<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *IEEE Intelligent Vehicles Symposium*.

Citation for the original published paper:

Nyberg, T., Pek, C., Dal Col, L., Norén, C., Tumova, J. (2021)

Risk-aware Motion Planning for Autonomous Vehicles with Safety Specifications

In:

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-296090>

Risk-aware Motion Planning for Autonomous Vehicles with Safety Specifications

Truls Nyberg^{1,2}, Christian Pek¹, Laura Dal Col², Christoffer Norén², and Jana Tumova¹

Abstract—Ensuring the safety of autonomous vehicles (AVs) in uncertain traffic scenarios is a major challenge. In this paper, we address the problem of computing the risk that AVs violate a given safety specification in uncertain traffic scenarios, where state estimates are not perfect. We propose a risk measure that captures the probability of violating the specification and determines the average expected severity of violation. Using highway scenarios of the US101 dataset and Responsible Sensitive Safety (RSS) as an example specification, we demonstrate the effectiveness and benefits of our proposed risk measure. By incorporating the risk measure into a trajectory planner, we enable AVs to plan minimal-risk trajectories and to quantify trade-offs between risk and progress in traffic scenarios.

I. INTRODUCTION

Safe driving requires autonomous vehicles (AVs) to account for the uncertain nature of traffic scenarios. In many situations, the future motions of other traffic participants are uncertain, sensor measurements are noisy, and parts of the scenario may be partially or fully occluded. These uncertainties make it difficult to decide whether given trajectories of the AV are safe or if given specifications, such as safe distances, are respected (see Fig. 1). Misjudging these decisions may lead to unsafe or undesired situations, potentially harming passengers and other traffic participants.

Formal verification approaches address these issues by rigorously accounting for bounded sets of uncertainties. These approaches are able to provide safety guarantees with respect to a given specification, but usually only if the uncertainties are over-approximated. However, in reality, these over-approximations are large and thus, may lead to conservative behavior or even the freezing robot problem [1]. Moreover, some scenarios may demand AVs to purposely violate specifications to make progress, e.g., overtaking another vehicle when there might be the possibility of a cut-in (see Fig. 1). For instance, if an overtaking maneuver leads to a significant efficiency increase, the risk of a small safety violation may be acceptable. However, if little is to gain from the maneuver, the same risk might seem unacceptable. As a consequence, uncertain traffic scenarios will require autonomous vehicles to balance between risks and progress towards desired driving goals.

*This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

¹The authors are with the Division of Robotics, Perception and Learning, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden and Digital Futures. Mail addresses: {trulsny, pek2, tumova}@kth.se

²The authors are with Research and Development, Scania CV AB, 151 87 Södertälje, Sweden. Mail addresses: {truls.nyberg, laura.dal.col, christoffer.noren}@scania.com

Fig. 1: The uncertain nature of traffic scenarios will require autonomous vehicles to accept residual risks. The state of vehicle 3 is uncertain, illustrated by different uncertainty levels. To overtake vehicle 3, vehicle 1 needs to account for the possibility that vehicle 3 might change to the left lane (see uncertainty levels on left lane). On the other hand, vehicle 2 needs to keep a safe distance Δ to vehicle 3; however, due to the uncertainties, it is not clear where the rear of vehicle 3 is located (see vertical lines), which is why vehicle 2 might violate the safe distance.

In this paper, we address the problem of estimating the risk of violating safety specifications in uncertain traffic scenarios. To that end, we propose a new risk measure accounting for possible uncertainties in the environment, e.g., uncertain state estimates or predictions of other traffic participants. In contrast to most formal verification methods, our approach allows AVs to estimate the risk of planned trajectories and to balance between risk and progress. Our approach enables AVs to proactively perform minimal-risk motions to reduce potential risks in uncertain environments.

A. Related work

Formal verification describes the process of mathematically proving that a system fulfills a given specification at all times [2], [3]. Various approaches have been proposed over the years to verify the safety of an AV. For instance, in theorem proving it is checked whether all generated trajectories of an AV satisfy a given set of logical formulas that encode safety constraints [4]–[6]. Logical formulas can also be used to directly synthesize correct-by-construction planners that automatically fulfill the specification [7]–[9]. However, there are situations in which no trajectory fulfills the specification due to, e.g., misbehavior of other traffic participants. To address this problem, least-violation planning approaches aim at minimizing the duration or severity of violations in motion planning [10]–[12].

Set-based reachability analysis allows one to compute the set of states that a system is able to reach over time considering an initial set of states and all admissible control inputs [13]. This technique can be used to determine if the AV is able to reach any unsafe set of states [14], [15] or to

rigorously predict the possible occupied space of other traffic participants [16]–[18]. Reachability analysis is well suited to verify the safety of the AV during its operation as shown in [19]–[22]. However, in these approaches it is usually difficult to estimate whether the ego vehicle is close to a violation.

Invariant sets are sets of safe states which allow the AV to remain safe [23]. With these sets, one is able to detect when the AV performs trajectories that are close to violating safety constraints, simply by checking the distance to the boundary of the set [19]. Recent approaches show that invariant sets can be computed using under-approximations [24], reachability analysis [25], control barrier functions [26], [27] or classical control techniques [28], [29]. Yet, it is difficult to incorporate uncertainties in the generation of these sets or to estimate the probability of leaving the set in uncertain traffic scenarios.

