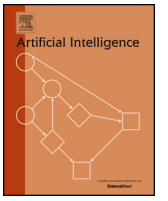
Artificial Intelligence 336 (2024) 104195

Contents lists available at ScienceDirect

Artificial Intelligence

journal homepage: www.elsevier.com/locate/artint



Sample-based bounds for coherent risk measures: Applications to policy synthesis and verification

Prithvi Akella *, Anushri Dixit, Mohamadreza Ahmadi, Joel W. Burdick, Aaron D. Ames

California Institute of Technology, Pasadena, USA

ARTICLE INFO

Keywords:
Sampling methods
Risk
Uncertain systems
Safety-critical control
Control barrier functions
Policy generation

ABSTRACT

Autonomous systems are increasingly used in highly variable and uncertain environments giving rise to the pressing need to consider risk in both the synthesis and verification of policies for these systems. This paper first develops a sample-based method to upper bound the risk measure evaluation of a random variable whose distribution is unknown. These bounds permit us to generate high-confidence verification statements for a large class of robotic systems in a sample-efficient manner. Second, we develop a sample-based method to determine solutions to non-convex optimization problems that outperform a large fraction of the decision space of possible solutions. Both sample-based approaches then permit us to rapidly synthesize risk-aware policies that are guaranteed to achieve a minimum level of system performance. To showcase our approach in simulation, we verify a cooperative multi-agent system and develop a risk-aware controller that outperforms the system's baseline controller. Our approach can be extended to account for any g-entropic risk measure.

1. Introduction

The problem of optimal policy generation under uncertainty has been well-studied in the learning community, most notably via Reinforcement Learning [1–5]. In this setting, the agent-environment interaction is modeled via a Markov Decision Process (MDP) where the state of the agent-environment pair is assumed to lie in some finite set [6]. Then, the agent's action at an initial state determines the set of states from which the successive state is randomly sampled. Each such transition is assigned a reward via a reward function, resulting in the traditional Reinforcement Learning policy generation problem - determining a policy that maximizes the expected, time-discounted reward achievable by the agent undergoing such uncertain transitions. The Partially Observable Markov Decision formulation of this problem is considered in [7–10].

However, for safety-critical control applications, expectation-maximization policies fail to consider variances in outcomes which could lead to catastrophic behavior [11]. This underscores the need to formally consider risk in both policy development and verification. To that end, Majumdar et al. [12] argue that coherent risk measures [13] like Conditional-Value-at-Risk (CVaR) and Entropic-Value-at-Risk (EVaR) which underlie the treatment of financial risk utilized by the vast majority of the world's banks [14], should similarly underlie the formal treatment of risk in robotics. This paradigm is also shared broadly within the learning and con-

* Corresponding author.
E-mail addresses: pakella@caltech.edu (P. Akella), adixit@caltech.edu (A. Dixit), mrahmadi@caltech.edu (M. Ahmadi), jwb@robotics.caltech.edu (J.W. Burdick), ames@caltech.edu (A.D. Ames).

https://doi.org/10.1016/j.artint.2024.104195
Received 15 April 2022; Received in revised form 21 March 2024; Accepted 27 July 2024

Available online 2 August 2024

trol communities, since these g-entropic risk measures like CVaR and EVaR exhibit nice mathematical properties, *e.g.* convexity, and permit a formal assessment of risk previously unaccounted for by non-coherent risk measures like Value-at-Risk. In short, this has led to the widespread adoption of these coherent risk measures as the de facto standard for formal consideration of risk in policy generation [12,15–21].

While there exist a plethora of policy synthesis techniques in a risk-sensitive setting, there exist few verification techniques - especially for g-entropic risk measures - that account for unstructured uncertainty. For example, numerous works propose risk-aware verification procedures for specific systems [22–24]. These methods verify their systems of interest against existing widespread standards, *e.g.* in [24] the authors verify a multi-agent collaborative robotic system against the international standards for safe robot interactions with humans, ISO 10218 [25,26]. As such, the verification analyses in these works are limited to their specific systems of interest, and the notion of risk is typically defined against the corresponding standard. In effect, these works do not use the same notion of risk as utilized in policy development. For more abstract, black-box approaches to verification, Corso et al. provide a very nice survey of existing techniques [27]. In short, however, the existing verification techniques that account for uncertain system measurements, *e.g.* Bayesian Optimization [28,29] or Reinforcement Learning [30,31], follow the same expectation-specific analysis that prompted the interest in a more risk-aware approach.

Motivating Questions: This inability to address risk-aware verification through the same risk measures utilized for policy development stems primarily from the inability to evaluate these measures over unknown probability distributions. While there exist concentration inequalities for specific coherent risk measures [32–35], there do not exist similar bounds for other risk measures, *e.g.* Value-at-Risk, Entropic-Value-at-Risk, *etc*. As such, the questions we aim to address are as follows. First, can we develop sample-based bounds for g-entropic risk measures like CVaR and EVaR and provide a fundamental requirement on the number of samples needed to generate our bounds even if the underlying probability distributions are unknown? Second, can we use these bounds to provide high-confidence statements on system performance in a risk-aware setting? Furthermore, the previously cited risk-aware controller generation works typically require *apriori* understanding of the underlying uncertainty whether via direct knowledge or in a distributionally-robust sense. Granted, risk-aware Reinforcement Learning does not require such knowledge, and existing convergence bounds guarantee that repeated iteration will eventually identify a satisfactory policy [18–21]. However, if we can determine a basic sample requirement for our risk measure bounds, and if the result of policy synthesis is the identification of satisfactory policies, can we provide similar sample requirements for the synthesis of satisfactory risk-aware policies and the relative complexity in identifying better policies?

Our Contributions: First, we develop a sample-based procedure that identifies a minimum sample requirement to upper bound the evaluation of a g-entropic risk measure for a random variable whose distribution is unknown. Second, we rephrase risk-aware verification as a risk measure determination problem for a random variable whose distribution is unknown. This permits us to use our prior results to generate high-confidence verification statements for arbitrarily complex robotic systems with limited system information. Third, to facilitate the rapid synthesis of risk-aware policies, we develop a sample-based procedure to identify "good" solutions to a large class of optimization problems, including the traveling salesman problem. For this problem, by a "good" solution we mean a path that is in the 99-th percentile of all possible paths with respect to minimizing the distance traveled along the path. By phrasing risk-aware synthesis as a similar type of optimization problem that employs the sample-based method developed for risk-aware verification, we develop a method to rapidly identify "good" risk-aware policies. To showcase the efficacy of our risk-aware verification and synthesis results, we verify and synthesize a controller for a cooperative three-agent robotic system that avoids self-collisions while each agent traverses to its goal. We also show that our synthesized controller outperforms the baseline controller with which the system is equipped by default, insofar as it more reliably satisfies both the safety-maintenance and goal-satisfaction objectives.

Structure: In Section 2 we provide a brief overview of some important topics - scenario optimization and g-entropic risk measures - and provide some general notation. In Section 3 we detail our sample-based upper-bounding procedure for g-entropic risk measures and provide a fundamental sample requirement for their generation. In Section 4 we detail our risk-aware verification approach as a specific example of the information developed in Section 3. Section 4.3 details the specific example of risk-aware verification for a cooperative, three-agent robotic system. Mirroring Section 3 which lays the foundation for Section 4, in Section 5 we detail a sample-based method to determine "good" solutions to a large class of optimization problems. Section 5.3 details the application to the traveling salesman problem, wherein we generate a path that is in the 99-th percentile by only having to evaluate 0.008% of the decision space to determine such a solution. Finally, in Section 6 we detail our risk-aware policy synthesis approach which employs the information in all prior sections. Section 6.3 details our risk-aware policy synthesis example for the same three-agent system verified in Section 4.3 and shows that our synthesized policy outperforms the baseline controller.

2. Preliminaries

Our analysis depends on two topics - scenario optimization and g-entropic risk measures. We will briefly describe both after some general notation.

General Notation: $|A|$ denotes the cardinality of the set A. 2^A denotes the set of all subsets of A. $\mathbb{Z}_+ = \{1, 2, \dots\}$ is the set of all positive integers, and $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$.

Scenario Optimization [36]: Scenario optimization identifies robust solutions to uncertain convex optimization problems of the following form:

$$\begin{aligned} \xi^* = \operatorname{argmin}_{\xi \in C \subset \mathbb{R}^d} \quad & c^T \xi, \\ \text{subject to} \quad & \xi \in C_\delta,\ \delta \in \Delta. \end{aligned} \tag{UP}$$

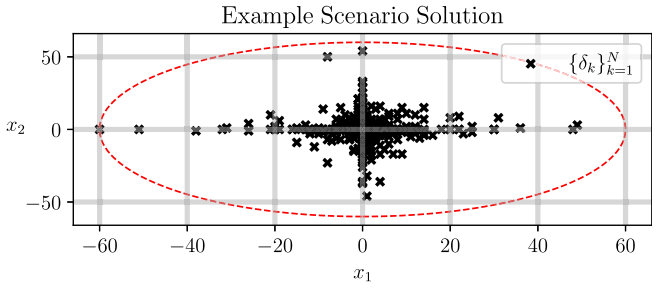


Fig. 1. An example of using scenario optimization to calculate the smallest circle radius required to encapsulate at least 0.987 of the probability mass of the underlying random variable with which samples were taken. In this case, samples δ are the randomly sampled points on the 2-d plane shown in black, the overarching decision space $C = \mathbb{R}_+$, and the sample-specific constraint spaces $C_\delta = \{r \in \mathbb{R}_+ \mid r \geq \|\delta\|\}$. Hence, the specific uncertain program (UP) would be to minimize the radius r subject to $r \in C_\delta$ for all sample-able δ.

Here, (UP) is an uncertain program as δ is a sample outcome from the probability space $\Sigma = (\Delta, \mathcal{F}, \mathbb{P})$, with sample space Δ, event space $\mathcal{F}$, and probability measure $\mathbb{P}$. Note, $\mathcal{F}$ and $\mathbb{P}$ may be unknown. Convexity is assured via an assumed convexity in the constraint spaces C and C_δ, where the latter is the subspace of C defined by the sample outcome δ. Making no further assumptions on the sample space Δ, *e.g.* finiteness, identification of a solution ξ^* such that $\xi^* \in C_\delta \ \forall \ \delta \in \Delta$ is typically infeasible. Fig. 1 describes an example uncertain program.

To resolve this issue, scenario optimization solves a related optimization problem formed from a set of N samples δ and provides a probabilistic guarantee on the robustness of the corresponding solution ξ_N^*. Specifically, given an N-sized set of samples $\{\delta_j\}_{j=1}^N$, we could construct the following scenario program:

$$\begin{aligned} \xi_N^* = \text{argmin}_{\xi \in C \subset \mathbb{R}^d} \quad & c^T \xi, \\ \text{subject to} \quad & \xi \in C_{\delta_i}, \ \forall \ \delta_i \in \{\delta_j\}_{j=1}^N. \end{aligned} \tag{RP-N}$$

Then, we require the following assumption.

Assumption 1. The scenario program (RP-N) is solvable for any N-sample set $\{\delta_j\}_{j=1}^N$ and has a unique solution ξ_N^*.

Assumption 1 guarantees the existence of a scenario solution ξ_N^* for (RP-N) for any provided sample set $\{\delta_j\}_{j=1}^N$. As such, we can define a set containing those samples $\delta \in \Delta$ for which the scenario solution ξ_N^* does not lie in the corresponding constraint set C_δ, *i.e.* $F(\xi) = \{\delta \in \Delta \mid \xi \notin C_\delta\}$. With this set definition we can formally define the *violation probability* of our solution.

Definition 1. The *violation probability* $V(\xi)$ of a given $\xi \in C$ is defined as the probability of sampling a constraint δ to which ξ is not robust, *i.e.* $V(\xi) = \mathbb{P}[\delta \in F(\xi)]$.

Then, we have the following theorem:

Theorem 1 *(Adapted from Theorem 1 in [36]). Let Assumption 1 hold. The following inequality is true:*

$$\mathbb{P}^N[V(\xi_N^*) > \epsilon] \leq \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}.$$

In Theorem 1 above, N is the number of sampled constraints δ for the scenario program (RP-N); ξ_N^* is the scenario solution to the corresponding scenario program; $V(\xi_N^*)$ is the violation probability of that solution as per Definition 1; d is the dimension of the decision variable ξ; and $\mathbb{P}^N$ is the induced probability measure over sets of N-samples of δ given the probability measure $\mathbb{P}$ for δ. An example scenario solution is shown in Fig. 1.

g-Entropic Risk Measures [37,13]: Risk measures ϕ map scalar random variables X to the real line. More accurately, consider a probability space $(\Omega, \mathcal{F}, P)$ with the sample space Ω, the event space $\mathcal{F}$, and a probability measure P. A scalar random variable (R.V.) X is a mapping from the sample space to the real-line, *i.e.* $X : \Omega \to \mathbb{R}$. The space of all scalar random variables defined for this probability space is $\tilde{X} = \{X \mid X : \Omega \to \mathbb{R}\}$. Then, a risk measure ϕ maps from a subset $\mathbf{X} \subseteq \tilde{X}$ to the extended real-line $\mathbb{R} \cup \{-\infty, \infty\}$, *i.e.* $\phi : \mathbf{X} \to \mathbb{R} \cup \{-\infty, \infty\}$. A coherent risk measure is defined as follows[1] [13].

