AGH University of Krakow

**FIELD OF SCIENCE: ENGINEERING AND TECHNOLOGY**

SCIENTIFIC DISCIPLINE: AUTOMATION, ELECTRONICS, ELECTRICAL ENGINEERING AND SPACE TECHNOLOGIES

SUMMARY OF ACCOMPLISHMENTS

# Real-Time Generation of Safe Trajectories for Autonomous Vehicles in Dynamic Environments


Author: *Wojciech Turlej*

First supervisor: *prof. dr hab. inż. Wojciech Mitkowski*

Auxiliary supervisor: *dr inż. Krzysztof Kogut*

Completed at: AGH University of Krakow, Faculty of Electrical Engineering, Automatics, Computer Science, and Biomedical Engineering


Kraków, 2024

# 1. Abstract

Driving is a task that is very difficult to automatize, mainly due to uncertain, often unpredictable behaviors of other road users. Complex interactions between traffic participants, imperfections of perception systems, and the need for long-term strategic planning, are all difficult to address in a deterministic, algorithmic manner. For this reason, motion planning approaches based on Machine Learning are often proposed as a potentially powerful alternative to previously utilized methods.

The thesis focuses on challenges related to the design and evaluation of motion planning systems based on Machine Learning approaches, helping to ensure their reliability and robustness. To address these challenges, several novel algorithms and solutions are proposed and evaluated.

To enable planning a safe trajectory in particularly challenging uncertain situations, a Multiple Hypothesis Planning method is proposed. The method allows to plan a set of optimal trajectories, that take into account several, possibly conflicting hypotheses regarding the future motion of other road users, as well as regarding the state of the controlled vehicle's environment. Several examples of the method's applications are presented, most notably in fail-safe planning, where a worst-case emergency trajectory is generated alongside the vehicle's nominal trajectory, ensuring the existence of a feasible collision-avoidance maneuver.

The described method is intended to be used in conjunction with a vehicle behavior planning method based on a Reinforcement Learning approach. Since motion planning systems in such setups are trained in a simulation environment, they are often susceptible to perception errors present in real systems. To address this issue, a set of efficient sensor models is introduced for training and evaluation of such systems. Proposed models are utilized to train a set of driving policies, that are extensively tested in various simulation environments.

Finally, to ensure the robustness of motion planning systems in rare, challenging scenarios, a novel method for the automatic generation of test scenarios is proposed. The introduced method utilizes stochastic optimization techniques to generate a wide set of scenarios that pose a particular challenge to a tested system, allowing to explore its limitations, and to prepare scenarios for further training. The effectiveness of the proposed method is demonstrated in the task of generating adversarial scenarios for machine learning-based driving policies.

# 2. Motivation

Despite the considerable amount of effort already invested in the development of vehicle motion planning systems, the field of Autonomous Driving remains an active research area. Among the proposed approaches, methods based on Reinforcement Learning (RL) emerge as one of the most promising research directions. RL-based methods typically utilize neural networks trained in a traffic simulation environment to select the appropriate behavior of the vehicle and/or generate its motion or control trajectories. Such a machine-learning-based system that utilizes data from the vehicle's sensors to plan its future motion can be referred to as a driving policy.

RL-based driving policies have been demonstrated to achieve promising reliability and efficiency in driving tasks, being able to account for complex interactions between other road users, plan long-term driving strategies, predict the behavior of other vehicles, and exhibit human-like negotiation skills on the road.

While the capabilities of RL-based driving policies are promising, integration of such algorithms into commercial vehicles requires solving several challenges related to their safety and reliability. Most notably, the following issues must be addressed before such algorithms can be deployed on a large scale.

- Uncertainty handling. RL-based policies often struggle with situations that can rarely be observed in the simulation environment, especially if they require consideration of several possible outcomes. Such policies often prioritize the most plausible outcomes of a given situation, disregarding less plausible ones. To address this issue, a transparent trajectory planning technique can be used in conjunction with the RL-based driving policy to plan a trajectory in situations, in which either the future behavior of other road users or the current state of the environment is uncertain.