Probabilistic approaches for motion planning aim to incorporate uncertainties. The task can be formulated as a Partially Observable Markov Decision Process (POMDP) [30]–[32]. Recent work shows promising results, however providing real-time performance and considering all possible states of a distribution remain a challenge. A framework for control with probabilistic guarantees on safety properties is proposed in [33]. It can be embedded in a receding horizon controller, but is restricted to safety constraints on random variables with unimodal distributions. The framework also uses a constant threshold for the accepted violation probability, a threshold we believe is difficult to tune to imitate human reasoning. As discussed previously, small risks can be justifiable, but only when something is to gain, and the potential severity of the violation is acceptable. We therefore argue that the threshold should also depend on the potential severity of a violation. This allows for fine-granular trade-offs, resembling how we humans assess risk in decision making.

To address the trade-off between safety and progress, risk can be incorporated in the motion planning problem. For instance, minimal-risk or risk-aware planning aim at balancing efficiency and risk costs to find optimal trajectories [34]–[39]. However, it is challenging to rationally define suitable risk measures for AVs. In [40], the authors present an axiomatic theory of risk in robotics, addressing the challenges of choosing a good risk metric and avoiding common pitfalls. The authors define risk as a function mapping a cost random variable to a real value, formed from the state space. The value is interpreted as the risk over an estimated world state. Axioms and requirements on risk measures are proposed and defined in [40] or [41].

Recent approaches for planning with coherent risk measures have been presented for various robotic tasks [42]–[44]. Our work is inspired by these recent approaches on risk, however, our risk measure is tailored to specifically work well with methods for trajectory planning in autonomous driving. It can handle a general family of safety constraints and can tackle multimodal state estimates that typically arise in AVs.

B. Contributions

Autonomous vehicles need to be able to estimate the risk of violating a safety specification given state estimates of itself and surrounding traffic participants. To address this problem for a given safety specification, this paper

- 1) proposes a risk measure to estimate the probability and expected severity of safety violations considering all uncertainties in the state estimates,
- 2) shows how the risk measure can be efficiently computed even if the state estimates are multi-modal, and
- 3) demonstrates the benefits of our risk measure in highway scenarios by evaluating the risk of human driven trajectories and planning minimal-risk trajectories.

II. PROBLEM

A. Preliminaries

We consider lane-based scenarios with N traffic participants $\mathcal{V}_i, i \in \{1, \dots, N\}$, and $\mathcal{V}_e$ describes the ego vehicle. The state of a traffic participant i at time t is a column vector denoted as $x_i(t) \in \mathcal{X}_i \subseteq \mathbb{R}^{n_i}, n_i \in \mathbb{N}_+, t \in \mathbb{R}_{\geq 0}$, and the collection of all states in the scenario is

$$x(t) := \begin{pmatrix} x_1(t) \\ \vdots \\ x_N(t) \end{pmatrix}.$$

We model the estimated state $\hat{x}(t)$ as a random variable with a multivariate Gaussian distribution, $\hat{x} \sim \mathcal{N}(\mu(t), \Sigma(t))$, where $\mu(t)$ and $\Sigma(t)$ are the mean vector and the covariance matrix of $\hat{x}(t)$.

We consider safety specifications given as safety constraints in the form of $h(x(t)) \leq 0, \forall t$. A specification is said to be satisfied when the expression is true, and violated if not. The *severity function*, $\ell_h(x(t))$ is defined as $\ell_h(x(t)) = \ell(h(x(t)))$ if $h(x(t)) > 0$, and $\ell_h(x(t)) = 0$ otherwise, where $\ell(x)$ is a monotonically increasing function. The *severity of violation* is a random variable L obtained by applying the severity function ℓ_h to the random variable $\hat{x}$, i.e. $L = \ell_h(\hat{x})$. A *risk measure* $\rho(L)$ maps the distribution of the severity to a real value interpreted as the current risk.

A sequence of states over a time interval is referred to as a trajectory $\mathcal{T}$. To ensure desired and safe motions, the states along a planned trajectory must satisfy a given specification. Examples of specifications are adhering to traffic rules or maintaining safe distances to other traffic participants.

B. Problem formulation

Trajectory planning typically aims to maximize some efficiency measure, while fulfilling constraints, such as safety constraints that need to hold at all times [45], [46]. Verifying whether a safety constraint is satisfied requires to account for uncertainty in state estimates. An overly conservative approach is to check the safety constraint for all possible states given the current measurement. However, this approach may lead to the freezing robot problem, where no plan makes progress while satisfying the constraints [47]. In contrast, an overly optimistic approach is checking the safety constraint

only for likely states, potentially leading to discomfort or harm to the AV or other traffic participants.

In this work, we aim to define a risk measure that will allow to account for uncertainty and severity of violating the safety constraints (see Sec. II-A), and facilitate risk-aware planning. The authors in [40, Sec. III] propose a number of axioms that should be satisfied by a suitable risk metric in robotics. We argue that on top of them, a risk measure for safe autonomous driving should satisfy the following requirements R1-R4. The risk should:

- (R1) be zero when there is zero probability of violating safety constraints;
- (R2) increase with increasing probability of violating safety constraints;
- (R3) increase with increasing average severity of violation;
- (R4) increase with increasing uncertainty.

Our motivation behind these requirements is to capture the human intuition of risk of violating safety constraints in uncertain traffic situations, given a severity measure.

Running Example

To illustrate the presented concepts, we introduce a running example. We consider that the ego vehicle $\mathcal{V}_e$ is following a leading vehicle $\mathcal{V}_l$ (as in the case of vehicle 2 and 3 in Fig. 1). Let the state be $x = [p_e, v_e, p_l, v_l]^\top$, where the state variables v_e and v_l refer to the velocity of vehicle e and l , respectively, p_e to the ego vehicle's front bumper position, and p_l to the leading vehicle's rear bumper position.

The components of $\hat{x}$ are independent, i.e. $\Sigma = \text{diag}(\sigma_{p_e}^2, \sigma_{v_e}^2, \sigma_{p_l}^2, \sigma_{v_l}^2)$, where $\sigma_\diamond^2$ denotes the variance of a variable $\diamond$. Without loss of generality, we use a curvilinear coordinate system [48] where both $p_e < p_l$ and $\hat{p}_e < \hat{p}_l$.

