s.t. } & \mathbb{E}\left[z + \frac{(\mathrm{dist}(f(\boldsymbol{x}, \boldsymbol{u}, w_t), A) - z)^+}{1-\alpha} \right] \leq \delta \end{aligned} \tag{8}$$

for $t \in \mathcal{T}$ with $v_T(\boldsymbol{x}) = q(\boldsymbol{x})$.

Proof. Using the dynamic programming principle, we have

$$\begin{aligned} v_t(\boldsymbol{x}) = \inf_{\boldsymbol{u} \in \mathbb{U}} \ & r(\boldsymbol{x}, \boldsymbol{u}) + \mathbb{E}[v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, w_t))] \\ \text{s.t } & \mathrm{CVaR}_\alpha[\mathrm{dist}(f(\boldsymbol{x}, \boldsymbol{u}, w_t), A)] \leq \delta, \end{aligned}$$

which is equivalent to (7). We denote the right-hand side of (8) as $\hat{v}_t(\boldsymbol{x})$ and show that $\hat{v}_t = v_t$. To show that $\hat{v}_t(\boldsymbol{x}) \leq v_t(\boldsymbol{x})$ fixing an arbitrary $\boldsymbol{x} \in \mathbb{R}^n$, we first note that for any $\epsilon > 0$ there exists $\boldsymbol{u}^\star \in \mathbb{U}$ such that

$$v_t(\boldsymbol{x}) + \epsilon > r(\boldsymbol{x}, \boldsymbol{u}^\star) + \mathbb{E}[v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}^\star, w_t))] \tag{9}$$

and $\mathrm{CVaR}_\alpha[\mathrm{dist}(f(\boldsymbol{x}, \boldsymbol{u}^\star, w_t), A)] \leq \delta$. By the extremal representation (3) of CVaR, the second inequality is equivalent to

$$\min_{z \in \mathbb{R}} \mathbb{E}\left[z + \frac{(\mathrm{dist}(f(\boldsymbol{x}, \boldsymbol{u}^\star, w_t), A) - z)^+}{1-\alpha} \right] \leq \delta,$$

which implies that there exists $z^\star \in \mathbb{R}$ such that

$$\mathbb{E}\left[z^\star + \frac{(\mathrm{dist}(f(\boldsymbol{x}, \boldsymbol{u}^\star, w_t), A) - z^\star)^+}{1-\alpha} \right] \leq \delta.$$

Combining this with the inequality (9), we have

$$\begin{aligned} & v_t(\boldsymbol{x}) + \epsilon > \\ & \inf_{\boldsymbol{u} \in \mathbb{U}, z \in \mathbb{R}} r(\boldsymbol{x}, \boldsymbol{u}) + \mathbb{E}[v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, w_t))] \\ & \quad \text{s.t. } \mathbb{E}\left[z + \frac{(\mathrm{dist}(f(\boldsymbol{x}, \boldsymbol{u}, w_t), A) - z)^+}{1-\alpha} \right] \leq \delta. \end{aligned}$$

Letting $\epsilon \to 0$, we have $\hat{v}_t(\boldsymbol{x}) \leq v(\boldsymbol{x})$.

We now show that $\hat{v}_t(\boldsymbol{x}) \geq v(\boldsymbol{x})$. For any $\epsilon > 0$, there exists $(\hat{\boldsymbol{u}}, \hat{\boldsymbol{z}}) \in \mathbb{U} \times \mathbb{R}$ such that

$$\hat{v}_t(\boldsymbol{x}) + \epsilon > r(\boldsymbol{x}, \hat{\boldsymbol{u}}) + \mathbb{E}[v_{t+1}(f(\boldsymbol{x}, \hat{\boldsymbol{u}}, w_t))]$$

and

$$\mathbb{E}\left[\hat{z} + \frac{(\mathrm{dist}(f(\boldsymbol{x}, \hat{\boldsymbol{u}}, w_t), A) - \hat{z})^+}{1-\alpha} \right] \leq \delta.$$