- Robustness to perception errors. Performance of perception systems and sensors used in the automotive industry is limited, and often susceptible to degradation in adverse weather conditions. For this reason, driving policies trained in a perfect simulation environment may struggle with generalization to real-life conditions, possibly leading to erratic or unsafe behavior. Further research is thus needed to evaluate the impact of perception errors on such driving policies and to efficiently model sensing systems in the simulation environment used for RL training.

- Validation and verification. Lack of transparency inherent to ML-based methods requires careful evaluation of their performance in difficult situations. Manual definition of the test scenarios however is not only tiresome but also may be insufficient to uncover potential problems in such systems, as erroneous decisions of RL-based driving policies may not necessarily be correlated with an objective difficulty of the test scenario. For this reason, further research into automated adversarial testing techniques is needed to ensure a thorough exploration of potential issues in the evaluated systems.

The thesis focuses on addressing these issues, proposing several algorithms that improve the safety of driving policies and help to evaluate them efficiently.

# 3. Research hypotheses

Three main hypotheses were formulated concerning the described issues.

1. It is possible to create a safe driving plan for an automated vehicle that considers several hypotheses regarding the future state of the vehicle's surroundings. In particular, reasonably foreseeable worst-case assumptions regarding the behavior of other road users can be taken into account in the motion planning algorithm, ensuring the existence of feasible collision avoidance maneuvers during the execution of the motion plan.

2. The use of stochastic models of perception systems in the training process of a Reinforcement-Learning driving policy improves the policy's robustness to perception errors.

3. Optimization-based adversarial scenario generation methods can be used in simulation-based validation of motion planning algorithms to expose potential weaknesses or issues in the evaluated systems.

# 4. Methods and results

The research presented in this thesis spans three interconnected topics, each related to one of the research hypotheses. The first topic, the Multiple Hypothesis Trajectory Planning is related to the first hypothesis, and covers a novel algorithm for trajectory planning in uncertain situations, that can generate a vehicle control trajectory, that takes into account an arbitrary number of hypotheses regarding current and/or future state of the vehicle's environment. The second topic is closely related to the hypothesis, that stochastic models of perception systems can improve an RL-based policy's robustness if used in the training process. High-level models were proposed for systems that perceive both static and dynamic environments of the vehicle, and their usefulness has been presented in the simulation experiments, where driving policy trained with proposed models has been compared to policies trained in simpler training environments. The last major topic covers a novel algorithm for optimization-based generation of adversarial test scenarios, that can be used to effectively explore deficiencies in an arbitrary autonomous driving system.

## 4.1. Multiple Hypothesis Trajectory Planning

Handling uncertainties is an inherent part of a driving task. Both human drivers and autonomous driving systems must deal with uncertainty regarding the current state of the vehicle's environment (as a consequence of limited perception capabilities), as well as take into account possible future decisions of other road users, that often cannot be precisely predicted.

Multiple Hypothesis Trajectory Planning algorithm proposed in this thesis is capable of planning the vehicle's control and state trajectories while taking into account two or more conflicting hypotheses regarding the current state and future behavior of other road users. The algorithm is particularly well suited to be used in conjunction with multi-modal trajectory prediction software, which often provides a multitude of plausible predictions of others' behavior.

The proposed planning algorithm utilizes a numerical optimization library to solve a non-linear optimization problem, where optimized parameters describe two or more vehicle control trajectories, each related to a single hypothesis regarding the current or future state of the environment. The trajectories, apart from fulfilling constraints related to the physical feasibility of generated motions, must fulfill a series of constraints, that enforce their equality in a predefined initial period. As a result, a single optimization problem produces several trajectories, where each trajectory constitutes an effective collision-free solution of a short-term planning task, assuming that its relative hypothesis is true. As all the generated trajectories overlap for a certain time, the decision of which one should be executed can be postponed, allowing to gather further observations that may trivialize such a decision by proving one of the hypotheses.

To illustrate a possible application of the described method, the situation presented in Fig. 4.1 can be considered.