In this example we will consider a single safety constraint corresponding to keeping a safe distance. The ego vehicle needs at all times to keep the safe distance Δ to the leading vehicle, i.e. $h(x) = \Delta - (p_l - p_e) \leq 0$, where $\Delta = \Delta(v_e, v_l)$ depends on the velocity of both vehicles.

III. APPROACH

A. Severity of Violations

To define our risk measure, we use the severity L introduced in Sec. II-A. The choice is motivated, for instance, by a governing agency issuing fines depending on the severity of violations on a regulated safety specification, such as disobedience of a speed limit. Any monotonically increasing function $\ell(x)$ can be used to shape the severity. However, for the simplicity of further presentation, we assume here that the severity function $\ell(x)$, is linear with increasing distance to satisfaction and simply use $L = \ell_h(\hat{x}) = h(\hat{x})$ for $h(\hat{x}) > 0$, and $L = 0$ otherwise.

Yet, the safety constraint may be quite complex and computing the probability density function (PDF) of L may not be feasible. We demonstrate how the PDF of L can be approximated in the following running example, where we use the Delta method through Taylor series expansions.

Running Example cont.

For the sake of demonstration, we consider a simplified safe distance specified as $\Delta(v_e, v_l) = v_e^2/b_e - v_l^2/b_l + m$, where the parameters b_e and b_l are the vehicles assumed maximum absolute braking acceleration and m is an extra margin. With a Gaussian distribution of the velocity estimate $\hat{v}_e$, the estimated braking distance $\hat{v}_e^2/b_e$ is distributed as a scaled noncentral chi-square distribution. With the Delta method [49], we can approximate the distribution as Gaussian where the mean and variance are estimated with a first order Taylor expansion. A linear approximation of the function $g(v_e) = v_e^2/b_e$ around the estimated $\hat{v}_e$ usually leads to small errors when $\sigma_{v_e} \ll \hat{v}_e$. By approximating the squared terms as Gaussians, Δ becomes a sum of Gaussians, yielding $\Delta \sim \mathcal{N}(\mu_\Delta, \sigma_\Delta^2)$, with

$$\mu_\Delta = \frac{\hat{v}_e^2}{b_e} - \frac{\hat{v}_l^2}{b_l} + m, \quad (1)$$

$$\sigma_\Delta^2 = g'(\hat{v}_l)^2 \sigma_{v_l}^2 + g'(\hat{v}_e)^2 \sigma_{v_e}^2 = \frac{4\hat{v}_e^2 \sigma_{v_e}^2}{b_e^2} + \frac{4\hat{v}_l^2 \sigma_{v_l}^2}{b_l^2}. \quad (2)$$

Since the position estimates are also assumed to be Gaussian, the distance between the vehicles and the difference to the safety distance are all Gaussian. With $h(\hat{x}) = \Delta - (\hat{p}_l - \hat{p}_e)$, the loss $L = \ell_h(\hat{x})$ is a Rectified Gaussian, with PDF

$$f(x; \mu_L, \sigma_L^2) = \Phi(-\frac{\mu_L}{\sigma_L})\delta(x) + \phi(\frac{x - \mu_L}{\sigma_L})U(x), \quad (3)$$

where ϕ and Φ are the standard normal PDF and CDF,

$$\mu_L = \mu_\Delta - (\hat{p}_l - \hat{p}_e), \quad (4)$$

$$\sigma_L^2 = \sigma_\Delta^2 + \sigma_{x_l}^2 + \sigma_{x_e}^2, \quad (5)$$

$\delta(x)$ is the Dirac delta function

$$\delta(x) = \begin{cases} +\infty, & x = 0, \\ 0, & x \neq 0, \end{cases} \quad (6)$$

and $U(x)$ is the unit step function

$$U(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0. \end{cases} \quad (7)$$

The distribution of the severity of violation L can be seen for a numerical example in Fig. 2.

B. Choice of Risk Measure

With the severity L modeled as a random variable, we are now interested in determining a risk measure suitable for AVs. Popular choices for risk management in finance are the risk measures VaR and CVaR [50], defined as

$$\text{VaR}_\alpha(L) = F_L^{-1}(\alpha), \quad (8)$$

$$\text{CVaR}_\alpha(L) = E[L|L > \text{VaR}_\alpha(L)], \quad (9)$$

where F_L^{-1} is the inverse of the cumulative distribution. These measures depend on the tail of the distribution of L , i.e., the potential loss in rare events. Another possible risk measure is the expected value $E[L]$. Figure 2 illustrates $\text{VaR}(L)$, $\text{CVaR}(L)$, and $E[L]$ for the running example.

Fig. 2: Severity L modeled as a Rectified Gaussian distribution illustrated as a blue line. The height of the Dirac is the probability of satisfaction, $\Phi_L(-\mu_L/\sigma_L^2) \approx 0.80$, i.e. the area under the dashed line $h(\hat{x}) < 0$. The shaded area is the probability of violation of the safety constraint $1 - \Phi_L(-\mu_L/\sigma_L^2)$. Risk measures $\text{VaR}_{0.9}$, $\text{CVaR}_{0.9}$, and Expected Violation $E[L]$, are illustrated in yellow, red, and black.