Due to the extremal formula (3) of CVaR, the second inequality implies that $\mathrm{CVaR}_\alpha[\mathrm{dist}(f(\boldsymbol{x}, \hat{\boldsymbol{u}}, w_t), A)] \leq \delta$. Therefore, $\hat{v}_t(\boldsymbol{x}) + \epsilon > v_t(\boldsymbol{x})$, which implies that $\hat{v}_t(\boldsymbol{x}) \geq v_t(\boldsymbol{x})$ as $\epsilon \to 0$. □

The minimization problem in the reformulated Bellman equation (8) is a computationally tractable stochastic program while the original Bellman equation (7) involves a nontrivial CVaR constraint. A similar reformulation approach has been proposed by Krokhmal *et al.* [24] in the context of single-stage optimization with CVaR constraints. Unlike their method based on the Karush-Kuhn-Tucker conditions, however, our proof does not assume the existence of an

optimal solution $\boldsymbol{u}^{opt}$ or the convexity of the objective and constraint functions. In other words, the proposed method not only yields a computationally tractable version of the Bellman equation but also broadens the applicability of the efficient reformulation method in [24] for CVaR-constrained optimization. We will further enhance the computational tractability of (8) in Section III-C.

B. Convexity of Value Functions

We now provide conditions under which the stochastic program in the Bellman equation (8) and the value function v_t are convex.

Proposition 1. *Suppose that $(\boldsymbol{x}, \boldsymbol{u}) \mapsto f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w})$ is affine on $\mathbb{R}^n \times \mathbb{U}$ for each $\boldsymbol{w} \in \mathbb{W}$, r is convex on $\mathbb{R}^n \times \mathbb{U}$, and q is convex on $\mathbb{R}^n$. Then, the value function v_t is convex on $\mathbb{R}^n$ for all $t \in \bar{\mathcal{T}}$.*

Since v_t is convex for all $t \in \mathcal{T}$ under the conditions in Proposition 1, the objective function of the stochastic program in the Bellman equation (8) is convex. The constraint is also convex since $\boldsymbol{u} \mapsto f(\boldsymbol{x}, \boldsymbol{u}, w_t)$ is affine for each $(\boldsymbol{x}, w_t)$ and $\boldsymbol{x} \mapsto \mathrm{dist}(\boldsymbol{x}, A)$ is convex. Therefore, the stochastic program in (8) is convex. This convexity is also used in our numerical experiments in Section V to approximate v_t as the convex envelope of v_t discretized over $\boldsymbol{x}$.

C. Finitely Supported Disturbance Distributions

We now consider the case of finitely supported disturbance distributions. This case is practically important as most empirical distributions directly obtained from data have a finite support. Furthermore, the popular sample average approximation (e.g., [25]) reduces the control problem (5) with an infinite support to the case of finitely supported disturbance distributions. Suppose that the support $\mathbb{W}$ of the disturbance distribution is given by

$$\mathbb{W} := \{\boldsymbol{w}^{(i)} \in \mathbb{R}^l \mid i = 1, \cdots, N\}, \tag{10}$$

which is a finite set. In this case, we can further simplify the Bellman equation (8) as the following deterministic optimization problem by removing the set distance function from the constraints.

Theorem 2 (Bellman equation II)**.** *Suppose that the disturbance distribution has a finite support of the form* (10)*. Then, the Bellman equation* (8) *is equivalent to*

$$v_t(\boldsymbol{x}) = \inf_{\boldsymbol{u}\in\mathbb{U}, y\in A^N, z\in\mathbb{R}} r(\boldsymbol{x}, \boldsymbol{u}) + \frac{1}{N}\sum_{i=1}^{N} v_{t+1}(f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}^{(i)})) \quad \text{s.t. } z + \frac{\sum_{i=1}^{N}(\|f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}^{(i)}) - y^{(i)}\| - z)^+}{N(1-\alpha)} \leq \delta$$

for $t \in \mathcal{T}$ with $v_T(\boldsymbol{x}) = q(\boldsymbol{x})$.