[1] Sometimes, we see the translation invariance property in Definition 2 written as $\phi(X + C) = \phi(X) - c$ with C and c distinct but sharing a mapping, and similarly, we see the monotonicity property written as $\phi(X_1) \geq \phi(X_2)$ subject to the same conditions. We make the aforementioned changes to align with our definitions of CVaR and EVaR as per Definitions 5 and 6, which represents a shift in the intuitive meaning of these measures as compared to the financial literature.

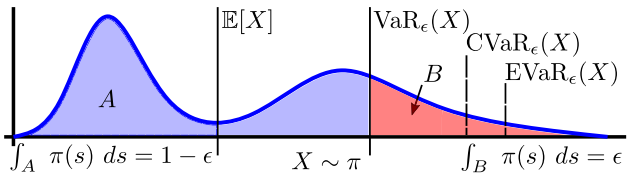


Fig. 2. Example of three common risk measures - Value-at-Risk $\mathrm{VaR}_\epsilon(X)$, Conditional-Value-at-Risk $\mathrm{CVaR}_\epsilon(X)$, and Entropic-Value-at-Risk $\mathrm{EVaR}_\epsilon(X)$ - for a scalar random variable X. Per their definitions in Section 2, $\mathrm{VaR}_\epsilon(X) \leq \mathrm{CVaR}_\epsilon(X) \leq \mathrm{EVaR}_\epsilon(X)$.

Definition 2. A *coherent risk measure* $\phi : \mathbf{X} \subseteq \tilde{X} \to \mathbb{R} \cup \{-\infty, \infty\}$ satisfies the following four properties:

1. Translation Invariance: $\phi(X + c) = \phi(X) + c$,
2. Sub-Additivity: $\phi(X_1 + X_2) \leq \phi(X_1) + \phi(X_2),\ X_1, X_2 \in \mathbf{X}$,
3. Monotonicity: If $X_1, X_2 \in \mathbf{X}$ and $X_1(\omega) \leq X_2(\omega)\ \forall\ \omega \in \Omega$, then $\phi(X_1) \leq \phi(X_2)$,
4. Positive Homogeneity: $\phi(\lambda X) = \lambda\phi(X),\ \forall\ X \in \mathbf{X},\ \lambda \geq 0$.

Ahmadi-Javid [37] defines g-entropic risk measures as those risk measures satisfying the following definition:

Definition 3 *(Adapted from Definition 5.1 in [37]).* Let g be a convex function with $g(1) = 0$ and let $\beta \geq 0$. The *g-entropic risk measure with divergence level β* $\mathrm{ER}_{g,\beta}$ is defined as follows, with P the probability measure for X, "$Q \ll P$" denoting Q is absolutely continuous with respect to P, and $\frac{dQ}{dP}$ the Radon-Nikodym derivative:

$$\mathrm{ER}_{g,\beta}(X) \triangleq \sup_{Q \in \mathcal{P}} \mathbb{E}_Q[X], \quad \mathcal{P} = \left\{ Q \;\middle|\; Q \ll P,\ \int g\left(\frac{dQ}{dP}\right) dP \leq \beta \right\}.$$

All such g-entropic risk measures are coherent risk measures as expressed in Theorem 3.2 in [37]. Of more immediate use, however, will be their representation as infimum problems, which requires a notion of L_p spaces. Specifically, for $p \geq 1$, we define L_p as the space of all random variables with p-bounded expectation, *i.e.* $L_p = \{X \in \tilde{X} \mid \mathbb{E}[|X|^p] < \infty\}$, and L_∞ is the set of all bounded scalar random variables. Then the following theorem expresses g-entropic risk measures as an infimum over a function of the expectation of the random variable X:

Theorem 2 *(Adapted from Theorem 5.1 in [37]). Let g be a closed convex function with $g(1) = 0$ and $\beta \geq 0$. For a random variable $X \in L_\infty$ with probability measure P and a g-entropic risk measure $\mathrm{ER}_{g,\beta}$ as per Definition 3, the following equivalency holds with g^* the convex-conjugate [38] of g:*

$$\mathrm{ER}_{g,\beta}(X) = \inf_{t>0, \mu \in \mathbb{R}} t\left[\mu + \mathbb{E}_P\left[g^*\left(\frac{X}{t} - \mu + \beta\right)\right]\right] \tag{1}$$

Finally, a few common risk measures - Value-at-Risk, Conditional-Value-at-Risk, and Entropic-Value-at-Risk - are defined below. Notably, Value-at-Risk is not a coherent risk measure whereas Conditional-Value-at-Risk and Entropic-Value-at-Risk are.

Definition 4. The *Value-at-Risk* level $\epsilon \in [0, 1]$, denoted as $\mathrm{VaR}_\epsilon(X)$, is the infimum over $\zeta \in \mathbb{R}$ of ζ such that samples x of the scalar random variable X with distribution π lie below ζ with probability greater than or equal to $1 - \epsilon$, *i.e.*

$$\mathrm{VaR}_\epsilon(X) = \inf\{\zeta \mid \mathbb{P}_\pi[x \leq \zeta] \geq 1 - \epsilon\}.$$

Definition 5. The *Conditional-Value-at-Risk* level $\epsilon \in (0, 1]$, denoted as $\mathrm{CVaR}_\epsilon(X)$, is the expected value of all samples x of the scalar random variable X with distribution π that are greater than or equal to the Value-at-Risk level ϵ for X, *i.e.*

$$\mathrm{CVaR}_\epsilon(X) = \mathbb{E}_\pi[x \mid x \geq \mathrm{VaR}_\epsilon(X)] = \inf_{z \in \mathbb{R}} z + \frac{\mathbb{E}_\pi[\max(X - z, 0)]}{\epsilon}.$$

Definition 6. The *Entropic-Value-at-Risk* level $\epsilon \in (0, 1]$ denoted as $\mathrm{EVaR}_\epsilon(X)$ is defined as the infimum over $z > 0$ of the Chernoff bound for the scalar random variable X with distribution π, *i.e.*

$$\mathrm{EVaR}_\epsilon(X) = \inf_{z>0} \frac{1}{z} \ln\left(\frac{\mathbb{E}_\pi\left[e^{zX}\right]}{\epsilon}\right)$$

Finally, the relationship between the three risk measures is shown in Fig. 2. Notably, for a random variable X and some risk level $\epsilon \in (0, 1)$, $\mathrm{VaR}_\epsilon(X) \leq \mathrm{CVaR}_\epsilon(X) \leq \mathrm{EVaR}_\epsilon(X)$, as shown in [37].

3. A scenario approach to bounding g-entropic risk measures

To start, by (1) we note that every g-entropic risk measure can be represented as the infimum of a loss function that linearly depends on the expectation of a function g^* of the random variable X of interest. As a result, if we could identify a general procedure to upper bound the expectation of a function of a random variable of interest, we could exploit such a procedure to develop upper bounds on a large class of g-entropic risk measures. Our formal problem statement will follow.

Problem 1. For a scalar random variable X with samples x and unknown distribution π, and a function $f : \mathbb{R} \to \mathbb{R}$, determine a method to upper bound the expected value of $f(x)$, *i.e.* to upper bound $\mathbb{E}_\pi[f(X)]$.

This section will build on our prior work wherein we employ a specific scenario program of the form (RP-N) that identifies the upper bound of a set of samples $\{x_k\}_{k=1}^N$ of a scalar R.V. X [39].

$$\begin{aligned} \zeta_N^* = \operatorname{argmin}_{\zeta \in \mathbb{R}} \quad & \zeta, \\ \text{subject to} \quad & \zeta \geq x_i, \ \forall \ x_i \in \{x_k\}_{k=1}^N. \end{aligned} \tag{UB-RP-N}$$

Then, the theorem we proved regarding high-confidence upper bounding of the Value-at-Risk level ϵ for a random variable X is as follows [39].

Theorem 3 *(Adapted from Theorem 2 in [39]). Let ζ_N^* be the solution to (UB-RP-N) for a set of N samples $\{x_k\}_{i=k}^N$ of a random variable X with unknown distribution π. The following statement is true $\forall\ \epsilon \in [0,1]$ and with $\mathrm{VaR}_\epsilon(X)$ as defined in Definition 4:*

$$\mathbb{P}_\pi^N[\zeta_N^* \geq \mathrm{VaR}_\epsilon(X)] \geq 1 - (1-\epsilon)^N.$$

3.1. Upper bounding the expected value of functions of a random variable

We will split this procedure into two steps. For the first step, if we have a scalar R.V. X with upper bound $u_b \in \mathbb{R}$, a cutoff $c \in \mathbb{R}$, and the probability mass of samples x of X that lie below this cutoff, *i.e.* $\mathbb{P}_\pi[x \leq c]$, then we can upper bound the expected value of X with this cutoff c and upper bound u_b:

$$\mathbb{E}_\pi[X] = \int_{\mathbb{R}} s\pi(s)\, ds \leq c\mathbb{P}_\pi[x \leq c] + u_b(1 - \mathbb{P}_\pi[x \leq c]). \tag{2}$$

Theorem 4 to follow, formally states that we can employ Theorem 3 to identify such a cutoff c, to upper bound $\mathbb{E}_\pi[X]$ *without knowledge of its distribution π* provided we know a (perhaps) loose upper bound $u_b \in \mathbb{R}$.

Theorem 4. *Let X be a scalar R.V. with samples x, unknown distribution function π, and upper bound $u_b \in \mathbb{R}$ such that $\mathbb{P}_\pi[x \leq u_b] = 1$. Furthermore, let ζ_N^* be the solution to (UB-RP-N) for a set of N independent and identically drawn samples $\{x_k\}_{k=1}^N$ of X. The following statement is true $\forall\ \epsilon \in [0,1]$:*

$$\mathbb{P}_\pi^N\left[\mathbb{E}_\pi[X] \leq \zeta_N^*(1-\epsilon) + u_b\epsilon\right] \geq 1 - (1-\epsilon)^N. \tag{3}$$

Proof. First, by Theorem 3 and the definition of $\mathrm{VaR}_\epsilon(X)$ in Definition 4,

$$\mathbb{P}_\pi^N\left[\mathbb{P}_\pi[x \leq \zeta_N^*] \geq 1 - \epsilon\right] \geq 1 - (1-\epsilon)^N.$$

Then by equation (2) we have the following inequality:

$$\mathbb{E}_\pi[X] \leq \zeta_N^*\mathbb{P}_\pi[x \leq \zeta_N^*] + u_b(1 - \mathbb{P}_\pi[x \leq \zeta_N^*]). \tag{4}$$

To finish the proof, we note that the right-hand side of the inequality (4) is maximized when this probability $\mathbb{P}_\pi[x \leq \zeta_N^*]$ equals its lower bound $1-\epsilon$, as $u_b \geq \zeta_N^*$. As this lower bound holds with minimum probability $1-(1-\epsilon)^N$, the result holds with the same probability, resulting in (3). □

The second step needed to bound functions of the random variable X is to note that if we have a measurable function $f : \mathbb{R} \to \mathbb{R}$, then $f(X) = Y$ is another random variable with samples y and distribution π_Y. Provided Y has an upper bound u_b, then we can upper bound $\mathbb{E}_{\pi_Y}[Y] = \mathbb{E}_\pi[f(X)]$. This result does not require X to be bounded, only the image of X under f needs to be bounded as expressed in the following corollary.