Let us now investigate which of the three possible risk measures is the most suitable for motion planning of AVs. In [40], the authors argue for the use of *coherent* risk measures, such as $E[L]$ or CVaR . We note that VaR is not a coherent risk measure since it fails to fulfill the subadditivity axiom, $\rho(L + L') \leq \rho(L) + \rho(L')$. However, neither VaR or CVaR fulfills the requirements we advocated for in Sec. II-B, since they only depend on the tail of the distribution. For a general distribution of L , the probability of violating a specification can increase, without changing the shape of the tail of the distribution. In a multimodal distribution, VaR and CVaR might only depend on the worst mode of the distribution, for example only accounting the rightmost mode in Fig. 3. Increases in means of the other modes might thus both lead to increased probability of violation and increased overall mean of L , without an increase in the VaR and CVaR .

The simpler risk measure, $E[L]$, is a coherent risk measure fulfilling all the axioms stated in [40]. Furthermore, it fulfills all the requirements listed in Sec. II-B, which we demonstrate in Sec. IV-B. This is achieved since the expected value depends on the complete distribution over all possible states, as opposed to VaR or CVaR which may only consider the worst mode in L . Moreover, the subadditivity axiom is fulfilled with equality since $E[L + L'] = E[L] + E[L']$. This property is a favourable since it allows us to compute the risk in parallel, independently for a set of constraints. We conclude that $\rho(\hat{x}) = E[L]$ is the most suitable risk measure.

Running Example cont.

To further illustrate why the expected severity of violation, $E[L]$, is an appropriate risk measure, we extend the example with an uncertain choice of lead vehicle. Given probabilities of candidate lead vehicles, π_i , the PDF of L is now

$$f(x; \pi, \mu_L, \sigma_L^2) = \sum_i \pi_i f_i(x; \mu_{L_i}, \sigma_{L_i}^2) \quad (10)$$

Fig. 3: Multimodal distribution of the violation and risk measures when considering 3 lead vehicle candidates, each seen as a mode in the distribution. The dashed line is the distribution of $h(\hat{x}) < 0$. The height of the Dirac is the probability of satisfaction, $F_L(0) \approx 0.41$, i.e. the area under the dashed line. The shaded area is the probability of violation, $1 - F_L(0)$.

and the expected severity of violation thus

$$E[L] = 0 + P(L > 0)E[L|L > 0] \quad (11)$$

$$= P(L > 0) \int_0^\infty \sum_i \pi_i f_i(x; \mu_{L_i}, \sigma_{L_i}^2) dx \quad (12)$$

$$= \left(\sum_i \pi_i \Phi\left(\frac{\mu_{L_i}}{\sigma_{L_i}}\right) \right) \sum_i \pi_i \int_0^\infty f_i(x; \mu_{L_i}, \sigma_{L_i}^2) dx \quad (13)$$

$$= \sum_i \pi_i \Phi\left(\frac{\mu_{L_i}}{\sigma_{L_i}}\right) \left(\mu_{L_i} + \sigma_{L_i} \frac{\phi\left(\frac{0 - \mu_{L_i}}{\sigma_{L_i}}\right)}{1 - \Phi\left(\frac{0 - \mu_{L_i}}{\sigma_{L_i}}\right)} \right) \quad (14)$$

$$= \sum_i \pi_i \left(\mu_{L_i} \Phi\left(\frac{\mu_{L_i}}{\sigma_{L_i}}\right) + \sigma_{L_i} \phi\left(\frac{\mu_{L_i}}{\sigma_{L_i}}\right) \right) \quad (15)$$

where the integral term in (13) is the inverse Mills ratio for a Gaussian, see the term in the large parenthesis in (14). The multimodal distribution of L with its $\text{VaR}_{0.9}(L)$, $\text{CVaR}_{0.9}(L)$, and expected severity, $E[L]$, can be seen in Fig. 3 for an example with 3 candidate lead vehicles. Here, the $\text{VaR}_{0.9}$ and $\text{CVaR}_{0.9}$ only depend on the rightmost mode of the distribution associated to the least likely lead vehicle candidate, whereas $E[L]$ weighs in the combined risk stemming from each vehicle. Note that (15) can also be efficiently computed using the standard normal CDF and PDF, as opposed to computing the $\text{VaR}_{0.9}$ and $\text{CVaR}_{0.9}$, where the inverse $F_L^{-1}(0.9)$ is unknown.

IV. EXPERIMENTS

The aim of the experiments is to demonstrate that the chosen risk measure satisfies the requirements listed in Sec. II-B. Further, we aim to show that creating a risk measure from a safety constraint allows for a finer characterization of safety in the state space. Finally, we demonstrate the benefits of using our risk measure for trajectory planning. Videos of our experiments can be found in the media attachments.

TABLE I: RSS parameters and assumed uncertainty of state estimates.

	Name	Value	Unit
RSS Parameters	τ	0.5	[s]
	α_{max}	3.5	[m/s ²]
	β_{min}	4	[m/s ²]
	β_{max}	8	[m/s ²]
Low uncertainty	$\sigma_{position}$	0.5	[m]
	$\sigma_{velocity}$	1	[m]
High uncertainty	$\sigma_{position}$	1	[m]
	$\sigma_{velocity}$	2	[m]

A. Considered specification and dataset

We demonstrate our risk measure on recorded real traffic scenarios from CommonRoad [51], specifically the US101 highway dataset. As our safety specification, we consider safe distances to preceding vehicles according to the Responsible Sensitive Safety (RSS) framework [52]. RSS is currently discussed to become a safety standard and thus, well suited to highlight the benefits of our approach. Similar to our running example, the considered constraint has the form $h(x) = (\Delta(x) - d(x)) \leq 0$, where $d(x)$ is the distance to a lead vehicle. Let us introduce τ as the reaction time of the ego vehicle $\mathcal{V}_e$ to start an emergency braking maneuver, α_{max} as the maximum acceleration during the reaction time, β_{min} as the minimum absolute braking deceleration of $\mathcal{V}_e$, and β_{max} as the maximum absolute braking deceleration of other traffic participants. The longitudinal safe distance $\Delta(x)$ to a lead vehicle $\mathcal{V}_l$ in RSS is defined as

$$\Delta(x) = \underbrace{\frac{2v_r\tau + \alpha_{max}\tau^2}{2}}_{\text{dist. due to reaction time}} + \underbrace{\frac{(v_r + \alpha_{max}\tau)^2}{2\beta_{min}}}_{\text{braking dist. } \mathcal{V}_e} - \underbrace{\frac{v_l^2}{2\beta_{max}}}_{\text{braking dist. } \mathcal{V}_l}.$$