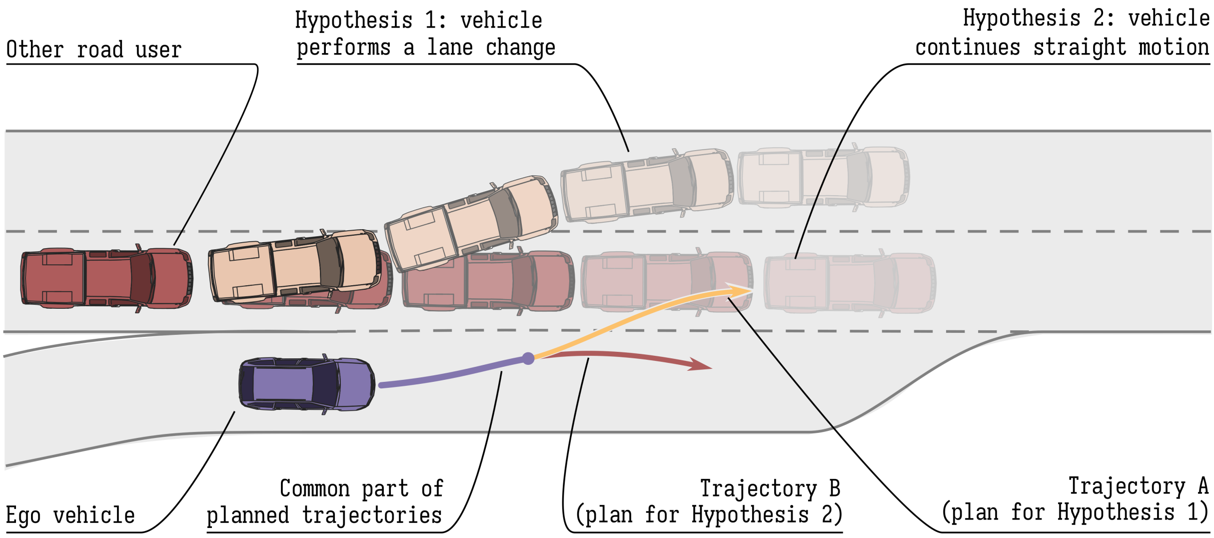


**Figure 4.1.** General idea of the multiple hypotheses planning algorithm. Multiple trajectories are planned simultaneously based on different hypotheses. The problem is formulated in a way that enforces the trajectories to remain identical for a certain duration.

In the given scenario, the ego vehicle merges into highway traffic, while another road user drives on the adjacent lane. The road user is anticipated to change lanes, potentially signaling this intention using an indicator light. Two hypotheses can be formulated regarding the behavior of the other road user: either they will successfully change lanes, allowing the ego vehicle to merge safely (Hypothesis 1), or they will fail to do so, e.g., due to an unsafe situation in the adjacent lane (Hypothesis 2). Based on these hypotheses, two trajectories are planned: one, in which Hypothesis 1 is assumed to be true, and the ego vehicle changes lanes (Trajectory A), and another considering the alternative hypothesis, where the ego vehicle brakes and stays in its current lane (Trajectory B). Planned trajectories overlap within a certain time frame, allowing to postpone a decision on which trajectory to execute until more information is gathered. It can be noted, that this way of planning vehicle trajectory results in the emergence of cautious behaviors - e.g., in the presented example, while Trajectory A describes an assertive merge-in maneuver, it is planned in a way that allows the safe execution of an emergency braking maneuver if merge-in is not possible.

While the hypotheses used for the planning may be provided by an external trajectory prediction module in the form of state trajectories, they can also take the form of spatio-temporal constraints (occupancy sets) that describe areas that the other vehicle may occupy over time in each hypothesis. Several methods for designing such hypotheses based on the simulation of plausible behaviors have been presented in the thesis.