Corollary 1. *Let X be a scalar R.V. with samples x and distribution π. Let $f : \mathbb{R} \to \mathbb{R}$ and let $Y = f(X)$ be another scalar R.V. with samples $y = f(x)$, distribution π_Y, and upper bound u_b such that $\mathbb{P}_{\pi_Y}[y \leq u_b] = 1$. Furthermore, let ζ_N^* be the solution to (UB-RP-N) for a set of N samples $\{y_k = f(x_k)\}_{k=1}^N$ of Y. The following statement is true $\forall\ \epsilon \in [0,1]$.*

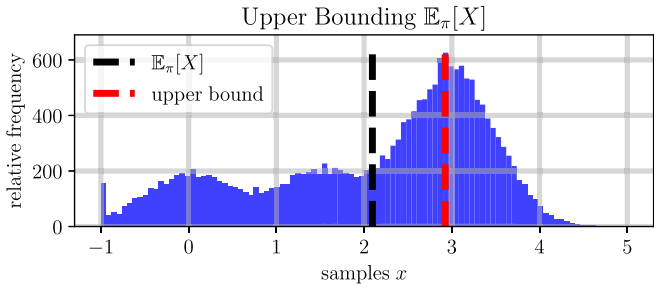


Fig. 3. To upper bound g-entropic risk measures, we motivate that we first must be able to upper bound the expected value of a scalar random variable X. The figure above is an example of our ability to do this as per Theorem 4. Here, we upper bound (red) the expected value (black) of a multi-modal random variable X by taking $N = 20$ samples of X and knowing its upper bound $u_b = 5$. The true distribution (blue) was calculated numerically by taking 20000 samples of X. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

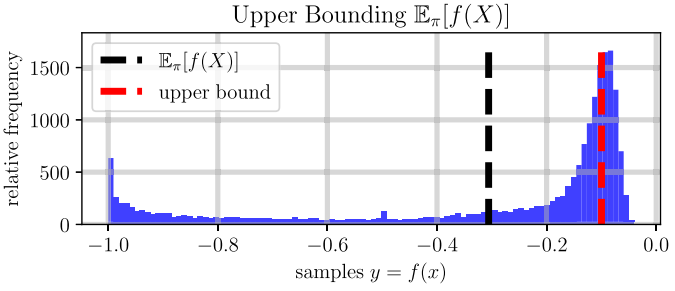


Fig. 4. The second step to upper bounding g-entropic risk measures is upper bounding the expected value of a function f of a scalar random variable X with unknown distribution π. Above is an example of us upper bounding $\mathbb{E}_\pi[f(X)]$ where $f(x) = \frac{-1}{x^2+1}$ for the multi-modal random variable X whose expectation we upper bounded in Fig. 3. To generate this upper bound we took $N = 20$ samples of the random variable $Y = f(X)$. We formally prove our ability to do this upper bound in Corollary 1. As before, the true expected value is shown in black and the upper bound is shown in red.

$$\mathbb{P}_\pi^N \left[\mathbb{E}_\pi[f(X)] \leq \zeta_N^*(1-\epsilon) + u_b\epsilon\right] \geq 1 - (1-\epsilon)^N.$$

Proof. This result stems via Theorem 4 where $\mathbb{E}_{\pi_Y}[Y] = \mathbb{E}_\pi[f(X)]$ (see Fig. 4). □

3.2. Upper bounding g-entropic risk measures

With Corollary 1, we next develop a procedure to upper bound a large class of g-entropic risk measures, including Conditional-Value-at-Risk and Entropic-Value-at-Risk. To do so for general g-entropic risk measures as per Definition 3, we require that the convex conjugate g^* of g maps from the reals to the reals, *i.e.* $g^* : \mathbb{R} \to \mathbb{R}$, and also has an upper bound when applied to samples x of the random variable X.

Assumption 2. X is a scalar random variable with samples x, distribution π, and upper bound $\ell \in \mathbb{R}$ such that $\mathbb{P}_\pi[x \leq \ell] = 1$. $\text{ER}_{g,\beta}$ is a g-entropic risk measure as per Definition 3 with respect to some $\beta \geq 0$ and closed convex function g such that $g(1) = 0$. $g^* : \mathbb{R} \to \mathbb{R}$ is the convex-conjugate for g, and g^* has an upper bounding function $u_b : D \subset \mathbb{R}^2 \to \mathbb{R}$ satisfying the following probabilistic equality $\forall\ \mu \in \mathbb{R},\ t > 0$:

$$L(x,\mu,t) \triangleq t\left(\mu + g^*\left(\frac{x}{t} - \mu + \beta\right)\right), \tag{5}$$
$$\mathbb{P}_\pi\left[L(x,\mu,t) \leq u_b(\mu,t)\right] = 1.$$

Finally, $\zeta_N^*(\mu,t)$ is the solution to (UB-RP-N) for a set of N-samples $\{y_k = L(x_k,\mu,t)\}_{k=1}^N$ of the random variable $Y = L(X,\mu,t)$ for some $\mu \in \mathbb{R},\ t > 0$.

For context, this assumption is not too restrictive, as it is easily satisfied by both CVaR and EVaR for any risk level $\alpha \in (0,1]$, and we will show this in a later section. Then via Corollary 1, we can upper bound the objective function in (1) and transform (1) into an optimization problem over a set of sampled values $\{x_k\}_{k=1}^N$ of X which is easily solvable. The formal statement of this procedure is provided in the following lemma, theorem, and corollary. The lemma will state that we can upper bound the expected value of the convex conjugate g^* applied to X, and the theorem will utilize this lemma to provide a high confidence upper bound on the g-entropic risk measure of interest. The corollary will formalize a relationship between the confidence in our estimate and the number of samples we take of the random variable.

Lemma 1. *Let Assumption 2 hold. Then,* $\forall\ \epsilon \in [0,1],\ \mu \in \mathbb{R},\ t > 0$,

$$\mathbb{P}_\pi^N \left[\mathbb{E}_\pi[L(X,\mu,t)] \leq \zeta_N^*(\mu,t)(1-\epsilon) + u_b(\mu,t)\epsilon \right] \geq 1 - (1-\epsilon)^N.$$

Proof. This is a direct application of Corollary 1. □

Theorem 5. *Let Assumption 2 hold. Then,* $\forall\ \epsilon \in [0,1]$,

$$\mathbb{P}_\pi^N \left[ER_{g,\beta}(X) \leq \inf_{t>0,\ \mu\in\mathbb{R}} \zeta_N^*(\mu,t)(1-\epsilon) + u_b(\mu,t)\epsilon \right] \geq 1 - (1-\epsilon)^N.$$

Proof. Via Theorem 2, the g-entropic risk measure $\mathrm{ER}_{g,\beta}$ can be represented via an infimum, and via linearity of the expectation operator and the definition of L in (5), we have the following equality:

$$\mathrm{ER}_{g,\beta}(X) = \inf_{t>0,\mu\in\mathbb{R}} \mathbb{E}_\pi[L(X,\mu,t)].$$

Then the result holds via Lemma 1. □

Corollary 2. *Let Assumption 2 hold,* $\gamma \in [0,1)$*, and* $\epsilon \in (0,1)$*. If for all* $\mu \in \mathbb{R}$ *and* $t > 0$, $\zeta_N^*(\mu,t)$ *is the solution to* (UB-RP-N) *for a set of* N *samples* $\{x_k\}_{k=1}^N$ *of the random variable* X *where* $N \geq \frac{\log(1-\gamma)}{\log(1-\epsilon)}$*, then*

$$\mathbb{P}_\pi^N \left[ER_{g,\beta}(X) \leq \inf_{t>0,\ \mu\in\mathbb{R}} \zeta_N^*(\mu,t)(1-\epsilon) + u_b(\mu,t)\epsilon \right] \geq \gamma.$$

Proof. Use Theorem 5 where $N \geq \frac{\log(1-\gamma)}{\log(1-\epsilon)} \implies 1-(1-\epsilon)^N \geq \gamma$. □

To summarize, Theorem 5 states that we can upper bound the g-entropic risk measure of a random variable X with unknown distribution π provided that we know an upper bounding function u_b for a function L of the random variable X. Corollary 2 provides the minimum number of samples N required to determine this upper bound with confidence $\gamma \in [0,1)$.

3.3. Specializing to CVaR and EVaR

Ahmadi-Javid [37] identifies the convex conjugate function g^* and parameter β with which CVaR can be recast as a g-entropic risk measure.

Remark 1. The Conditional-Value-at-Risk level $\alpha \in (0,1]$ can be recast as a g-entropic risk measure with convex conjugate function $g^*(x) = \frac{1}{\alpha}\max\{x,0\}$ and scalar parameter $\beta = 0$ [37].

Then, to use Theorem 5 we must show that the CVaR satisfies Assumption 2.

Lemma 2. *The Conditional-Value-at-Risk for any risk-level* $\alpha \in (0,1]$ *satisfies Assumption 2.*

Proof. To start, CVaR for any risk-level α is a g-entropic risk measure with $g^*(x) = \frac{1}{\alpha}\max\{x,0\}$ and $\beta = 0$. As a result, a solution $\zeta_N^*(\mu,t)$ will always exist for (UB-RP-N) as it is the solution to a linear program minimizing a scalar decision variable subject to a finite set of lower bounds taking values in $\mathbb{R}$. Then, to prove that the Conditional-Value-at-Risk at any risk level $\alpha \in (0,1]$ satisfies Assumption 2, it suffices to identify an upper bounding function $u_b : D \subset \mathbb{R}^2 \to \mathbb{R}$ under the assumption that the scalar random variable X with distribution π and samples x has an upper bound $\ell \in \mathbb{R}$ such that $\mathbb{P}_\pi[x \leq \ell] = 1$. This function u_b will be defined as follows:

$$L(x,\mu,t) = t\left(\mu + \frac{1}{\alpha}\max\left\{\frac{x}{t} - \mu, 0\right\}\right), \quad u_b(\mu,t) = L(\ell,\mu,t). \tag{6}$$

Since this upper bounding function satisfies the probabilistic equality in Assumption 2, CVaR $\forall\ \alpha \in (0,1]$ satisfies Assumption 2. □

As the Conditional-Value-at-Risk satisfies Assumption 2, we can use u_b (6) to provide high-confidence estimates on $\mathrm{CVaR}_\alpha(X)\ \forall\ \alpha \in (0,1]$.

Corollary 3. *Let* X *be a scalar random variable with samples* x*, distribution* π*, and upper bound* $\ell \in \mathbb{R}$ *such that* $\mathbb{P}_\pi[x \leq \ell] = 1$*. Let* $\alpha \in (0,1]$*,* $\epsilon \in [0,1]$*, and* L, u_b *be as defined in* (6) *with respect to this upper bound* ℓ *and constant* α*. Furthermore, let* $\zeta^*(\mu,t)$ *be the solution to* (UB-RP-N) *for a set of* N*-samples* $\{y_k = L(x_k,\mu,t)\}_{k=1}^N$ *of the random variable* $Y = L(X,\mu,t)$*. Then,*

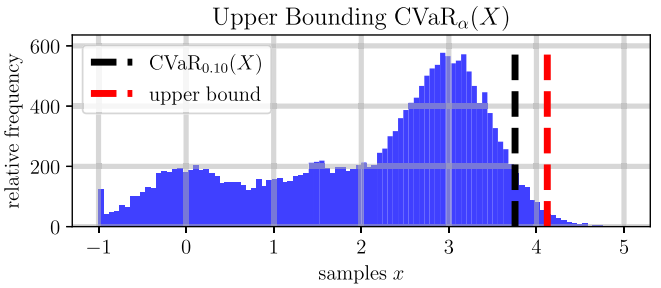


Fig. 5. A specific g-entropic risk measure, the Conditional-Value-at-Risk for any risk-level $\alpha \in (0,1]$ should be upper-boundable via our scenario approach as expressed in Corollary 3. Shown above is an example of our scenario approach in bounding $\text{CVaR}_{\alpha=0.1}(X)$ for the R.V. X whose distribution is shown in blue. To generate the upper bound shown in red, we required $N = 45$ samples of X. The true $\text{CVaR}_{0.1}(X)$ is shown in black.

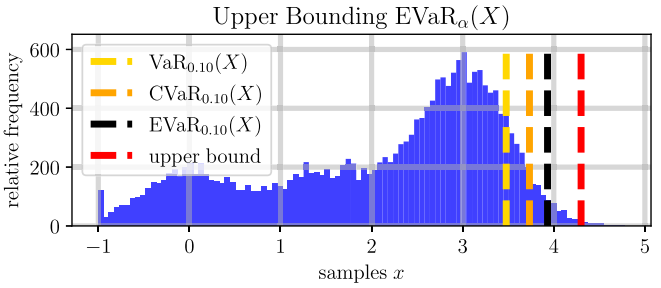


Fig. 6. A culmination of our approach to upper-bounding g-entropic risk measures. Shown above is our attempt to upper bound the Entropic-Value-at-Risk at risk level $\alpha = 0.1$ of the scalar multi-modal random variable X whose distribution is shown in blue. To calculate this upper bound (red) with $N = 20$ samples of X, we used the same method that we used to determine the upper bound for $\text{CVaR}_\alpha(X)$ in Fig. 5. We formally state our capacity to do this in Corollary 4. Notice that the true $\text{VaR}_\alpha(X) \leq \text{CVaR}_\alpha(X) \leq \text{EVaR}_\alpha(X)$ as also shown in Fig. 2. As before, the true $\text{EVaR}_\alpha(X)$ is shown in black.

$$\mathbb{P}_\pi^N \left[\text{CVaR}_\alpha(X) \leq \inf_{\mu \in \mathbb{R},\ t>0} \zeta_N^*(\mu,t)(1-\epsilon) + u_b(\mu,t)\epsilon \right] \geq 1 - (1-\epsilon)^N.$$

Proof. This is a direct application of Theorem 5 due to Lemma 2 (see Fig. 5). □

To use Theorem 5 to provide an upper bound on the Entropic-Value-at-Risk for any risk level $\alpha \in (0,1]$, we will follow a similar procedure as we followed for the Conditional-Value-at-Risk. First, Ahmadi-Javid [37] identifies the convex conjugate function g^* and parameter β which enables Entropic-Value-at-Risk to be cast as a g-entropic risk measure.