The distribution of the safe distance $\Delta(\hat{x})$ to a lead vehicle is estimated as in the running example in Sec.III-A. To approximate the mean and variance similar to (1), we use a first order Taylor series expansion:

$$\mu_\Delta = \frac{2\hat{v}_e\tau + \alpha_{max}\tau^2}{2} + \frac{(\hat{v}_e + \alpha_{max}\tau)^2}{2\beta_{min}} - \frac{\hat{v}_l^2}{2\beta_{max}}, \quad (16)$$

$$\sigma_\Delta^2 = \sigma_{v_e}^2\tau^2 + \frac{\sigma_{v_e}^2\hat{v}_e^2 + \sigma_{v_e}^2\tau^2\alpha_{max}^2}{\beta_{min}^2} + \frac{\sigma_{v_l}^2\hat{v}_l^2}{\beta_{max}^2}. \quad (17)$$

The severity ℓ of violation corresponds to the amount of meters to satisfaction.

To compute longitudinal distances, we use a curvilinear coordinate system attached to the center-line of the ego vehicle's current lane [48]. We consider a vehicle as a candidate lead vehicle if it is ahead of the ego vehicle and the difference in lateral positions are within the vehicles' averaged widths. Out of the set of all candidate lead vehicles, we choose the lead vehicle to be the one where the ego vehicle is closest to violating the safe distance.

The used parameters are summarized in Tab. I. The parameters are similar to the ones that have been validated in other studies considering RSS [53]–[55]. However, we adjust the reaction time to 0.5s as a compromise between reasonable reaction times for humans and AVs [56], [57].

B. Evaluation of requirements

To demonstrate that our risk measure fulfills the requirements in Sec. II-B, we consider the illustrated scenario Fig. 4, particularly the orange vehicle, $\mathcal{V}_1$.

- (R1) The risk of a vehicle violating a safety constraint is considered to be zero when the estimated probability of violation is zero. With our risk measure as the expected value $E[L]$, this property holds by definition, since the severity is defined as $L = 0$ when the safety specification is satisfied. In our example, the estimated probability of violation is never zero, since Gaussian distributions are used. The risk is therefor never zero, but at the initial time step $t = 0s$ in Fig. 5, the estimated probability of $\mathcal{V}_1$ violating the RSS safe distance is close to zero. The risk is therefor also close to zero.
- (R2) An increase of the probability of violation implies that more samples from the estimated distribution violates the specification. The probability of violation $P(h(\hat{x}) > 0)$ increases over time for the orange vehicle, due to it straying into the neighbouring lane resulting in a change of lead vehicle. The increase in probability of violation is seen in Fig. 5, and the risk measure $E[L] = P(L > 0)E[L|L > 0]$ thus increases too.
- (R3) Our risk measure determines the expected average severity given all possible states of the distribution that violate the specification. With unbiased estimates, the expected value of the severity is the average seen from estimates of the severity $L = h(x) = (\Delta(x) - d(x))$, when $h(x) > 0$. In Fig. 5 the expected value of the severity is seen to rapidly increase after $t = 6s$ when the distance $d(\hat{x})$ decreases due to a change in the orange vehicle's lead vehicle.
- (R4) Increasing the uncertainty in state estimates lowers the estimated probability of violation. However, since the expected severity of violation increases faster, the risk measure increases with less certain state estimates. In Fig. 6 the risk is seen for several vehicles in the scenario when two different levels of uncertainties in the state estimates are considered. The risk is always higher with higher uncertainty, but the difference in risk decreases with increasing probability of violation. When the probability of violation approaches one the risk measure tends towards $h(x) = (\Delta(x) - d(x))$, i.e. the required distance increase in meters to satisfy the RSS specification.

C. Risk-aware trajectory planning

Our risk measure can easily be integrated in many planning frameworks to enable risk-aware planning. To demonstrate the benefit of risk-aware planning, we integrated our risk measure into a sampling-based trajectory planner [45]. This planner generates trajectories by sampling goal states and connecting them with the ego vehicle's initial state through quintic polynomials. The cost function J of this planner is modified to only punish the cumulative risk along a trajectory $\mathcal{T} := (\hat{x}(t_0), \dots, \hat{x}(t_h))$ from time t_0 to t_h .

Fig. 4: Four snapshots of a US 101 traffic scenario (CommonRoad ID: USA_US101-11.4-T-1:2018) at different time steps $t \in \{0s, 2.5s, 4.5s, 7s\}$. The orange vehicle $\mathcal{V}_1$ is straying right into a neighbouring lane, the purple vehicle $\mathcal{V}_2$ experiences a cut-in from the left adjacent lane, and the red vehicle $\mathcal{V}_3$ might experience a cut-in from the left adjacent lane since a vehicle is driving close to the lane boundary.

Fig. 5: The risk $\rho(L)$, the mean RSS safe distance violation, and probability of violation for the orange vehicle $\mathcal{V}_1$ in Fig 4.