Under the conditions in Proposition 1, the simplified Bellman equation involves a deterministic convex program, which can be efficiently solved by several existing convergent algorithms. The dimension of its optimization variable linearly increases the cardinality N of the support $\mathbb{W}$. It is worth mentioning that our focus is not to resolve the fundamental scalability issue in dynamic programming: the computational complexity of our approach scales exponentially with the dimension n of state space as in standard dynamic programming. The major advantage of our method is to reformulate the triple-level optimization problem in the original Bellman equation (7) as a tractable single-level optimization problem.

IV. Risk-Constrained Safe Sets

So far, we have viewed the CVaR-based safety risk as a constraint of an optimal control problem. In this section, we illustrate how the safety risk can be used to verify the safety of stochastic systems.

A. Connection to Probabilistic Reachability Analysis

We begin by establishing a few interesting connections between our risk-based approach and the probabilistic safety/reachability specifications. To verify that a stochastic system starting from a particular initial point can be controlled to operate in a safe range A with a pre-specified probability α, one can use the *probabilistic safe set*, defined as

$$S_\alpha(A) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \exists \pi \in \Pi \text{ s.t. } x_0 = \boldsymbol{x}, \ \mathbb{P}^\pi(x_t \in A,\ t = 1, \cdots, T) \geq \alpha\}.$$

If $x_0 \in S_\alpha(A)$, then there exists a control strategy that guarantees the system safety with probability greater than or equal to α. Dynamic programming-based tools to compute such probabilistic safe sets have been developed for stochastic hybrid systems (SHS) [4], [26], [27], partially observable SHS [28], and stochastic systems under distributional ambiguity [6]. Departing from these tools for probabilistic safe sets, our risk-constrained method provides the following novel safe sets that can also be used for safety specification and verification:

Definition 1 (Risk-constrained safe set)**.** *We define the* risk-constrained safe set *for A as*

$$RS_{\alpha,\delta}(A) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \exists \pi \in \Pi \text{ s.t. } x_0 = \boldsymbol{x}, \ \mathrm{CVaR}_\alpha^\pi[\mathrm{dist}(x_{t+1}, A) \mid \mathcal{F}_{t-1}] \leq \delta, t \in \mathcal{T}\}$$

for some $\alpha \in (0, 1)$ and $\delta \geq 0$.

In words, whenever $x_0 \in RS_{\alpha,\delta}(A)$, we can control the system to satisfy the CVaR-based safety constraint for all stages. We will introduce a method to compute the risk-constrained safe sets in the next subsection. Before this, we take a close look at the CVaR constraint to relate $RS_{\alpha,\delta}(A)$ with $S_\alpha(A)$. Our first observation is that when $\delta = 0$, $\mathrm{dist}(x_{t+1}, A) = 0$ with probability 1 because (i) its $(1-\alpha)$ worst-case quantile is less than or equal to zero and (ii) the distance is greater than or equal to zero by definition. This observation leads to the following proposition:

Proposition 2. *If the risk threshold parameter $\delta = 0$, then the risk-constrained safe set $RS_\alpha(A)$ is a subset of the probabilistic safe set $S_\alpha(A)$. Furthermore,*

$$RS_{\alpha,0}(A) = S_1(A) \subseteq S_\alpha(A) \quad \forall \alpha \in (0,1).$$

Proposition 2 implies that $RS_{\alpha,0}(A)$ can be used for very conservative decision-making in terms of safety via hard constraints. When $\delta > 0$, we have another interesting connection between risk-constrained and probabilistic safe sets as follows:

Proposition 3. *Let $A_\delta := \{\boldsymbol{x} \in \mathbb{R}^n \mid \mathrm{dist}(\boldsymbol{x}, A) \leq \delta\}$ for any $\delta > 0$. Then, the risk-constrained safe set $RS_{\alpha,\delta}(A)$ is a subset of the probabilistic safe set $S_{\alpha^\tau}(A_\delta)$, i.e.,*

$$RS_{\alpha,\delta}(A) \subseteq S_{\alpha^\tau}(A_\delta).$$

Due to the definition of set distance-based safety risk, Proposition 3 compares $RS_{\alpha,\delta}(A)$ with the probabilistic safe set for a relaxed desirable set A_δ. To compare with $S_\alpha(A)$ instead of $S_\alpha(A_\delta)$, it is often useful to consider $RS_{\alpha,\delta}(A_{-\delta})$, which is contained in $S_{\alpha^\tau}(A)$, where $A_{-\delta} := \{\boldsymbol{x} \in \mathbb{R}^n \mid \mathrm{dist}(\boldsymbol{x}, A^c) \geq \delta\}$ for $\delta > 0$.

B. From Value Functions to Risk-Constrained Safe Sets

We now propose a simple approach to computing the risk-constrained safe sets by using the value function of (5). The key idea is that $\boldsymbol{x} \in RS_{\alpha,\delta}(A)$ if the control problem (5) with $x_0 = \boldsymbol{x}$ has a non-empty feasible set.

Theorem 3. *Suppose that*

$$r(\boldsymbol{x}, \boldsymbol{u}) < +\infty \;\; \forall \boldsymbol{u} \in \mathbb{U}, \quad q(\boldsymbol{x}) < +\infty$$

for each $\boldsymbol{x} \in \mathbb{R}^n$. Then, the risk-constrained safe set can be computed as

$$RS_{\alpha,\delta}(A) = \{\boldsymbol{x} \in \mathbb{R}^n \mid v_0(\boldsymbol{x}) < +\infty\}.$$

By Theorem 3, we can use our dynamic programming solution of (5) in two useful ways. First, we can verify whether a given initial state x_0 will satisfy all the safety risk constraints by checking the value $v_0(x_0)$. Second, we can explicitly construct an optimal risk-averse policy π^{opt} of (5) by solving associated Bellman equations backward in time. In particular, under the measurable selection condition (e.g., [29]), the Bellman equation admits an optimal solution $\boldsymbol{u}^{opt}$ for each $(t, \boldsymbol{x})$ and thus one can construct a non-randomized Markov policy, which is optimal, by letting $\pi_t^{opt}(\boldsymbol{x}) := \boldsymbol{u}^{opt}$.

V. Application to Inventory Control

To demonstrate the advantages of using our approach in a realistic problem, we examine an inventory control model. We define the state evolution function as

$$x_{t+1} = x_t + u_t - w_t,$$

where u_t is the quantity ordered/received at stage t, w_t is the demand at stage t, and x_t is the current inventory level. The control is bounded by $u \in [0, 32]$, and we use a time horizon of one week, i.e., $\mathcal{T} := \{0, 1, \cdots, 7\}$. Any demand that is left unsatisfied is backlogged for the next stage, which is represented as a negative state value. We define the stage-wise cost as

$$r(x_t, u_t, w_t) := c_o(x_t + u_t - w_t)^+ + c_u(w_t - x_t - u_t)^+,$$

where $c_o = 1$ represents the holding or storage cost and $c_u = 1$ represents the cost of lost sales due to unavailable inventory. The desired set for safety is chosen as $A = [0, 100]$. We use $N = 40$ samples of w_t, generated from the distribution $w_t \sim \mathcal{N}(20, 6)$.

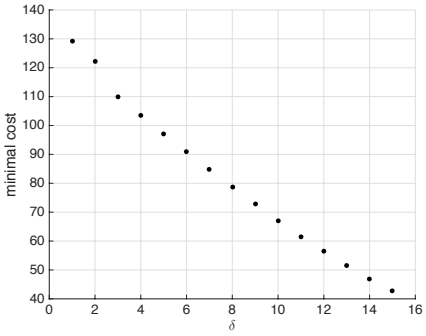

Fig. 1: Minimal expected cost over independent simulations for several different parameters δ.