Remark 2. The Entropic-Value-at-Risk level $\alpha \in (0,1]$ can be recast as a g-entropic risk measure with convex conjugate function $g^*(x) = e^{x-1}$ and scalar parameter $\beta = -\ln(\alpha)$ [37].

Then, we note that EVaR for any risk-level $\alpha \in (0,1]$ satisfies Assumption 2.

Lemma 3. *The Entropic-Value-at-Risk for any risk-level $\alpha \in (0,1]$ satisfies Assumption 2.*

Proof. This proof follows the proof of Lemma 2 insofar as it suffices to identify an upper bounding function $u_b : D \subset \mathbb{R}^2 \to \mathbb{R}$ satisfying the probabilistic inequality in Assumption 2 where $g^*(x) = e^{x-1}$ and $\beta = -\ln(\alpha)$. The following function u_b suffices, with ℓ the upper bound for X:

$$L(x,\mu,t) = t\left(\mu + e^{\frac{x}{t} - \mu - \ln(\alpha) - 1}\right), \quad u_b(\mu,t) = L(\ell,\mu,t). \quad \square \tag{7}$$

Since $\text{EVaR}_\alpha\ \forall\ \alpha \in (0,1]$ satisfies Assumption 2, we can use u_b (7) to provide high-confidence estimates on $\text{EVaR}_\alpha(X)\ \forall\ \alpha \in (0,1]$.

Corollary 4. *Let X be a scalar random variable with samples x, distribution π, and upper bound $\ell \in \mathbb{R}$ such that $\mathbb{P}_\pi[x \leq \ell] = 1$. Let $\alpha \in (0,1]$, $\epsilon \in [0,1]$, and L, u_b be as defined in (7) with respect to this upper bound ℓ and constant α. Furthermore, let $\zeta_N^*(\mu,t)$ be the solution to (UB-RP-N) for a set of N-samples $\{y_k = L(x_k,\mu,t)\}_{k=1}^N$ of the random variable $Y = L(X,\mu,t)$. Then,*

$$\mathbb{P}_\pi^N \left[\text{EVaR}_\alpha(X) \leq \inf_{\mu \in \mathbb{R},\ t>0} \zeta_N^*(\mu,t)(1-\epsilon) + u_b(\mu,t)\epsilon \right] \geq 1 - (1-\epsilon)^N.$$

Proof. This is an application of Theorem 5 via Lemma 3 (see Fig. 6). □

4. Risk-aware verification

An important application of the sample-based bounds derived in the prior section arises in safety-critical system verification where system evolution is partially stochastic due to unmodeled dynamics, noise, *etc*. In this section, we will detail how we can reformulate safety-critical system verification as a risk measure determination problem. This reformulation lets us use our prior results to easily bound system performance in a cooperative multi-agent system setting as illustrated in Section 4.3.

4.1. Notation and problem setting

For our system under study, $\mathcal{X}$ is the state space, $\mathcal{U}$ is the input space, and Θ is a known space of parameters θ influencing the system's controller U. Furthermore, we assume the system is subject to stochastic noise ξ with an unknown distribution $\pi_\xi(x,u,t)$ over $\mathbb{R}^n$.

$$\begin{aligned} &\dot{x} = f(x,u) + \xi, && x \in \mathcal{X} \subset \mathbb{R}^n,\ u \in \mathcal{U} \subset \mathbb{R}^m, \\ &u = U(x,\theta), && \theta \in \Theta \subset \mathbb{R}^p, \\ &\xi \sim \pi_\xi(x,u,t), && \int_{\mathcal{X}} \pi_\xi(x,u,t,s)\, ds = 1 \ \forall\ x,u,t. \end{aligned} \tag{8}$$

We denote x_t^θ as our closed-loop system solution at time t - note that the parameter θ does not change over a trajectory - and x^θ will correspond to our closed-loop state signal:

$$\dot{x}_t^\theta = f\left(x_t^\theta, U\left(x_t^\theta, \theta\right)\right) + \xi, \quad x^\theta \in \mathcal{S}^{\mathbb{R}^n},\ \mathcal{S}^{\mathbb{R}^n} = \{s : \mathbb{R}_{\geq 0} \to \mathbb{R}^n\}$$

Verification work typically assumes the existence of a robustness metric ρ - a function that maps state trajectory signals to the real line, with positive evaluations of the metric indicating system objective satisfaction [40].

Definition 7. A *robustness metric* ρ is a function that maps from the signal space to the reals, *i.e.* $\rho : \mathcal{S}^{\mathbb{R}^n} \to [-a, b],\ a, b \in \mathbb{R}_{++}$. Furthermore, a signal s is "accepted" by the robustness metric ρ, *i.e.* exhibits the desired properties encoded by the robustness metric ρ, if and only if $\rho(s) \geq 0$.

Examples of robustness metrics ρ include the minimum value of a control barrier function h over some pre-specified time horizon [41, 42], or the robustness metrics of Signal Temporal Logic [40,43]. As the existence and construction of these functions have been well-studied, we will simply assume their existence for the time being [44–46].

For risk-aware verification, the uncertainty arises through the uncertainty ξ entering the system dynamics in (8). As a result, the robustness ρ of a closed loop trajectory x^θ, *i.e.* $\rho(x^\theta)$, is a scalar random variable $R(x_0, \theta)$ with some distribution $\pi_R(x_0, \theta)$ that depends on the initial condition x_0 and parameter θ for our closed-loop system trajectory x^θ. By further uniformly sampling initial conditions and parameters from their respective spaces *i.e.* sampling (x_0, θ) from $\mathrm{U}[\mathcal{X}_0 \times \Theta]$, we generate the scalar randomized robustness variable R with distribution π_R that was the subject of study in [39].

Definition 8. The *holistic system robustness*, R, is a scalar random variable[2] with distribution π_R and samples r denoting the closed-loop robustness $\rho(x^\theta)$ of trajectories x^θ whose initial condition and parameter (x_0, θ) were sampled uniformly from $\mathcal{X}_0 \times \Theta$.

The goal of risk-aware verification then would be to identify risk measures for this holistic system robustness random variable R as per Definition 8, *i.e.* identify $\mathrm{CVaR}_\alpha(-R)$ or $\mathrm{EVaR}_\alpha(-R)$ for some $\alpha \in (0, 1]$. Since we do not know the distribution π_R of R, however, direct identification of these risk measures is difficult. Therefore, we will instead identify upper bounds for these risk measures - r_C^* for $\mathrm{CVaR}_\alpha(R)$ and r_E^* for $\mathrm{EVaR}_\alpha(R)$ - that are upper bounds with some minimum probability - ϵ_C, ϵ_E respectively. The formal statement of this problem follows.

Problem 2. For the scalar random variable R with distribution π_R as per Definition 8, devise a method to determine upper bounds $r_C^*, r_E^* \in \mathbb{R}$ with corresponding probabilities $\epsilon_C, \epsilon_E \in [0, 1]$ such that for some $\alpha \in (0, 1]$,

$$\mathbb{P}_{\pi_R}[r_C^* \geq \mathrm{CVaR}_\alpha(-R)] \geq 1 - \epsilon_C,\ \mathbb{P}_{\pi_R}[r_E^* \geq \mathrm{EVaR}_\alpha(-R)] \geq 1 - \epsilon_E.$$

Here, $\mathrm{CVaR}_\alpha(-R)$, $\mathrm{EVaR}_\alpha(-R)$ are the coherent risk measures defined in Definitions 5 and 6 respectively.

[2] Strictly speaking, this definition assumes that the robustness metric ρ is a measurable function. This is typically the case for robustness metrics in robotics, however, so this assumption has been omitted. That being said, determining a class of robustness metrics satisfying Definition 7 that are provably measurable is an open problem.

4.2. Upper bounds for risk-aware verification

First, we note that via Definitions 7 and 9, the negation of the holistic system robustness R has a probabilistic upper bound $a \in \mathbb{R}$, i.e. $\mathbb{P}_{\pi_R}[-r \leq a] = 1$. Therefore, we can directly apply Corollaries 3 and 4 to solve Problem 2. To formally state our results in this vein, we will redefine the loss function L and upper bounding function u_b for each risk measure in this specific application. Here, a is the upper bound for $-R$:

$$L(x,\mu,t) = t\left(\mu + \frac{1}{\alpha}\max\left\{\frac{x}{t} - \mu, 0\right\}\right), \qquad u_b(\mu,t) = L(a,\mu,t), \tag{CVaR}$$

$$L(x,\mu,t) = t\left(\mu + e^{\frac{x}{t}-\mu-\ln(\alpha)-1}\right), \qquad u_b(\mu,t) = L(a,\mu,t). \tag{EVaR}$$

First, we have that we can upper bound $\mathrm{CVaR}_\alpha(-R)\ \forall\ \alpha \in (0,1]$.

Corollary 5. *Let R be a scalar random variable as per Definition 8, let $\alpha \in (0,1]$, let L, u_b be as defined in equation (CVaR) with respect to this α, let $\epsilon \in (0,1)$ and $\gamma \in [0,1)$, and let $\zeta_N^*(\mu,t)$ be the solution to (UB-RP-N) for a set of N-samples $\{y_k = L(-r_k,\mu,t)\}_{k=1}^N$ of the random variable $Y = L(-R,\mu,t)$, where $N \geq \frac{\log(1-\gamma)}{\log(1-\epsilon)}$. Then,*

$$\mathbb{P}_{\pi_R}^N\left[r_C^* \triangleq \inf_{\mu\in\mathbb{R},\ t>0} \zeta_N^*(\mu,t)(1-\epsilon) + u_b(\mu,t)\epsilon \geq \mathrm{CVaR}_\alpha(-R)\right] \geq \gamma.$$

Proof. This is an application of Corollary 3 with L and u_b as per equation (CVaR). The R.V. $-R$ has as its upper bound $a \in \mathbb{R}$ as per Definitions 7 and 8. Finally, as $N \geq \frac{\log(1-\gamma)}{\log(1-\epsilon)}$, $1-(1-\epsilon)^N \geq \gamma$. □

Likewise, we can also upper bound $\mathrm{EVaR}_\alpha(-R)\ \forall\ \alpha \in (0,1]$.