As an example, the purple vehicle $\mathcal{V}_2$ from the scenario in Fig. 4 is chosen as the ego vehicle. As seen in Fig. 6, the vehicle is at a high risk of violating the RSS safe distance since it experiences a cut-in maneuver at $t = 1s$. The cumulative RSS violation for the human driven trajectory is over 1400 m, meaning an average of 20 meters violation severity per time step of $0.1s$. We sample over 550 trajectories in total with a time step of $0.1s$ and time horizon of $7s$ with goal states at different lateral positions and velocities.

Fig. 7a illustrates the 80 sampled trajectories with lowest cumulative risk for vehicle $\mathcal{V}_2$. The risk for each trajectory is also plotted over time (see Fig. 7b). The color of each trajectory in Fig. 7 encodes the cumulative risk over the scenario's time horizon of $7s$. To not significantly decrease the velocity of $\mathcal{V}_2$, we only consider trajectories with an average velocity above $10 m/s$. The cumulative risk of violation for the planned trajectories are all below $300m$, meaning an average below 5 meters of violation severity

Fig. 6: The risk $\rho(\hat{x})$, the mean RSS safe distance violation, and the probability of violation for the vehicles $\mathcal{V}_1$, $\mathcal{V}_2$, and $\mathcal{V}_3$ in Fig 4 for two different levels of assumed uncertainty in the state estimates.

per time step. The least risky trajectory in Fig. 7 is the trajectory with the lowest velocity. However, several lane changing trajectories are planned with higher velocities while still significantly reducing the risk when comparing to the human driven trajectory contained in the dataset.

Lastly, we investigate the need for AVs to balance risk and progress by analyzing how the risk of the planned trajectories compare to the human driven trajectory of vehicle $\mathcal{V}_2$ (see Fig. 8). Therefore, we evaluate the increase of velocity and

Fig. 7: Planning results with our proposed risk measure. (a) shows the planned trajectories of vehicle V_2 . (b) illustrates the change of risk of each trajectory over time. (c) the trajectories are color coded according to the cumulative risk.

Fig. 8: For each planned trajectory, we plot the velocity increase over the cumulative risk increase compared to the human driven trajectory (in which the risk is not increasing and thus, the value of 1).

cumulative risk for each trajectory over its entire planning horizon (including the human driven). For instance, the human driven trajectory (see red sample in Fig. 8) increases the velocity by around 3% and has a risk increase factor of 1, meaning that its risk does not change. Our planned trajectory T_1 is able to increase the velocity by 10% while at the same time reducing the risk to 0.95 compared to the human. Trajectory T_2 , on the other hand, has the lowest risk of 0.05 over all trajectories, but reduces the velocity by 50%. The highest velocity increase of 21% is achieved with trajectory T_3 , however, at the cost of increasing the risk by 25%. Lastly, our risk measure also allows to determine that even if the velocity of trajectory T_4 is slightly lower at 0.95%, the risk increases significantly by almost 50%.

V. CONCLUSIONS

In this paper we propose a risk measure for AVs that estimates the average expected violation severity of a safety constraint. The risk is modeled over a random variable that describes the severity of violating the safety constraint, given by an associated severity function. We show how the risk can be computed even when its distribution is multi-modal. To demonstrate that our risk measure fulfills the requirements of risk measures, we evaluate the risk of human driven vehicles in the US1010 dataset. Moreover, we incorporate our risk measure into a sampling based planner to generate minimal-risk trajectories. Our results affirm that autonomous vehicles need to balance between progress and risk. In this regard, our risk measure provides valuable insights on how the risk of trajectories changes over the planning horizon.

For future work, we endeavour to explore the optimal trade-off between progress and risk. We would also like to

investigate and formalize acceptable thresholds for risk and perceived safety. Further requirements on safety constraints and severity measures may be needed to fully model how risk should be assessed in traffic, however, we believe this work can serve as a starting point on how a risk measure can be built from a regulatory framework for safety in AVs.

REFERENCES

- [1] P. Trautman and A. Krause, "Unfreezing the robot: Navigation in dense, interacting crowds," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2010, pp. 797–803.
- [2] D. Walter, H. Täubig, and C. Lüth, "Experiences in applying formal verification in robotics," in *Proc. of the Int. Conf. on Computer Safety, Reliability, and Security*, 2010, pp. 347–360.
- [3] E. M. Clarke and J. M. Wing, "Formal methods: State of the art and future directions," *ACM Computing Surveys*, vol. 28, no. 4, pp. 626–643, 1996.
- [4] W. Damm, H.-J. Peter, J. Rakow, and B. Westphal, "Can we build it: formal synthesis of control strategies for cooperative driver assistance systems," *Mathematical Structures in Computer Science*, vol. 23, no. 04, pp. 676–725, 2013.
- [5] M. Hilscher, S. Linker, and E.-R. Olderog, "Proving safety of traffic manoeuvres on country roads," in *Theories of Programming and Formal Methods*. Springer, 2013, pp. 196–212.
- [6] S. Mitsch, S. M. Loos, and A. Platzer, "Towards formal verification of freeway traffic control," in *Proc. of the IEEE Int. Conf. on Cyber-Physical Systems*, 2012, pp. 171–180.
- [7] V. Raman, A. Donzé, D. Sadigh, R. M. Murray, and S. A. Seshia, "Reactive synthesis from signal temporal logic specifications," in *Proc. of the Int. Conf. on hybrid systems: Computation and control*, 2015, pp. 239–248.
- [8] A. Ulusoy, T. Wongpiromsarn, and C. Belta, "Incremental controller synthesis in probabilistic environments with temporal logic constraints," *The Int. Journal of Robotics Research*, vol. 33, no. 8, pp. 1130–1144, 2014.
- [9] A. Nikou, D. Boskos, J. Tumova, and D. V. Dimarogonas, "On the timed temporal logic planning of coupled multi-agent systems," *Automatica*, vol. 97, pp. 339–345, 2018.
- [10] J. Tumova, G. C. Hall, S. Karaman, E. Frazzoli, and D. Rus, "Least-violating control strategy synthesis with safety rules," in *Proceedings of the Int. Conf. on Hybrid systems: computation and control*, 2013, pp. 1–10.
- [11] T. Wongpiromsarn, K. Slutsky, E. Frazzoli, and U. Topcu, "Minimum-violation planning for autonomous systems: Theoretical and practical considerations," *arXiv preprint arXiv:2009.11954*, 2020.
- [12] J. Karlsson and J. Tumova, "Intention-aware motion planning with road rules," in *Proc. of the IEEE Int. Conf. on Automation Science and Engineering*, 2020, pp. 526–532.
- [13] A. Girard, C. Le Guernic, and O. Maler, "Efficient computation of reachable sets of linear time-invariant systems with inputs," in *Proc. of the Int. Conf. on Hybrid Systems: Computation and control*, 2006, pp. 257–271.
- [14] M. Althoff and J. M. Dolan, "Online verification of automated road vehicles using reachability analysis," *IEEE Transactions on Robotics*, vol. 30, no. 4, pp. 903–918, 2014.
- [15] C. Pek, S. Manzing, M. Koschi, and M. Althoff, "Using online verification to prevent autonomous vehicles from causing accidents," *Nature Machine Intelligence*, vol. 2, no. 9, pp. 518–528, 2020.