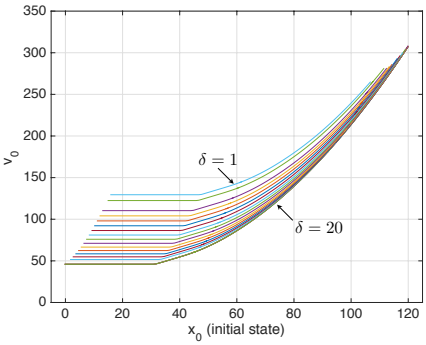

Fig. 2: Value function v_0 (minimal expected cost) for $\delta = 1, \cdots, 20$ on $\{x_0 : v_0(x_0) < +\infty\}$.

We first examine the tradeoff between the mean performance and the risk tolerance of our controller. Fig. 1 shows the mean total cost over independent simulations for several different risk tolerance values δ when $\alpha = 0.90$. As the constraint is tightened by decreasing δ, the mean total cost increases. Having a larger δ corresponds to a larger allowable deviation from the desired set A, so the total cost decreases. The choice of optimal δ thus depends on the designer's preference for either a low-risk or a low-cost controller. Fig. 2 plots the value function at stage $t = 0$, for various values of δ. Here, we can see how higher values of δ generate lower expected costs. This figure also shows another effect of tightening the CVaR constraint: the set $RS_{\alpha,\delta}(A)$ of feasible initial states becomes smaller as δ decreases.

Let $RS_{\alpha,\delta,t}(A)$ be the time-dependent risk-constrained safe set initialized at stage t, that is the set of x_t's for which the CVaR constraint can be satisfied at all future times. As shown in Fig. 3, the time-dependent safe set shrinks as we move backwards in time. A state at stage t is feasible and is in the time-dependent risk-constrained safe set if a

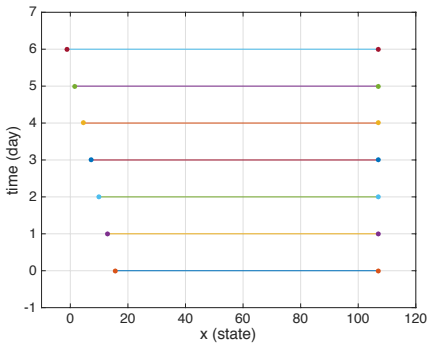

Fig. 3: Time-dependent risk-constraint safe set $RS_{\alpha,\delta,t}$ for $t = 0, \cdots, 6$.

control value can be found that satisfies two constraints: (i) $\mathrm{CVaR}_\alpha[\mathrm{dist}(x_{t+1}, A) \mid \mathcal{F}_{t-1}]$ must be no greater than δ and (ii) the future state must fall within the safe set in the next stage. The second constraint is strict and must hold for even the largest possible demand w_t. The state at stage t must be large enough that even if we encounter the maximum demand $\max\{w^{(i)}\} > u_{\max} := 32$, which leads to $x_{t+1} = x_t + u_{\max} - \max\{w^{(i)}\} < x_t$, the future state remains within the safe set. This leads to the minimum feasible state at stage t being larger than the minimum feasible state at stage $t+1$. By a similar logic, the upper bound of $RS_{\alpha,\delta,t}(A)$ is no smaller at time t than $t+1$ since $\min\{w^{(i)}\} > 0$. In this case, however, the CVaR constraint (i) is tighter than the constraint (ii)–the state at stage t must be small enough to guarantee the probability of exceeding $A_{\max} := 100$ is small. We observe that the largest state for which the CVaR constraint is satisfied is constant across all of the stages. For this reason, the upper bound of $RS_{\alpha,\delta,t}(A)$ is also constant across all stages in this example.

VI. Conclusions and Future Work

A risk-based approach has been proposed for safety-aware optimal control of stochastic systems. We developed a computationally tractable dynamic programming solution, which provides a risk-constrained optimal controller and safe set. The latter can be used for verifying the safety of stochastic systems in a risk-constrained manner while enjoying useful connections with probabilistic safe sets. We also identified the tradeoff between the risk tolerance and mean performance of our controller through a numerical example. In future research, this approach can be extended by (i) adopting a larger class of risk measures, and (ii) employing approximation and simulation-based methods.