Corollary 6. *Let R be a scalar random variable as per Definition 8, let $\alpha \in (0,1]$, let L, u_b be as defined in equation (EVaR) with respect to this α, let $\epsilon \in (0,1)$ and $\gamma \in [0,1)$, and let $\zeta_N^*(\mu,t)$ be the solution to (UB-RP-N) for a set of N-samples $\{y_k = L(-r_k,\mu,t)\}_{k=1}^N$ of the random variable $Y = L(-R,\mu,t)$, where $N \geq \frac{\log(1-\gamma)}{\log(1-\epsilon)}$. Then,*

$$\mathbb{P}_{\pi_R}^N\left[r_E^* \triangleq \inf_{\mu\in\mathbb{R},\ t>0} \zeta_N^*(\mu,t)(1-\epsilon) + u_b(\mu,t)\epsilon \geq \mathrm{EVaR}_\alpha(-R)\right] \geq \gamma.$$

Proof. This is an application of Corollary 4 and follows in the footsteps of the proof for Corollary 5. □

4.3. Examples

To showcase the results of Corollary 5 and 6, we will verify a cooperative multi-agent robotic system. Specifically, we will identify lower bounds on expected worst-case system performance - *i.e.* upper bounds on both $\mathrm{CVaR}_{0.1}(-R)$ and $\mathrm{EVaR}_{0.1}(-R)$ - for a multi-agent system that is to avoid self-collisions while each agent reaches their respective goal. As a case study, we will use the Georgia Tech Robotarium, wherein all the robots can be modeled as unicycles [47]:

$$x = \begin{bmatrix} x, \\ y, \\ \theta \end{bmatrix},\ \dot{x} = \begin{bmatrix} v\cos(\theta), \\ v\sin(\theta), \\ \omega, \end{bmatrix},\ u = [v,\omega]^T, \qquad \mathcal{X} = [-1,1]\times[-0.6,0.6]\times[0,2\pi],\ M = [I_2,\ \mathbf{0}_{2x1}]. \tag{9}$$

As mentioned in [47], a Lyapunov-based controller drives each agent x^i to their desired orientation $x_d^i \in \mathcal{X}$, and control inputs are filtered in a barrier-based quadratic program to ensure that robots do not collide when multiple robots are moving simultaneously [41]. As a result, this barrier-based filter provides a natural robustness measure as per Definition 7. Let the state vector for all 3 robots $\mathbf{x}^T = [x^{1T}, x^{2T}, x^{3T}]$. Then, we can generate two, key functions. One function, h_g, that the system hopes to keep positive, and another function, h_f, that the system hopes to make positive over its trajectory. Here, x^i, x_d^i are the states and desired poses for robot i:

$$h_g(\mathbf{x}) = \min_{i\neq j,\ i,j\in[1,2,3]} \|M(x^i - x^j)\| - 0.15. \tag{10}$$

$$h_f(\mathbf{x}) = \max_{i\in[1,2,3]} 0.1 - \|M(x^i - x_d^i)\|. \tag{11}$$

Intuitively, $h_g(\mathbf{x}) \geq 0$ implies that the robots are sufficiently far apart, and $h_f(\mathbf{x}) \geq 0$ means that all robots are sufficiently close to their goal. From these two functions, we construct our robustness measure ρ for a state-signal $\mathbf{x}^\theta$ where $\theta^T = [x_d^{1T}, x_d^{2T}, x_d^{3T}]$:

$$\rho_g(\mathbf{x}^\theta) = \min_{t\in[0,30]} h_g\left(\mathbf{x}_t^\theta\right), \quad \rho_f(\mathbf{x}^\theta) = \max_{t\in[0,30]} h_f\left(\mathbf{x}_t^\theta\right),$$

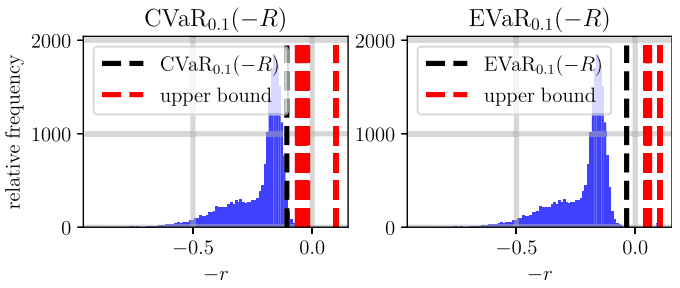


Fig. 7. As an application of the results in Section 3, we upper bound the risk measures of a multi-agent robotic system when its state trajectory is evaluated through a robustness metric (Definition 7). Shown above is this upper bounding procedure performed in Section 4.3 for both CVaR (left) and EVaR (right). For each of the 50 trials, the upper bounds (red) for both risk measures are indeed greater than or equal to their "true" counterparts (black). These "true" counterparts were calculated by taking 20000 samples of the randomized system robustness R (Definition 8), and the distribution of samples is shown (blue). The fact that the upper bounds are indeed upper bounds over all trials serves as a numerical confirmation of Corollaries 5 and 6. They also support the repeatability of our procedure in identifying upper bounds to g-entropic risk measures with high probability.

$$\rho(\mathbf{x}^\theta) = \begin{cases} \rho_g(\mathbf{x}^\theta), & \text{if } \rho_g(\mathbf{x}^\theta), \rho_f(\mathbf{x}^\theta) \geq 0, \\ \max\left\{\rho_g(\mathbf{x}^\theta), -0.1\right\}, & \text{if } \rho_g(\mathbf{x}^\theta) < 0, \\ \max\left\{\rho_f(\mathbf{x}^\theta), -0.1\right\}, & \text{else.} \end{cases} \tag{12}$$

If $\rho(\mathbf{x}^\theta) \geq 0$ then, all 3 robots stayed at least 0.15 meters from each other for 30 seconds - as $h_g(\mathbf{x}_t^\theta) \geq 0,\ \forall\, t \in [0, 30]$ - and reached within 0.1 meters of their goal within 30 seconds - as $h_f(\mathbf{x}_t^\theta) \geq 0$, for some $t \in [0, 30]$. Furthermore, we also know this robustness measure ρ outputs values that are always greater than or equal to -0.1. This robustness measure ρ also lets us specifically define our holistic system robustness R in this case. For this example, samples r of R arise when we first uniformly randomly sample initial locations and parameters $(\mathbf{x}_0, \theta)$ from their respective spaces below:

$$\mathcal{X}_0 = \{\mathbf{x} \in \mathcal{X}^3 \mid h_g(\mathbf{x}) \geq 0.3\}, \quad \Theta = \{\mathbf{x} \in \mathcal{X}^3 \mid h_g(\mathbf{x}) \geq 0.3\}.$$

Then, we record the corresponding state trajectory of the multi-agent system $\mathbf{x}^\theta$ for at-least 30 seconds. The holistic system robustness sample $r = \rho(\mathbf{x}^\theta)$ then, with ρ as defined in (12).

Verifying a 3 Robot System: Our goal is to determine an upper bound on both the $\mathrm{CVaR}_\alpha(-R)$ and $\mathrm{EVaR}_\alpha(-R)$ with $\alpha = 0.1$ for the three robots as they carry out their task. Furthermore, we hope to determine this upper bound with 95% confidence, *i.e.* $\gamma = 0.95$, and we require that $\epsilon = 0.02$. For this case, Corollaries 5 and 6 require that we take at minimum $N \geq 149$ samples of our holistic system robustness R to determine this upper bound. As a result, we uniformly sampled $N = 149$ initial conditions and parameters $(\mathbf{x}_0, \theta)$ from $\mathcal{X}_0 \times \Theta$, recorded the corresponding state trajectories $\mathbf{x}^\theta$, and recorded their robustnesses $r = \rho(\mathbf{x}^\theta)$. We also repeated this data collection procedure 50 times to ensure that our results are repeatable, and to identify the true $\mathrm{CVaR}_\alpha(-R)$ and $\mathrm{EVaR}_\alpha(-R)$ we repeated this data collection procedure once more but took $N = 20000$ samples instead. If Corollaries 5 and 6 are correct, then with 95% confidence, we expect that our sampled upper bounds for the 50 trials performed are greater than or equal to their true counterparts. All this data is shown in Fig. 7, and as prior, the black lines indicate the true risk measure evaluation and all red lines are the upper bounds generated per trial. As can be seen, all generated upper bounds are indeed upper bounds for their counterparts which serves as a numerical confirmation of Corollaries 5 and 6 and also of the repeatability of our procedure in identifying upper bounds to these risk measures.

Brief Remark on Relative Bound Tightness: Fig. 7 does prompt one question, however. Specifically, it appears the upper bounds for CVaR are tighter than those for EVaR. We anticipate this might arise as CVaR is a less conservative risk measure than EVaR. As a result, for a given confidence level γ in our upper bound, the upper bound for EVaR will likely be looser on account of this conservatism.

5. Identification of "good" solutions to optimization problems

The prior two sections detailed a sample-based approach to determine the probability with which a controlled system realizes a behavior of interest. Here, the uncertain quantities were the initial states and inputs available to the controlled system and the system's evolution itself. Similarly, by viewing policy synthesis as an optimization problem wherein the policy is now the uncertain quantity, the following two sections will build on the prior sampled-based approaches to develop a technique that outputs provably sub-optimal risk-aware policies with some minimum, user-specified confidence. To that end, this section will detail our efforts in developing a sample-based approach to identifying "good" solutions to a wide class of optimization problems. Then, Section 6 will phrase policy synthesis as one such optimization problem and make use of the results to be described.

5.1. Notation and setting

To simplify the setting, we assume we have a reward function $J : \mathbb{X} \subseteq \mathbb{R}^n \to \mathbb{R}$ for some positive integer n and general domain $\mathbb{X}$. Here, we note that typically this reward function is referred to via an R in the learning community. However, to prevent notational overload with the holistic system robustness variable R in Definition 8, we will use J as our reward function:

$$\max_{x \in D \subseteq \mathbb{X}} \; J(x). \tag{13}$$

For (13), the decision space D and reward function J satisfy the following assumption:

Assumption 3. The decision space D has bounded volume, *i.e.* $\int_D 1 \, dx = M < \infty$, and the reward function J attains its maximum value over D, *i.e.* $\exists \, x^* \in D$ such that $J(x^*) = J^* < \infty$ and $J(x^*) \geq J(x), \; \forall \, x \in D$.

This assumption is not too restrictive, as a large class of optimization problems falls under this setting, including risk-aware policy synthesis. Additionally, $\mathcal{V} : 2^D \to \mathbb{R}$ outputs the volume fraction of a subset A of D:

$$\mathcal{V}(A) = \frac{\int_A 1 \, dx}{\int_D 1 \, dx}. \tag{14}$$

Finally, let $F : D \to 2^D$ be a function that identifies the space of decisions $x' \in D$ that are "better" than the provided decision x, *i.e.*

$$F(x) = \{x' \in D \mid J(x') > J(x)\}. \tag{15}$$

Then this section considers the following problem.

Problem 3. For any $\epsilon \in (0, 1)$ devise a method to find a decision $x \in D$ such that x is at least in the $100(1-\epsilon)$-th percentile of all possible decisions $x \in D$ with respect to the reward function J, *i.e.* find a decision x such that $\mathcal{V}(F(x)) \leq \epsilon$, with $\mathcal{V}$ defined in equation (14) and F defined in equation (15).

5.2. Sampling methods for identification of "good" decisions

First, we note that for a uniform distribution over the decision space D, the probability of sampling a decision $x \in A \subset D$ is equivalent to the volume fraction of A with respect to D.

Lemma 4. *Let Assumption 3 hold. The following is true with $\mathcal{V}$ as in (14):*

$$\mathbb{P}_{\mathrm{U}[D]}[x \in A \subseteq D] = \mathcal{V}(A).$$

Proof. As the distribution is uniform,

$$\mathbb{P}_{\mathrm{U}[D]}[x \in A] = \frac{\int_A 1 \, dx}{\int_D 1 \, dx} = \mathcal{V}(A). \quad \square$$

Second, we note that (13) can be recast as an uncertain program similar to (UP). Specifically, let X be a random variable whose distribution is the uniform distribution over D, *i.e.* $\mathrm{U}[D]$. Then, the corresponding random reward is $Y = J(X)$. This random reward variable Y has its own samples y and (unknown) distribution π_Y, letting us construct an uncertain program of the form in (UP), with respect to samples y of Y:

$$\begin{aligned} \zeta^* = \operatorname{argmin}_{\zeta \in \mathbb{R}} \quad & \zeta, \\ \text{subject to} \quad & \zeta \geq y, \; y \in \mathbb{R}. \end{aligned} \tag{UP-G}$$

This uncertain program (UP-G) also has an analogous scenario program:

$$\begin{aligned} \zeta_N^* = \operatorname{argmin}_{\zeta \in \mathbb{R}} \quad & \zeta, \\ \text{subject to} \quad & \zeta \geq y_i, \; y_i \in \{y_k = J(x_k)\}_{k=1}^N, \; X \sim \mathrm{U}[D] \text{ with samples } x. \end{aligned} \tag{RP-G}$$

This scenario program is crucial to our approach for two reasons which we will prove in the theorem to follow. First, there exists a sampled decision x_i in the set of all sampled decisions $\{x_k\}_{k=1}^N$ such that the reward of this decision is equivalent to the scenario solution ζ_N^*, *i.e.* $J(x_i) = \zeta_N^*$. Second, via Theorem 1 we can upper bound the probability of sampling another decision $x' \in D$ such that $J(x') > \zeta_N^* = J(x_i)$. Via Lemma 4, such a probability corresponds to the volume fraction of those decisions $x' \in D$ that are "better" than x_i, *i.e.* this probability corresponds to $\mathcal{V}(F(x_i))$. As a result, to determine a decision $x \in D$ that is "better" than a predetermined volume fraction $\epsilon \in (0, 1)$ of all possible decisions $x' \in D$ with confidence $\gamma \in [0, 1)$, we just need to take enough samples of X and solve (RP-G).