To illustrate the possible use of the occupancy sets to describe plausible behaviors of other road users, the example presented in Fig. 4.1 can be extended to take into account three hypotheses formulated in the form of occupancy sets:

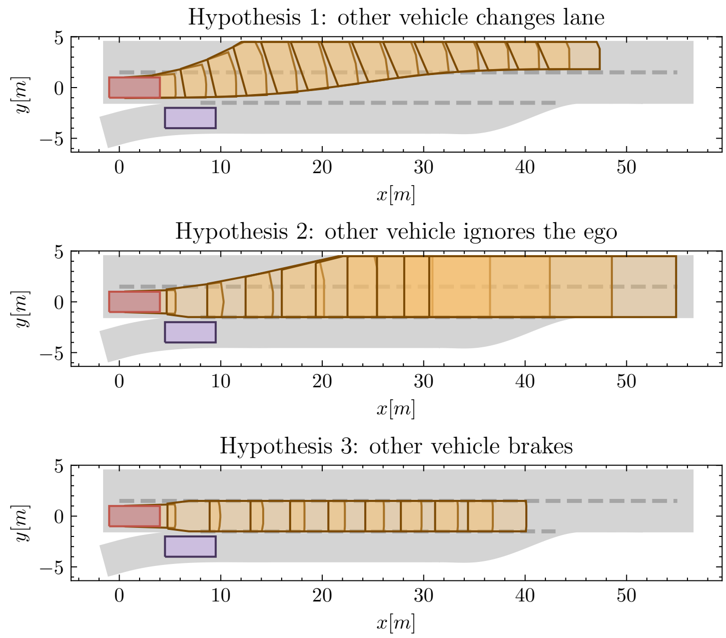


**Figure 4.2.** Constraints (occupancy sets) generated for different hypotheses. The ego vehicle is marked blue, the other vehicle - red, and the constraints - yellow.

Results of the execution of the proposed trajectory planning method taking into account these three hypotheses are shown in Fig 4.3.

As shown in Fig. 4.3, the use of the proposed method in a given scenario resulted in a generation of three trajectories, that overlap in the initial duration of 0.75 s. Each trajectory remains collision-free for their respective hypotheses. Trajectories generated for Hypotheses 1 and 3, where the other vehicle yields to allow the ego to merge into traffic result in the ego's assertive lane change, while the second trajectory constitutes an emergency braking maneuver, that ensures collision avoidance in case the other road user fails to yield.

The proposed method can also be used to implement a Fail-Safe Planning system. In this setup, one trajectory (a nominal trajectory) is planned based on the most plausible hypothesis regarding the behavior of other road users, while another one (a fail-safe trajectory) assumes that other road users can perform any action within certain limits, effectively describing a reasonable worst-case scenario. Such a control scheme, executed repeatedly in a Model-Predictive Control scheme allows to ensure, that a collision-free trajectory always exists (assuming accurate perception data and worst-case assumptions) and can be executed to avoid potential accidents.