References

[1] J. Hu, M. Prandini, and S. Sastry, "Aircraft conflict prediction in the presence of a spatially correlated wind field," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 3, pp. 326–340, 2005.

[2] 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.

[3] S. Prajna, A. Jadbabaie, and G. J. Pappas, "A framework for worst-case and stochastic safety verification using barrier certificates," *IEEE Transactions on Automatic Control*, vol. 52, no. 8, pp. 1415–1429, 2007.

[4] 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.

[5] N. Kariotoglou, M. Kamgarpour, T. H. Summers, and J. Lygeros, "The linear programming approach to reach-avoid problems for Markov decision processes," *Journal of Artificial Intelligence Research*, accepted.

[6] I. Yang, "A dynamic game approach to distributionally robust safety specifications for stochastic systems," *Automatica*, accepted.

[7] K. Lesser and A. Abate, "Multiobjective optimal control with safety as a priority," *IEEE Transactions on Control Systems Technology*, accepted.

[8] M. Svoreňová, I. Černá, and C. Belta, "Optimal control of MDPs with temporal logic constraints," in *Proceedings of the 52nd IEEE Conference on Decision and Control*, 2013.

[9] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distribution," *Journal of Banking & Finance*, vol. 26, pp. 1443–1471, 2002.

[10] S. Samuelson and I. Yang, "Safety-aware optimal control of stochastic systems using conditional value-at-risk," *arXiv preprint arXiv:1802.07903*, 2018.

[11] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Mathematical Finance*, vol. 9, no. 3, pp. 203–228, 1999.

[12] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of Risk*, vol. 2, pp. 21–42, 2000.

[13] A. Ruszczyński and A. Shapiro, "Conditional risk mappings," *Mathematics of Operations Research*, vol. 31, no. 3, pp. 544–561, 2006.

[14] A. Ruszczyński, "Risk-averse dynamic programming for markov decision processes," *Mathematical Programming*, vol. 125, pp. 235–261, 2010.

[15] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku, "Coherent multiperiod risk adjusted values and Bellman's principle," *Annals of Operations Research*, vol. 152, pp. 5–22, 2007.

[16] N. Bäuerle and J. Ott, "Markov decision processes with average-value-at-risk criteria," *Mathematical Methods of Operations Research*, vol. 74, pp. 361–379, 2011.

[17] V. Borkar and R. Jain, "Risk-constrained Markov decision processes," *IEEE Transactions on Automatic Control*, vol. 59, no. 9, pp. 2574–2579, 2014.

[18] Y. Chow, A. Tamar, S. Mannor, and M. Pavone, "Risk-sensitive and robust decision-making: a CVaR optimization approach," in *NIPS*, 2015.

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

[20] G. C. Pflug and A. Pichler, "Time-inconsistent multistage stochastic programs: Martingale bounds," *European Journal of Operational Research*, vol. 249, no. 1, pp. 155–163, 2016.

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

[22] Y. L. Chow and M. Pavone, "Stochastic optimal control with dynamic, time-consistent risk constraints," in *Proceedings of 2013 American Control Conference*, 2013.

[23] O. Çavuş and A. Ruszczyński, "Risk-averse control of undiscounted transient Markov models," *SIAM Journal on Control and Optimization*, vol. 52, no. 6, pp. 3935–3966, 2014.

[24] P. Krokhmal, J. Palmquist, and S. Uryasev, "Portfolio optimization with conditional value-at-risk objective and constraints," *Journal of Risk*, vol. 4, pp. 43–68, 2002.

[25] S. M. Robinson, "Analysis of sample-path optimization," *Mathematics of Operations Research*, vol. 21, no. 3, pp. 513–528, 1996.

[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] 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, no. 9, pp. 2665–2674, 2013.

[28] K. Lesser and M. Oishi, "Approximate safety verification and control of partially observable stochastic hybrid systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 81–96, 2017.

[29] O. Hernández-Lerma and J. B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer, 2012.