Theorem 6. *Let $\epsilon \in (0,1)$, let $\gamma \in [0,1)$, let Assumption 3 hold, let ζ_N^* be the solution to (RP-G) for an N-sample set $\{x_k\}_{k=1}^N$ of X, let $\mathcal{V}$ be as defined in (14), and let F be as defined in (15). If $N \geq \frac{\log(1-\gamma)}{\log(1-\epsilon)}$ then with minimum probability γ, there exists at least one sampled decision $x_i \in \{x_k\}_{k=1}^N$ that is at-least in the $100(1-\epsilon)$-th percentile of all possible decisions $x \in D$, i.e.,*

$$\exists\ x_i \in \{x_k\}_{k=1}^N \text{ s.t. } \mathbb{P}^N_{\mathrm{U}[D]}[\mathcal{V}(F(x_i)) \leq \epsilon] \geq \gamma, \text{ and } \zeta_N^* = J(x_i).$$

Proof. First, for any finite sample set $\{x_k\}_{k=1}^N$ there exists a decision $x_i \in \{x_k\}_{k=1}^N$ such that $\zeta_N^* = J(x_i)$. This is due to the fact that (RP-G) is a linear program minimizing a scalar decision variable subject to a series of lower bounds that take values in $\mathbb{R}$. As a result, ζ_N^* must equal one of its lower bounds to be a valid solution to (RP-G), and as such, it must be equal to $J(x_i)$ for at least one $x_i \in \{x_k\}_{k=1}^N$.

Then via Theorem 1, the following inequality holds for our scenario solution ζ_N^*, where the violation probability $V(\zeta_N^*) = \mathbb{P}_{\mathrm{U}[D]}[J(x) > \zeta_N^*]$:

$$\mathbb{P}^N_{\mathrm{U}[D]}[\mathbb{P}_{\mathrm{U}[D]}[J(x) > \zeta_N^*] \leq \epsilon] \geq 1 - (1-\epsilon)^N. \tag{16}$$

By definition of F in (15), (16) can be recast as follows for some $x_i \in \{x_k\}_{k=1}^N$:

$$\mathbb{P}^N_{\mathrm{U}[D]}[\mathbb{P}_{\mathrm{U}[D]}[x \in F(x_i)] \leq \epsilon] \geq 1 - (1-\epsilon)^N.$$

Via Lemma 4 we can replace the inner probability with $\mathcal{V}(F(x_i))$:

$$\mathbb{P}^N_{\mathrm{U}[D]}[\mathcal{V}(F(x_i)) \leq \epsilon] \geq 1 - (1-\epsilon)^N.$$

Then, as $N \geq \frac{\log(1-\gamma)}{\log(1-\epsilon)}$, $\epsilon \in (0,1)$ and $\gamma \in [0,1)$, $1-(1-\epsilon)^N \geq \gamma$. □

To summarize, Theorem 6 states that if we want to find a "good" solution to (13) with minimum probability γ - "good" insofar as it is in the $100(1-\epsilon)$-th percentile of all decisions $x \in D$ for maximizing J - that we are required to evaluate at minimum $N \geq \frac{\log(1-\gamma)}{\log(1-\epsilon)}$ uniformly randomly chosen decisions $x \in D$. If we evaluate the reward function J for each such sampled decision in our sample set $\{x_k\}_{k=1}^N$, at least one decision $x_i \in \{x_k\}_{k=1}^N$ is guaranteed, with minimum probability γ, to produce a reward $J(x_i)$ that is in the $100(1-\epsilon)$-th percentile of all possible rewards achievable. Hence, this decision x_i that produces this reward $J(x_i)$ is also guaranteed, with minimum probability γ, to be in the $100(1-\epsilon)$-th percentile of all possible decisions $x \in D$ with respect to the reward function J.

5.3. Examples

This section provides a few examples of applying Theorem 6 to finding paths for the traveling salesman problem (TSP) that are in the 99-th percentile of all possible paths achievable [48]. Specifically, TSP references the identification of a permutation of integers $P = [i_1, i_3, i_6, \dots]$ where such a permutation minimizes the summation of a function $d : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \mathbb{R}$. That is, if we define a finite set of integers $I \subset \mathbb{Z}_+$, we can also define a set $P \subset 2^I$ of all permutations of integers in I, *i.e.* $P = \{p = (i_1, \dots, i_{|I|}) \mid \forall\ i_j \in I,\ i_j \in p$ and i_j appears only once in $p\}$. Then, the optimization problem is:

$$\min_{s \in P} \sum_{k=1}^{|I|-1} d(s_k, s_{k+1}) + d(s_{|I|}, s_1). \tag{17}$$

While phrased as a minimization, (17) can be recast as a maximization problem of the form in (13) - the reward function J is the negation of the objective function above, and the decision space $D = P$. Furthermore, this problem satisfies Assumption 3, with $\int_P 1\ dx = |P| < \infty$, which means that we should be able to follow Theorem 6 to identify a "good" solution to this problem.

Specifically then, we will randomly generate 9 nodes $n_i \in [-5,5]^2 \subset \mathbb{R}^2$. Our cost function $d(i,j) = \|n_i - n_j\|$. As we have generated 9 nodes, the cardinality of the set of all possible permutations $|P| = 9! = 362880$. However, according to Theorem 6, if we want a permutation $p \in P$ that is in the 99-th percentile with 95% confidence, then we need to uniformly sample $N \geq \frac{\log(1-\gamma=0.05)}{\log(1-\epsilon=0.99)} \approx 299$ permutations $p \in P$ and evaluate their corresponding costs. Doing this procedure once, we identified a permutation $p^* \in P$ that was in the 99.9997-th percentile of all paths as shown on the left in Fig. 8. This one case shows the utility of our sample approach in providing "good" solutions to difficult optimization problems insofar as we only had to evaluate $299/362880 \approx 0.0008$ of all possible paths to generate a very good one. Indeed, the method also repeatably identifies "good" solutions as shown on the right in Fig. 8. In this case, we required a path in the 95-th percentile with minimum probability $1 - 10^{-6}$ for a case where we randomly generated 8 nodes $n_i \in [-5,5]^2 \subset \mathbb{R}^2$. In every case, the identified path p^* was in the 95-th percentile as evidenced by $\mathcal{V}(F(p^*)) \leq \epsilon$.

6. Risk-aware policy synthesis

Finally, we can now combine the sample-based approaches developed in all prior sections. Specifically, we will phrase risk-aware policy synthesis as an optimization problem where the objective function accounts for uncertainty in initial system states, controller

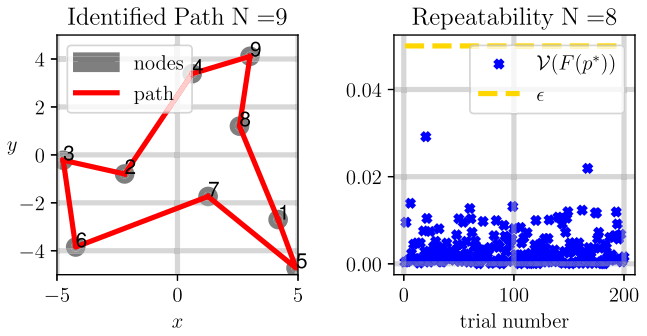


Fig. 8. Risk-Aware Policy Synthesis (Section 6) requires a method to identify good solutions to potentially non-convex optimization problems - this is the procedure detailed in Section 5. Shown above is an application of this procedure to identify "good" paths for a traveling salesman problem [48]. By uniformly sampling $N = 299$ paths from the set of all possible paths P with $|P| = 362880$, we can identify (left) a path that is in the 99.9997-th percentile of all paths. This procedure also repeatably identifies "good" paths, *a.k.a* "good" decisions, as shown in the figure on the right. If we take the minimum number of samples offered by Theorem 6 to identify a path that is in the 95-th percentile with minimum probability $1 - 10^{-6}$, we see that over 200 trials - taking $N = 270$ samples each time - all determined paths are in the 95-th percentile as $\mathcal{V}(F(p^*)) \leq \epsilon = 0.05$.

inputs, and system evolution, and the design space accounts for the uncertainty in the optimal policy. To "solve" this optimization problem, we will use the sample-based approach to identifying "good" policies as developed in Section 5 where each policy sample taken here is the result of our sample-based verification approach detailed in Section 4. As a result, we detail a risk-aware policy synthesis technique with tractable sample requirements, that easily scales to high-dimensional systems and potentially non-convex policy optimization objectives.

6.1. Setting and problem statement

Specifically, we still focus on a general class of systems with the state space, $\mathcal{X}$; the input space, $\mathcal{U}$; a space of problem parameters Θ that the controller U must account for. The discrepancy with (8) arises in the parameterization of the controller U via the parameter space P of controller parameters. As before, we assume the system is subject to some stochastic noise ξ with an unknown distribution $\pi_{\bar{\xi}}(x,u,t)$ over $\mathbb{R}^n$.

$$\begin{aligned} &\dot{x} = f(x,u) + \xi, && x \in \mathcal{X} \subset \mathbb{R}^n,\ u \in \mathcal{U} \subset \mathbb{R}^m, \\ &u = U(x,\theta,p), && \theta \in \Theta \subset \mathbb{R}^l,\ p \in P \subset \mathbb{R}^s, \\ &\xi \sim \pi_{\bar{\xi}}(x,u,t), && \int_{\mathcal{X}} \pi_{\bar{\xi}}(x,u,t,s)\, ds = 1\ \forall\ x,u,t. \end{aligned} \tag{18}$$

As before, we will define $x_t^{\theta,p}$ as the solution to our closed-loop system at time t, given the initial condition $x_0 \in \mathcal{X}_0 \subseteq \mathcal{X}$, an accounting parameter $\theta \in \Theta$, and a set of controller parameters $p \in P$.

$$\dot{x}_t^{\theta,p} = f\left(x_t^{\theta,p}, U\left(x_t^{\theta,p}, \theta, p\right)\right) + \xi,\ x^{\theta,p} \in \mathcal{S}^{\mathbb{R}^n},\ \mathcal{S}^{\mathbb{R}^n} = \{s : \mathbb{R}_{\geq 0} \to \mathbb{R}^n\}.$$

Then, our goal will be to identify a set of parameters $p \in P$ such that the ensuing controller U is better than, say, 90% of any possible controller we could develop, with respect to maximizing the lower-bound on robustness risk of the closed-loop system - the lower bound outputted by the sample-based risk-aware verification approach detailed in Section 4. To be specific, choosing a parameter $p \in P$ defines a specific closed-loop system - that system for which the controller U is parameterized via this parameter p. Then for this system, we can calculate a lower bound on the Conditional-Value-at-Risk of the system's robustness ρ over the entire space of initial conditions and accounting parameters $\mathcal{X}_0 \times \Theta$ - this is Corollary 5. We could follow a similar process for upper bounding the Entropic-Value-at-Risk as well. However, we will focus on CVaR for brevity. As such, the natural risk-aware policy synthesis aim is to maximize this lower bound.

Formally stating this problem first requires a modification of Definition 8 for the parameterized setting.

Definition 9. The *parameterized holistic system robustness*, R_p, is a scalar random variable with distribution π_{R_p} and samples r_p denoting the closed-loop robustness $\rho(x^{\theta,p})$ of trajectories $x^{\theta,p}$ whose initial condition and parameter (x_0, θ) were sampled uniformly from $\mathcal{X}_0 \times \Theta$ and whose controller U is parameterized via the controller parameter $p \in P$ as per equation (18).

Per Corollary 5, if we fix the parameter p for the controller U, we can generate an upper bound $r^*_{C,p}$ on $\text{CVaR}_\alpha(-R_p)$ for any $\alpha \in (0,1]$ with high-confidence. As we study only CVaR in this section, we will drop the subscript C for $r^*_{C,p}$ to simplify notation. Mathematically, this lets us generate a mapping $\mathcal{R}$ from controller parameter p, confidence γ, and risk-level α, to the upper bound r^*_p generated with confidence γ for $\text{CVaR}_\alpha(-R_p)$. Implicit in this function definition will be the number of samples $N_{\mathcal{R}}$ we take of the initial conditions and accounting parameters (x_0, θ) from the uniform distribution over $\mathcal{X}_0 \times \Theta$:

$$\mathcal{R}(p,\gamma,\alpha) = r_p^*, \text{ s.t. } \mathbb{P}^{N_{\mathcal{R}}}_{\pi_{R_p}}\left[r_p^* \geq \text{CVaR}_\alpha(-R_p)\right] \geq \gamma. \tag{19}$$

Then, by Definitions 4 and 5, we have the following proposition indicating that the risk map $\mathcal{R}$ serves as a high-confidence lower bound on the expected worst-case system performance.