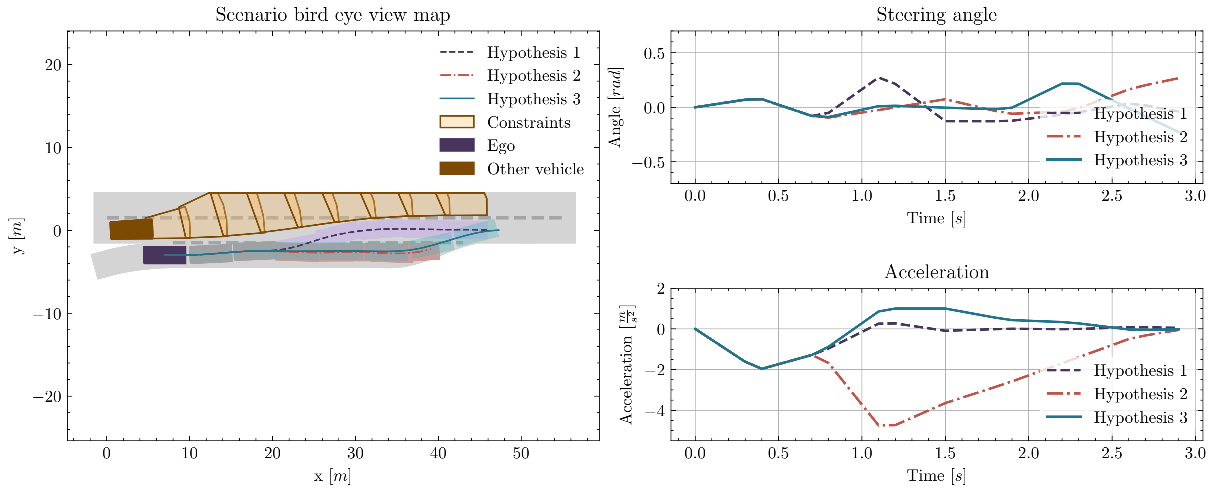


**Figure 4.3.** Trajectory generation results for the merge-in scenario. Note, that only the constraints based on *Hypothesis 1* were plotted for clarity.

An example of the Fail-Safe Planning application is presented in Fig. 4.4

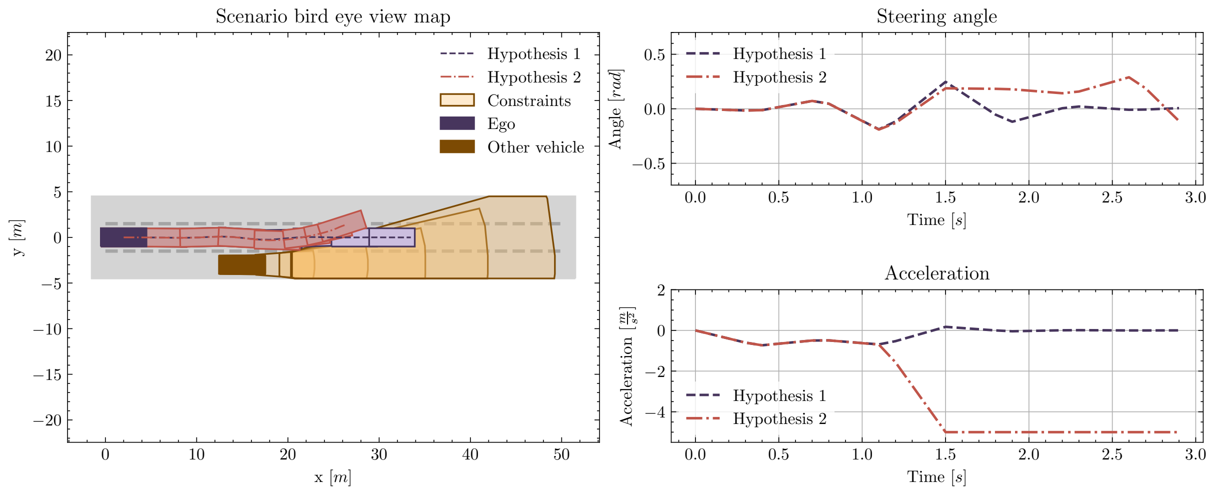


**Figure 4.4.** Trajectories generated for the highway fail-safe planning scenario. Hypothesis 1 assumes that the other vehicle remains on its lane, resulting in a straight nominal trajectory of the ego vehicle. Hypothesis 2 covers worst-case scenario behaviors, in which the other vehicle may perform an aggressive cut-in maneuver. An emergency maneuver composed of braking and evasive steering is planned for this scenario.

Unlike existing methods, in the proposed approach all the trajectories are planned simultaneously in a single optimization problem, and thus the nominal trajectory is influenced by the necessity of the fail-safe trajectory existence. In other words, the quality of the nominal trajectory (measured as a combination of passengers' comfort, fuel efficiency, and other desired characteristics) can be partially sacrificed to ensure the existence of a feasible collision avoidance maneuver.

## 4.2. Sensor Models for Reinforcement Learning

Autonomous vehicles perceive their surroundings using a set of sensors, such as automotive cameras and radars. The performance of such sensors is however inherently limited. Errors in the environment perception that may occur in such systems can be categorized into three main categories, described below.