- [16] M. Koschi and M. Althoff, "Set-based prediction of traffic participants considering occlusions and traffic rules," *IEEE Transactions on Intelligent Vehicles*, 2020.
- [17] P. F. Orzechowski, A. Meyer, and M. Lauer, "Tackling occlusions & limited sensor range with set-based safety verification," in *Proc. of the Int. Conf. on Intelligent Transportation Systems*, 2018, pp. 1729–1736.
- [18] H. Michael, F. Antonella, and W. Daniel, "Data-based reachability analysis for movement prediction of pedestrians and motion planning," in *Proc. of the IEEE Int. Conf. on Vehicular Electronics and Safety*, 2018, pp. 1–7.
- [19] C. Pek and M. Althoff, "Fail-safe motion planning for online verification of autonomous vehicles using convex optimization," *IEEE Transactions on Robotics*, 2020.
- [20] S. Vaskov, S. Kousik, H. Larson, F. Bu, J. R. Ward, S. Worrall, M. Johnson-Roberson, and R. Vasudevan, "Towards provably not-at-fault control of autonomous robots in arbitrary dynamic environments," in *Proc. of Robotics: Science and Systems*, 2019, pp. 1–10.
- [21] S. L. Herbert, M. Chen, S. Han, S. Bansal, J. F. Fisac, and C. J. Tomlin, "FaSTrack: a modular framework for fast and guaranteed safe motion planning," in *Proc. of the IEEE Conference on Decision and Control*, 2017, pp. 1517–1522.
- [22] K. Leung, E. Schmerling, M. Zhang, M. Chen, J. Talbot, J. C. Gerdes, and M. Pavone, "On infusing reachability-based safety assurance within planning frameworks for human–robot vehicle interactions," *The Int. Journal of Robotics Research*, vol. 39, no. 10–11, pp. 1326–1345, 2020.
- [23] F. Blanchini, "Set invariance in control," *Automatica*, vol. 35, no. 11, pp. 1747 – 1767, 1999.
- [24] C. Pek and M. Althoff, "Efficient computation of invariably safe states for motion planning of self-driving vehicles," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2018, pp. 3523–3530.
- [25] A. Lawitzky, A. Nicklas, D. Wollherr, and M. Buss, "Determining states of inevitable collision using reachability analysis," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2014, pp. 4142–4147.
- [26] A. D. Ames, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs with application to adaptive cruise control," in *Proc. of the IEEE Int. Conf. on Decision and Control*, 2014, pp. 6271–6278.
- [27] L. Lindemann and D. V. Dimarogonas, "Control barrier functions for signal temporal logic tasks," *IEEE Control Systems Letters*, vol. 3, no. 1, pp. 96–101, 2018.
- [28] M. Jalalmaab, B. Fidan, S. Jeon, and P. Falcone, "Guaranteeing persistent feasibility of model predictive motion planning for autonomous vehicles," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2017, pp. 843–848.
- [29] M. Korda, D. Henrion, and C. N. Jones, "Convex computation of the maximum controlled invariant set for polynomial control systems," *SIAM Journal on Control and Optimization*, vol. 52, no. 5, pp. 2944–2969, 2014.
- [30] S. Brechtel, T. Gindele, and R. Dillmann, "Probabilistic decision-making under uncertainty for autonomous driving using continuous POMDPs," in *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*, 2014, pp. 392–399.
- [31] C. Hubmann, J. Schulz, M. Becker, D. Althoff, and C. Stiller, "Automated Driving in Uncertain Environments: Planning With Interaction and Uncertain Maneuver Prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 1, pp. 5–17, Mar. 2018.
- [32] M. Bouton, J. Karlsson, A. Nakhaei, K. Fujimura, M. J. Kochenderfer, and J. Tumova, "Reinforcement Learning with Probabilistic Guarantees for Autonomous Driving," *arXiv:1904.07189 [cs]*, May 2019.
- [33] D. Sadigh and A. Kapoor, "Safe Control under Uncertainty with Probabilistic Signal Temporal Logic," in *Robotics: Science and Systems XII*. Robotics: Science and Systems Foundation, 2016.
- [34] J. Bernhard and A. Knoll, "Risk-constrained interactive safety under behavior uncertainty for autonomous driving," *arXiv preprint arXiv:2102.03053*, 2021.
- [35] M.-Y. Yu, R. Vasudevan, and M. Johnson-Roberson, "Occlusion-Aware Risk Assessment for Autonomous Driving in Urban Environments," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 2235–2241, Apr. 2019.
- [36] —, "Risk Assessment and Planning with Bidirectional Reachability for Autonomous Driving," *arXiv:1909.08059 [cs]*, Sep. 2019.