Proposition 1. *Let $\mathcal{R}$ be as defined in equation (19) for some $\gamma \in [0,1)$ and $\alpha \in (0,1]$. The following inequality holds, with π_{R_p} the distribution of the parameterized holistic system robustness R_p defined in Definition 9:*

$$\mathbb{P}^N_{\pi_{R_p}}\left[-\mathcal{R}(p,\gamma,\alpha) \leq \mathbb{E}_{\pi_{R_p}}[r \mid r \leq \text{VaR}_{1-\alpha}(R_p)]\right] \geq \gamma.$$

Proof. By definition of CVaR_α for any risk-level $\alpha \in (0,1]$ in Definition 5, we have the following equivalency:

$$\text{CVaR}_\alpha(-R_p) = \mathbb{E}_{\pi_{R_p}}[r_p \mid -r_p \geq \text{VaR}_\alpha(-R_p)].$$

Then by definition of VaR and a symmetry argument when flipping the random variable of interest, we have the following equality:

$$-\text{CVaR}_\alpha(-R_p) = \mathbb{E}_{\pi_{r_p}}[r_p \mid r_p \leq \text{VaR}_{1-\alpha}(R_p)].$$

Then the desired inequality stems from the definition of the risk map $\mathcal{R}$ in equation (19). □

As a result, we state that the goal of policy synthesis should be to minimize $\mathcal{R}(p,\gamma,\alpha)$. Specifically, minimization of $\mathcal{R}(p,\gamma,\alpha)$ over P for some fixed γ,α corresponds to the identification of a set of controller parameters $p \in P$ such that in the worst $100\alpha\%$ of cases, the expected minimum system robustness is maximized. Since the robustness measure ρ outputs positive numbers for those trajectories that satisfy the desired behavior (Definition 7), if this maximum lower bound were positive, then the expected performance in the worst $100\alpha\%$ of cases would still exhibit satisfactory behavior. This leads to the nominal, risk-aware synthesis problem:

$$\min_{p\in P} \mathcal{R}(p,\gamma,\alpha), \text{ for some } \gamma,\alpha \in [0,1), \tag{20}$$

which is similar to the general problem (13) studied in Section 5. The decision space $D = P$ and the reward function $R = -\mathcal{R}(\gamma,\alpha)$. We can also rewrite the volume fraction function $\mathcal{V}$ (14) and "better" policy map F (15):

$$\mathcal{V}(A) = \frac{\int_A 1\, dx}{\int_P 1\, dx}, \quad F(p,\gamma,\alpha) = \left\{p' \in P \mid \mathcal{R}(p',\gamma,\alpha) < \mathcal{R}(p,\gamma,\alpha)\right\}. \tag{21}$$

This lets us formally define the problem under study for this section.

Problem 4. Let $\mathcal{R}$ be as defined in (19) with respect to some $\alpha \in (0,1]$ and $\gamma \in (0,1)$. Let $\epsilon \in (0,1)$ as well. Find a set of controller parameters $p \in P$ such that the corresponding controller U is at-least in the $100(1-\epsilon)$-th percentile of all possible controllers with respect to minimization of $\mathcal{R}$, *i.e.* find $p \in P$ such that $\mathcal{V}(F(p)) \leq \epsilon$ where $\mathcal{V}, F$ are defined in (21).

6.2. Identification of "good" risk-aware policies

To start, Problem 4 is a refinement of Problem 3. As such, we start by constructing a similar scenario program to (RP-G):

$$\begin{aligned} \zeta_N^* = \text{argmin}_{\zeta\in\mathbb{R}} \quad & \zeta, \\ \text{subject to} \quad & \zeta \geq y_i,\ y_i \in \left\{y_k = \mathcal{R}(p_k,\gamma,\alpha)\right\}_{k=1}^N, \\ & p_k \text{ are drawn uniformly from } P. \end{aligned} \tag{RP-PS}$$

Then, identification of "good" policies is a direct consequence of Theorem 6, as stated in the following corollary.

Assumption 4. $\mathcal{R}$ is as defined in equation (19) with respect to some $\alpha \in (0,1]$ and $\gamma_1 \in [0,1)$. $\epsilon \in (0,1)$, $\gamma_2 \in [0,1)$, ζ_N^* is the solution to (RP-PS) for a set of N-samples $\{y_k = \mathcal{R}(p_k,\gamma_1,\alpha)\}_{k=1}^N$ where each p_k was drawn uniformly from P, and $\mathcal{V}, F$ are as defined in equation (21).

Corollary 7. *Let Assumption 4 hold. If $N \geq \frac{\log(1-\gamma_2)}{\log(1-\epsilon)}$, then with minimum probability γ_2 there exists at least one parameter set $p_i \in \{p_k\}_{k=1}^N$ whose associated controller U is at-least in the $100(1-\epsilon)$-th percentile of controllers possible with respect to minimizing $\mathcal{R}$, i.e.*

$$\exists\ p_i \in \{p_k\}_{k=1}^N, \text{ s.t. } \mathbb{P}^N_{\text{U}[P]}\left[\mathcal{V}(F(p_i)) \leq \epsilon\right] \geq \gamma_2, \text{ and } \zeta_N^* = \mathcal{R}(p_i,\gamma_1,\alpha).$$

Proof. This is a direct application of Theorem 6. □

To summarize, for each uniformly sampled controller parameterization $p \in P$, we evaluate the corresponding controller U against $N_{\mathcal{R}}$ uniformly sampled initial conditions and accounting parameters $(x_0, \theta) \in \mathcal{X}_0 \times \Theta$ and evaluate the robustness $\rho(x^{\theta,p})$ of the corresponding trajectories. By evaluating enough such trajectories, we generate - via Corollary 5 and Proposition 1 - a lower bound $-r^*$ on the expected worst-case system performance in the worst $100\alpha\%$ of cases. By repeating this procedure for enough samples - $N \geq \frac{\log(1-\gamma_2)}{\log(1-\epsilon)}$ samples - we are guaranteed with minimum probability γ_2 to find at least one parameter set p_i in the sampled set $\{p_k\}_{k=1}^N$ such that the corresponding controller U is at-least in the $100(1-\epsilon)$-th percentile of all possible controllers with respect to maximizing this lower bound. This probabilistic guarantee is due to Theorem 6. As we also provide a minimum sample requirement on the number of controllers required to be tested in this procedure, we state that we have a fundamental sample requirement to generate "good" risk-aware policies with high confidence. Furthermore, as this sample requirement is based on both our desired confidence γ and desired percentile level $1-\epsilon$, we state that this procedure also identifies the relative complexity in the identification of a "better" policy, insofar as it provides a sample requirement for doing so.

6.3. Examples

In this section, we will synthesize a risk-aware controller for the cooperative multi-agent system whose controller was verified in Section 4.3, and we will show that the synthesized controller outperforms the baseline controller when accounting for worst-case system performance. To reiterate, we will use the robotarium as a case study again, where the robots can be modeled as unicycles [47]. The dynamics are the same as in (9) in Section 4.3.

Different from the verification case in Section 4.3, however, we will now assume that the hybrid Lyapunov-barrier controller the robots are equipped with [47] is now to be optimized. Since we only assume knowledge of the controller parameters p and not the specific controller form U, we will refrain from specifying the controller in favor of stating that the controller parameters we can vary are the gains for the Lyapunov controller - approach angle gain p_1; desired angle gain p_2; rotation error gain p_3; and decay constant p_4 for the barrier filter. As such, the space of parameters

$$P = [0.2, 5]^3 \times [0.1, 200], \; p_{1,2,3} \in [0.2, 5]^3, \; p_4 \in [0.1, 200].$$

To determine a "good" parameter set $p \in P$ without knowledge of how these parameters affect the controller U, we require a robustness measure ρ (Definition 7) to act as a determiner of satisfactory system performance. As such, we will use the same robustness measure ρ in equation (12) in Section 4.3 to facilitate a risk-aware verification comparison between our to-be-calculated controller and the baseline controller for the Robotarium. To be clear, we will reproduce this robustness measure ρ in its application to our parameterized system trajectories $\mathbf{x}^{\theta,p}$, where h_g, h_f are as defined in equations (10) and (11) respectively:

$$\rho_g(\mathbf{x}^{\theta,p}) = \min_{t\in[0,30]} h_g\left(\mathbf{x}_t^{\theta,p}\right), \quad \rho_f(\mathbf{x}^{\theta,p}) = \max_{t\in[0,30]} h_f\left(\mathbf{x}_t^{\theta,p}\right),$$

$$\rho(\mathbf{x}^{\theta,p}) = \begin{cases} \rho_g(\mathbf{x}^{\theta,p}), & \text{if } \rho_g(\mathbf{x}^{\theta,p}), \rho_f(\mathbf{x}^{\theta,p}) \geq 0, \\ \max\left\{\rho_g(\mathbf{x}^{\theta,p}), -0.1\right\}, & \text{if } \rho_g(\mathbf{x}^{\theta,p}) < 0, \\ \max\left\{\rho_f(\mathbf{x}^{\theta,p}), -0.1\right\}, & \text{else.} \end{cases}$$

Then, to take a sample r_p of our randomized robustness R_p as per Definition 9, we first uniformly sample (x_0, θ) from the spaces below:

$$\mathcal{X}_0 = \{\mathbf{x} \in \mathcal{X}^3 \mid h_g(\mathbf{x}) \geq 0.3\}, \quad \Theta = \{\mathbf{x} \in \mathcal{X}^3 \mid h_g(\mathbf{x}) \geq 0.3\}.$$

Then we evaluate the robustness of the corresponding state trajectory, *i.e.* $r_p = \rho(\mathbf{x}^{\theta,p})$. Our goal is to identify a parameter set $p \in P$ that minimizes the upper bound r_p^* for $\text{CVaR}_{\alpha=0.1}(-R_p)$ identified with 95% confidence with $N_{\mathcal{R}} = 149$ samples. By definition of $\mathcal{R}$ in equation (19), this results in the following optimization problem, reminiscent of (20)

$$\min_{p\in P} \mathcal{R}(p, 0.95.0.1), \; N_{\mathcal{R}} = 149.$$

As such, we hope to identify a policy that is in the 99-th percentile with 99% confidence, *i.e.* $\gamma = 0.99$ and $\epsilon = 0.01$.

According to Corollary 7, to generate a parameter set $p_i \in P$ whose controller U is in the 99-th percentile of all controllers with minimum probability $\gamma = 0.99$, we are required to test $N \geq 459$ uniformly randomly sampled parameters $p \in P$. After doing so, we identified a parameter set $p_i \in P$ that realized an upper bound $\mathcal{R}(p_i, 0.95, 0.1) = -0.1489$. As can be seen in Fig. 9, the controller corresponding to this policy is indeed in at least the 99-th percentile with respect to all controllers achievable via this parameterization. This serves as a numerical confirmation of Corollary 7 and its parent, Theorem 6. We can also compare the distribution of the robustness of trajectories realized by this controller U_{calc} as compared to the baseline controller with which the robotarium comes equipped U_{base}. This information can be seen in Fig. 10. Notice that our identified controller outperforms the baseline robotarium controller for our robustness measure of interest.

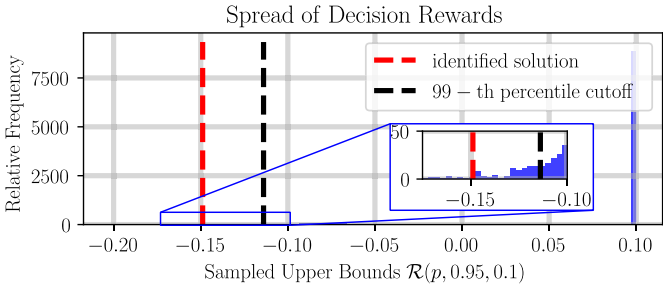


Fig. 9. The final use case of our sample-based bounds arises in Risk-Aware Policy Synthesis. Our goal is to identify a parameterized controller that maximizes a lower bound on worst-case system performance, *i.e.* minimizes an upper bound, $\mathcal{R}(p, 0.95, 0.1)$, over the set of controller parameters $p \in P$ (20). Shown above in blue is the distribution of this upper bound for all controller parameters $p \in P$ and was generated by taking 20000 uniform parameter samples p and evaluating $\mathcal{R}(p, 0.95, 0.1)$. As per the decision selection process detailed in Section 5, our goal is to identify a controller in the 99-th percentile with respect to minimization of this upper bound with the true 99-th percentile cutoff shown in black and all controllers yielding upper bounds to its left lying in the 99-th percentile. As can be seen, our identified solution (red) achieves an upper bound in at least the 99-th percentile. This serves as a numerical confirmation of both Theorem 6 and Corollary 7 insofar as we evaluated the minimum number of controllers prescribed, $N = 459$ controllers, to calculate our solution which meets our desired criteria.