- False negative detection errors. Perception systems may fail to detect other road users, traffic signs, lane markers, and other elements of the static environment, e.g. due to adverse weather conditions, software errors, or occlusions. False negative detection errors are particularly dangerous, as the subsequent behavior and trajectory planning systems may fail to plan a collision-free trajectory if the perception system fails to detect a nearby vehicle or obstacle.

- False positive detection errors. This type of failures refers to situations, where a non-existing environmental feature (vehicle, lane marker, obstacle, etc.) is reported to exist by the perception system. This type of errors may occur e.g., when the emitted radar wave reflected from an object is reflected by other environmental features before being registered by the radar's receiving antennas. While less critical than false negatives, false positive detection errors may trigger potentially dangerous evasive maneuvers.

- State estimation errors. Due to limited resolution and performance, automotive sensors typically provide only an approximation of other objects' states. In extreme cases, inaccuracies in the state estimation may also result in incorrect decisions of the planning system that may lead to hazardous situations.

Driving policies based on the Reinforcement Learning (RL) approach rely on extensive training in simulation environments, that typically provide an accurate description of the vehicle's surroundings. Policy trained in such an environment may however struggle in the presence of realistic errors, that were not encountered in the training. This problem of the discrepancies between the training environment and the target environment is commonly referred to as "sim to real gap".

One of the approaches commonly utilized to combat this issue is based on domain randomization - i.e., the introduction of random errors to the observations provided as an input to the neural network that performs key decisions in the RL-based planning system. It is however unclear, whether applying simple random errors is sufficient to alleviate this issue, as the errors produced by perception systems tend to be time-correlated and follow certain characteristics, that may require special care in the error modeling.

As an alternative to simple error models based on a Gaussian noise applied to state estimates and random occurrences of false detection errors, a set of high-level sensor models has been proposed in this thesis. Proposed models can be used to simulate an arbitrary dynamic objects detection and tracking module as well as a lane markers detection system.

The approach proposed for the dynamic objects implements a state estimation errors model based on Orstein-Uhlenbeck processes, that result in time-correlated inaccuracies that mimic ones observed in the radar-based and camera-based perception systems with subsequent tracking and fusion modules. Additionally, false negative detection errors are modeled as long-lasting events, with the occurrence probability and duration calculated based on their distance, time since detection, occlusions, and other characteristics. False positive object detection errors are on the other hand modeled with the use of kinematic motion models, to achieve error patterns similar to ones observed in detection systems with tracking or fusion modules based on the Kalman Filter algorithm, commonly used in such applications.

Lane marker detection errors are also modeled in a simple, yet relatively realistic manner, taking into account occlusions from other vehicles, dynamically limiting the detection distance, and modeling the markers' geometry estimation errors with the use of an Ornstein-Uhlenbeck noise.