- [37] S. G. McGill, G. Rosman, T. Ort, A. Pierson, I. Gilitschenski, B. Araki, L. Fletcher, S. Karaman, D. Rus, and J. J. Leonard, "Probabilistic Risk Metrics for Navigating Occluded Intersections," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4322–4329, 2019.
- [38] A. Pierson, W. Schwarting, S. Karaman, and D. Rus, "Navigating Congested Environments with Risk Level Sets," in *Proc. of the IEEE International Conference on Robotics and Automation*, May 2018, pp. 5712–5719.
- [39] F. Damerow and J. Eggert, "Balancing risk against utility: Behavior planning using predictive risk maps," in *Proc. of the IEEE Intelligent Vehicles Symposium*. Seoul, South Korea: IEEE, Jun. 2015, pp. 857–864.
- [40] A. Majumdar and M. Pavone, "How Should a Robot Assess Risk? Towards an Axiomatic Theory of Risk in Robotics," in *Robotics Research*, N. M. Amato, G. Hager, S. Thomas, and M. Torres-Torriti, Eds. Cham: Springer International Publishing, 2020, vol. 10, pp. 75–84.
- [41] D. Bertsimas and D. B. Brown, "Constructing Uncertainty Sets for Robust Linear Optimization," *Operations Research*, vol. 57, no. 6, pp. 1483–1495, Dec. 2009.
- [42] A. Majumdar, S. Singh, A. Mandlekar, and M. Pavone, "Risk-sensitive Inverse Reinforcement Learning via Coherent Risk Models," in *Robotics: Science and Systems XIII*. Robotics: Science and Systems Foundation, Jul. 2017.
- [43] S. Safaoui, L. Lindemann, D. V. Dimarogonas, I. Shames, and T. H. Summers, "Control Design for Risk-Based Signal Temporal Logic Specifications," *IEEE Control Systems Letters*, vol. 4, no. 4, pp. 1000–1005, Oct. 2020.
- [44] M. P. Chapman, R. Bonalli, K. M. Smith, I. Yang, M. Pavone, and C. J. Tomlin, "Risk-sensitive safety analysis using Conditional Value-at-Risk," *arXiv:2101.12086 [cs, eess]*, Jan. 2021.
- [45] M. Werling, S. Kammel, J. Ziegler, and L. Gröll, "Optimal trajectories for time-critical street scenarios using discretized terminal manifolds," *The Int. Journal of Robotics Research*, vol. 31, no. 3, pp. 346–359, 2012.
- [46] B. Paden, M. Čáp, S. Z. Yong, D. Yershov, and E. Frazzoli, "A survey of motion planning and control techniques for self-driving urban vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, p. 33–55, Mar 2016.
- [47] W. Schwarting, J. Alonso-Mora, and D. Rus, "Planning and Decision-Making for Autonomous Vehicles," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 187–210, May 2018.
- [48] E. Héry, S. Masi, P. Xu, and P. Bonnifait, "Map-based curvilinear coordinates for autonomous vehicles," in *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*, 2017, pp. 1–7.
- [49] J. M. V. Hoef, "Who invented the delta method?" *The American Statistician*, vol. 66, no. 2, pp. 124–127, 2012.
- [50] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Mathematical Finance*, vol. 9, no. 3, pp. 203–228, 1999.
- [51] M. Althoff, M. Koschi, and S. Manzing, "CommonRoad: Composable benchmarks for motion planning on roads," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2017, pp. 719–726.
- [52] S. Shalev-Shwartz, S. Shammah, and A. Shashua, "On a Formal Model of Safe and Scalable Self-driving Cars," *arXiv:1708.06374*, 2018.
- [53] B. Gassmann, F. Oboril, C. Buerkle, S. Liu, S. Yan, M. S. Elli, I. Alvarez, N. Aerrabotu, S. Jaber, P. van Beek, D. Iyer, and J. Weast, "Towards Standardization of AV Safety: C++ Library for Responsibility Sensitive Safety," in *Proc. of IEEE Intelligent Vehicles Symposium*, 2019, pp. 2265–2271.
- [54] X. Xu, X. Wang, X. Wu, O. Hassanin, and C. Chai, "Calibration and evaluation of the Responsibility-Sensitive Safety model of autonomous car-following maneuvers using naturalistic driving study data," *Transportation Research Part C: Emerging Technologies*, vol. 123, p. 17, Feb. 2021.
- [55] C. Chai, X. Zeng, X. Wu, and X. Wang, "Safety Evaluation of Responsibility-Sensitive Safety (RSS) on Autonomous Car-Following Maneuvers Based on Surrogate Safety Measurements," in *Proc. of the IEEE Intelligent Transportation Systems Conference*, Oct. 2019, pp. 175–180.
- [56] A. Rodionova, I. Alvarez, M. S. Elli, F. Oboril, J. Quast, and R. Mangharam, "How safe is safe enough? Automatic safety constraints boundary estimation for decision-making in automated vehicles," in *Proc. IEEE Intelligent Vehicles Symposium*, 2020, pp. 1457–1464.
- [57] M. Althoff and R. Losch, "Can automated road vehicles harmonize with traffic flow while guaranteeing a safe distance?" in *Proc. of the IEEE Int. Conf. on Intelligent Transportation Systems*, Nov. 2016, pp. 485–491.