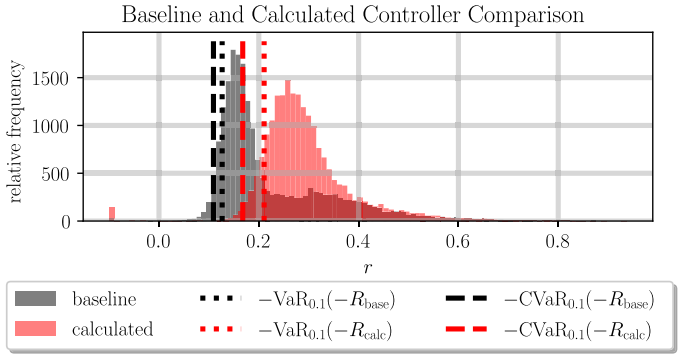


Fig. 10. A comparison between the baseline controller provided with the robotarium - the controller that was probabilistically verified in Section 4.3 - and our calculated controller identified in Fig. 9. Our risk-aware policy synthesis goal is to identify a controller in the 99-th percentile with respect to maximizing the lower bound on the expected worst-case system performance in the worst 10% of cases - *i.e.* maximizes a lower bound on $-\text{CVaR}_{0.1}(-R)$ as expressed in Proposition 1. Shown above is the distribution of this randomized robustness R for the calculated controller (R_{calc} in red) and the baseline controller provided with the robotarium (R_{base} black). As can be seen, the calculated controller outperforms the baseline controller insofar as the worst-case robustness value for 10% of cases $-\text{VaR}_{0.1}(-R_{\text{calc}}) \geq -\text{VaR}_{0.1}(-R_{\text{base}})$, and the expected value in the worst 10% of cases $-\text{CVaR}_{0.1}(-R_{\text{calc}}) \geq -\text{CVaR}_{0.1}(-R_{\text{base}})$ as well.

7. Conclusion and future work

To develop a systematic procedure for risk-aware verification, we first developed a procedure to generate sample-based bounds for g-entropic risk measures. Then, we reframed risk-aware verification as a risk-measure identification problem employing our prior results. To utilize these results to simplify risk-aware policy generation then, we developed a sample-based approximation for a general class of optimization problems. This approximation lets us repeatably and reliably identify decisions that outperform a large fraction of all possible decisions available with respect to the optimization of a provided objective. The prior sample-based bounds then provide the objective function for the risk-aware policy synthesis setting, with a "good" policy determined via the prior, sample-based approach.

In future work, we will analyze the tightness of the sample-based bounds we generate, and move towards generating sample-based concentration inequalities for g-entropic risk measures. Second, we aim to apply these tighter bounds to decrease conservatism in our sample-based verification approach. In addition, we aim to reduce the number of samples required for the generation of these bounds to facilitate the generation of high-confidence performance bounds on hardware. Finally, we aim to extend the risk-aware synthesis procedure for controller generation on hardware systems, for which we require a method to ensure that we only sample safe controllers. Furthermore, it is not entirely clear that a "good" risk-aware simulator controller will achieve similarly "good" performance on hardware. Here, we think that works addressing the Sim2Real gap might prove useful [49,50].

CRediT authorship contribution statement

Prithvi Akella: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Anushri Dixit:** Conceptualization, Formal analysis, Writing – review & editing. **Mohamadreza Ahmadi:** Conceptualization, Formal analysis, Supervision, Writing – review & editing. **Joel W. Burdick:** Formal analysis, Supervision, Validation, Writing – review & editing. **Aaron D. Ames:** Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by the AFOSR, grant FA9550-19-1-0302.

References

[1] T.P. Lillicrap, J.J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning, arXiv preprint arXiv:1509.02971, 2015.
[2] L.P. Kaelbling, M.L. Littman, A.W. Moore, Reinforcement learning: a survey, J. Artif. Intell. Res. 4 (1996) 237–285.
[3] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT Press, 2018.
[4] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller, Playing atari with deep reinforcement learning, arXiv preprint arXiv:1312.5602, 2013.
[5] D.P. Bertsekas, J.N. Tsitsiklis, Neuro-dynamic programming: an overview, in: Proceedings of 1995 34th IEEE Conference on Decision and Control, vol. 1, IEEE, 1995, pp. 560–564.
[6] M.L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, 1st edition, John Wiley & Sons, Inc., USA, 1994.
[7] G.E. Monahan, State of the art—a survey of partially observable Markov decision processes: theory, models, and algorithms, Manag. Sci. 28 (1) (1982) 1–16.
[8] S. Bhattacharya, S. Badyal, T. Wheeler, S. Gil, D. Bertsekas, Reinforcement learning for pomdp: partitioned rollout and policy iteration with application to autonomous sequential repair problems, IEEE Robot. Autom. Lett. 5 (3) (2020) 3967–3974.
[9] L.T. Dung, T. Komeda, M. Takagi, Reinforcement learning for pomdp using state classification, Appl. Artif. Intell. 22 (7–8) (2008) 761–779.
[10] S. Png, J. Pineau, Bayesian reinforcement learning for pomdp-based dialogue systems, in: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2011, pp. 2156–2159.
[11] H.A. Taha, Operations Research: an Introduction, vol. 790, Pearson/Prentice Hall, Upper Saddle River, NJ, USA, 2011.
[12] A. Majumdar, M. Pavone, How should a robot assess risk? Towards an axiomatic theory of risk in robotics, in: Robotics Research, Springer, 2020, pp. 75–84.
[13] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risk, Math. Finance 9 (3) (1999) 203–228.
[14] BCBS, Fundamental review of the trading book: a revised market risk framework, 2013.
[15] S. Singh, Y. Chow, A. Majumdar, M. Pavone, A framework for time-consistent, risk-sensitive model predictive control: theory and algorithms, IEEE Trans. Autom. Control 64 (7) (2018) 2905–2912.
[16] A. Hakobyan, G.C. Kim, I. Yang, Risk-aware motion planning and control using cvar-constrained optimization, IEEE Robot. Autom. Lett. 4 (4) (2019) 3924–3931.
[17] M. Ahmadi, A. Dixit, J.W. Burdick, A.D. Ames, Risk-averse stochastic shortest path planning, arXiv preprint arXiv:2103.14727, 2021.
[18] M. Heger, Consideration of risk in reinforcement learning, in: Machine Learning Proceedings 1994, Elsevier, 1994, pp. 105–111.
[19] Y. Chow, M. Ghavamzadeh, L. Janson, M. Pavone, Risk-constrained reinforcement learning with percentile risk criteria, J. Mach. Learn. Res. 18 (1) (2017) 6070–6120.
[20] O. Mihatsch, R. Neuneier, Risk-sensitive reinforcement learning, Mach. Learn. 49 (2) (2002) 267–290.
[21] P. Geibel, F. Wysotzki, Risk-sensitive reinforcement learning applied to control under constraints, J. Artif. Intell. Res. 24 (2005) 81–108.
[22] W. Korb, D. Engel, R. Boesecke, G. Eggers, R. Marmulla, N. O'Sullivan, J. Raczkowsky, S. Hassfeld, Risk Analysis for a Reliable and Safe Surgical Robot System, International Congress Series, vol. 1256, Elsevier, 2003, pp. 766–770.
[23] F. Vicentini, M. Askarpour, M.G. Rossi, D. Mandrioli, Safety assessment of collaborative robotics through automated formal verification, IEEE Trans. Robot. 36 (1) (2019) 42–61.
[24] R. Inam, K. Raizer, A. Hata, R. Souza, E. Forsman, E. Cao, S. Wang, Risk assessment for human-robot collaboration in an automated warehouse scenario, in: 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), vol. 1, IEEE, 2018, pp. 743–751.
[25] E. ISO, 10218: Robots and robotic devices-safety requirements for industrial robots-part 1: Robots, ISO, Geneve, Switzerland, 2011.
[26] I. ISO, 10218-2: 2011: Robots and robotic devices–safety requirements for industrial robots–part 2: Robot systems and integration, vol. 3, International Organization for Standardization, Geneva, Switzerland, 2011.
[27] A. Corso, R.J. Moss, M. Koren, R. Lee, M.J. Kochenderfer, A survey of algorithms for black-box safety validation, arXiv:e-prints, arXiv-2005, 2020.
[28] J. Deshmukh, M. Horvat, X. Jin, R. Majumdar, V.S. Prabhu, Testing cyber-physical systems through Bayesian optimization, ACM Trans. Embed. Comput. Syst. 16 (5s) (2017) 1–18.
[29] G.E. Mullins, P.G. Stankiewicz, R.C. Hawthorne, S.K. Gupta, Adaptive generation of challenging scenarios for testing and evaluation of autonomous vehicles, J. Syst. Softw. 137 (2018) 197–215.
[30] A. Corso, P. Du, K. Driggs-Campbell, M.J. Kochenderfer, Adaptive stress testing with reward augmentation for autonomous vehicle validatio, in: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE, 2019, pp. 163–168.
[31] M. Koren, M.J. Kochenderfer, Adaptive stress testing without domain heuristics using go-explore, in: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2020, pp. 1–6.
[32] P. Thomas, E. Learned-Miller, Concentration inequalities for conditional value at risk, in: International Conference on Machine Learning, in: PMLR, 2019, pp. 6225–6233.
[33] D.B. Brown, Large deviations bounds for estimating conditional value-at-risk, Oper. Res. Lett. 35 (6) (2007) 722–730.
[34] Z. Mhammedi, B. Guedj, R.C. Williamson, Pac-Bayesian bound for the conditional value at risk, Adv. Neural Inf. Process. Syst. 33 (2020) 17919–17930.
[35] A. Kagrecha, J. Nair, K. Jagannathan, Distribution oblivious, risk-aware algorithms for multi-armed bandits with unbounded rewards, arXiv preprint arXiv:1906.00569, 2019.
[36] M.C. Campi, S. Garatti, The exact feasibility of randomized solutions of uncertain convex programs, SIAM J. Optim. 19 (3) (2008) 1211–1230.
[37] A. Ahmadi-Javid, Entropic value-at-risk: a new coherent risk measure, J. Optim. Theory Appl. 155 (3) (2012) 1105–1123.
[38] W. Fenchel, On conjugate convex functions, in: Traces and Emergence of Nonlinear Programming, Springer, 2014, pp. 125–129.
[39] P. Akella, M. Ahmadi, A.D. Ames, A scenario approach to risk-aware safety-critical system verification, arXiv preprint arXiv:2203.02595, 2022.

[40] A. Donzé, O. Maler, Robust satisfaction of temporal logic over real-valued signals, in: International Conference on Formal Modeling and Analysis of Timed Systems, Springer, 2010, pp. 92–106.
[41] A.D. Ames, X. Xu, J.W. Grizzle, P. Tabuada, Control barrier function based quadratic programs for safety critical systems, IEEE Trans. Autom. Control 62 (8) (2016) 3861–3876.
[42] X. Xu, P. Tabuada, J.W. Grizzle, A.D. Ames, Robustness of control barrier functions for safety critical control, IFAC-PapersOnLine 48 (27) (2015) 54–61.
[43] V. Raman, A. Donzé, M. Maasoumy, R.M. Murray, A. Sangiovanni-Vincentelli, S.A. Seshia, Model predictive control with signal temporal logic specifications, in: 53rd IEEE Conference on Decision and Control, IEEE, 2014, pp. 81–87.
[44] L. Lindemann, J. Nowak, L. Schönbächler, M. Guo, J. Tumova, D.V. Dimarogonas, Coupled multi-robot systems under linear temporal logic and signal temporal logic tasks, IEEE Trans. Control Syst. Technol. 29 (2) (2019) 858–865.
[45] A. Puranic, J. Deshmukh, S. Nikolaidis, Learning from demonstrations using signal temporal logic, in: Conference on Robot Learning, in: PMLR, 2021, pp. 2228–2242.
[46] G.E. Fainekos, G.J. Pappas, Robustness of temporal logic specifications for continuous-time signals, Theor. Comput. Sci. 410 (42) (2009) 4262–4291.
[47] S. Wilson, P. Glotfelter, L. Wang, S. Mayya, G. Notomista, M. Mote, M. Egerstedt, The robotarium: globally impactful opportunities, challenges, and lessons learned in remote-access, distributed control of multirobot systems, IEEE Control Syst. Mag. 40 (1) (2020) 26–44, https://doi.org/10.1109/MCS.2019.2949973.
[48] M.M. Flood, The traveling-salesman problem, Oper. Res. 4 (1) (1956) 61–75.
[49] C. Doersch, A. Zisserman, Sim2real transfer learning for 3d human pose estimation: motion to the rescue, Adv. Neural Inf. Process. Syst. 32 (2019).
[50] A. Kadian, J. Truong, A. Gokaslan, A. Clegg, E. Wijmans, S. Lee, M. Savva, S. Chernova, D. Batra, Sim2real predictivity: does evaluation in simulation predict real-world performance?, IEEE Robot. Autom. Lett. 5 (4) (2020) 6670–6677.