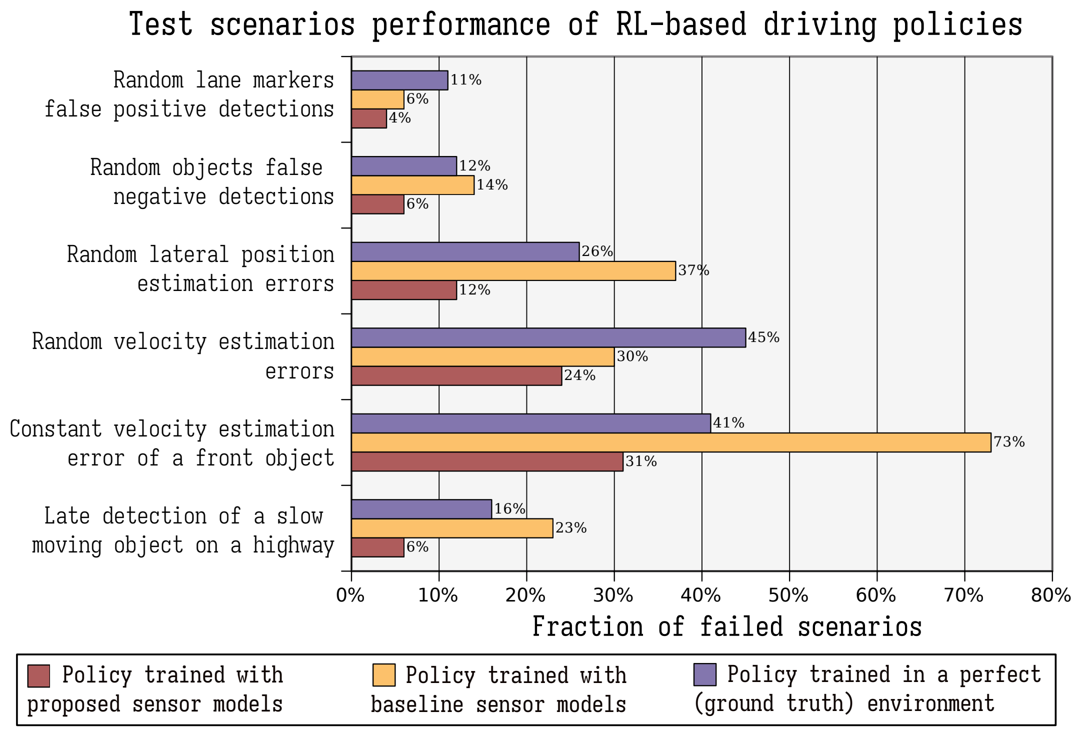


**Figure 4.5.** Comparison of the evaluated policies in predefined test scenarios. The use of the proposed sensor models in the training had a significant positive impact on the robustness of the trained policy.

Proposed sensor models have been utilized to train an RL-based vehicle control policy, alongside other policies trained in a perfect simulation environment, and with baseline sensor models based on commonly used domain randomization techniques. The performance of all the trained policies has been evaluated in a series of long-driving simulation experiments, as well as in a wide set of challenging test scenarios. Performed evaluation has shown a significant advantage of the policy trained with the use of proposed sensor models, compared to the other policies, proving its usefulness in increasing the robustness of developed algorithms.

## 4.3. Adversarial Scenarios Generation

Verification and validation of the vehicle behavior and path planning systems are among the most challenging aspects of their development. A variety of plausible traffic situations, weather conditions, perception errors, and road geometries result in a practically infinite number of scenarios in which the system may need to operate. Ensuring that it will do so is a costly and difficult endeavor, that typically involves various types of software tests, simulation tests, as well as on-road vehicle test drives.

Testing the system in real traffic, while providing the most accurate representation of the conditions in which the system will operate, tends to be extremely costly, and introduces significant risks if the tested system is in relatively early stages of development. For this reason, traffic simulators are often extensively utilized as a testing environment before performing the on-road tests.

Simulation-based tests may be based on either short, predefined scenarios, or large-scale traffic simulations, in which the model of the vehicle controlled by the tested system drives for potentially thousands of kilometers.

The main problem with manually defined test scenarios is the fact, that in Machine-Learning-based systems, there is no guarantee, that the perceived difficulty of the scenario matches the actual one from the system's perspective, i.e. the system may handle objectively challenging scenarios well, but have issues in specific scenarios that are not necessarily objectively difficult. For this reason, manually defined scenarios may fail to uncover potential faults in the system.

Large-scale driving tests in simulated traffic, while effectively exploring massive amounts of possible scenarios, may fail to generate atypical behavior of road users. Since the RL-based policy is also trained in simulation, this is especially dangerous, as it may fail in situations that are rarely observed in simulated environments, such as erratic driving of other road users.

To alleviate these issues, an automated adversarial scenario generation is introduced in the described thesis. The proposed method utilizes stochastic optimization methods to generate control and state trajectories of one or more road users in the vicinity of the ego vehicle. The optimization process incentivizes solutions that present a particular challenge to the evaluated driving policy. This is accomplished by assigning a negative cost value based on safety measures, such as the minimal distance observed between the ego vehicle and other vehicles within the simulated scenario.

The described formulation allows for the generation of individual adversarial scenarios. To expand this to create a wider scenario database, the entire generation process can be executed

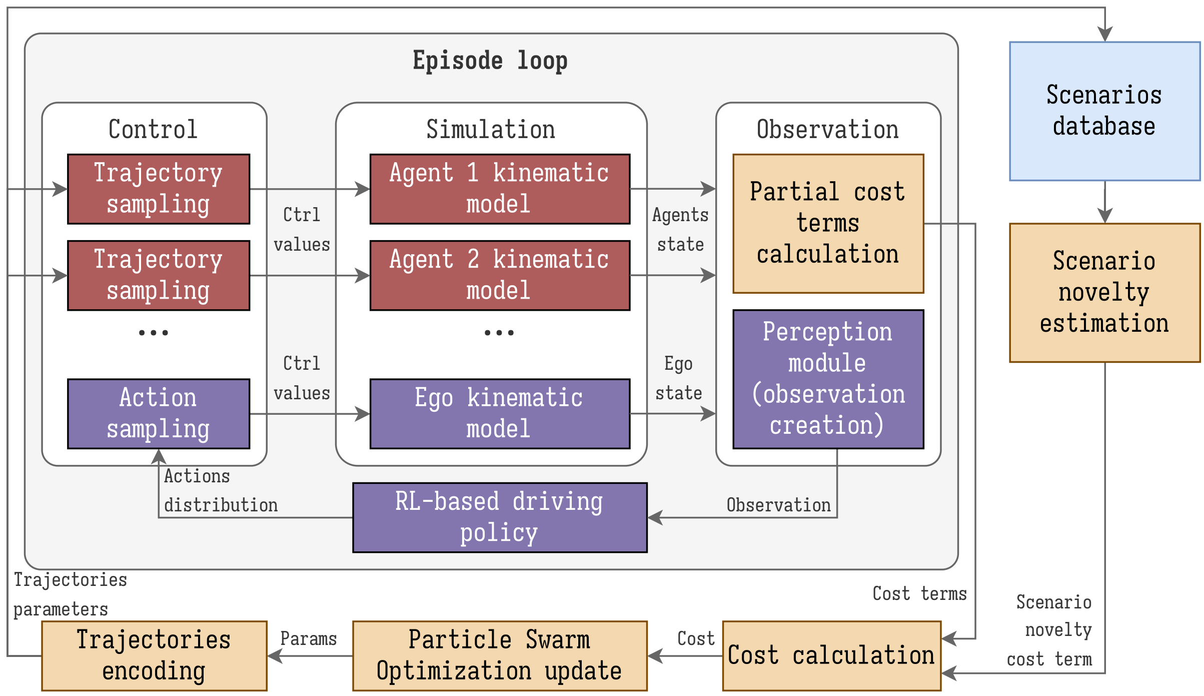


**Figure 4.6.** General idea of the adversarial trajectory generation method proposed in the thesis.

repeatedly with an additional cost term that promotes the discovery of unique scenarios different from those already generated. Adding the cost term that incentivizes large Euclidean distances between the parameters of the generated scenario and other scenarios in the database, each subsequent solution to the optimization problem produces a distinct and novel scenario.

The general idea of the proposed method is presented in Fig. 4.6.

Since, as mentioned in previous sections, perception errors may have a significant impact on planning systems' safety, an additional extension of the proposed method has been introduced to generate adversarial error patterns alongside the trajectories of other vehicles. In this setup, the optimization problem generates not only a state trajectory for each vehicle but also a perceived trajectory, i.e., a series of possibly erroneous state estimations produced by the tested vehicle's perception system. Additional cost terms are introduced to the problem to minimize the differences between the actual and perceived trajectory, allowing for the exploration of scenarios where even tiny errors in state estimation can result in significant safety failures.

Proposed methods were tested in the task of generating adversarial scenarios for an RL-based vehicle control system. The method generated a varied set of adversarial scenarios, exposing several issues in the tested system. An example of a generated adversarial scenario is shown in Fig. 4.7.

The proposed method in the presented example has been able to find a driving pattern of other road users, that led to an incorrect reaction from the evaluated system, leading to a collision. Repeated execution of the method produced further critical scenarios, including behaviors such

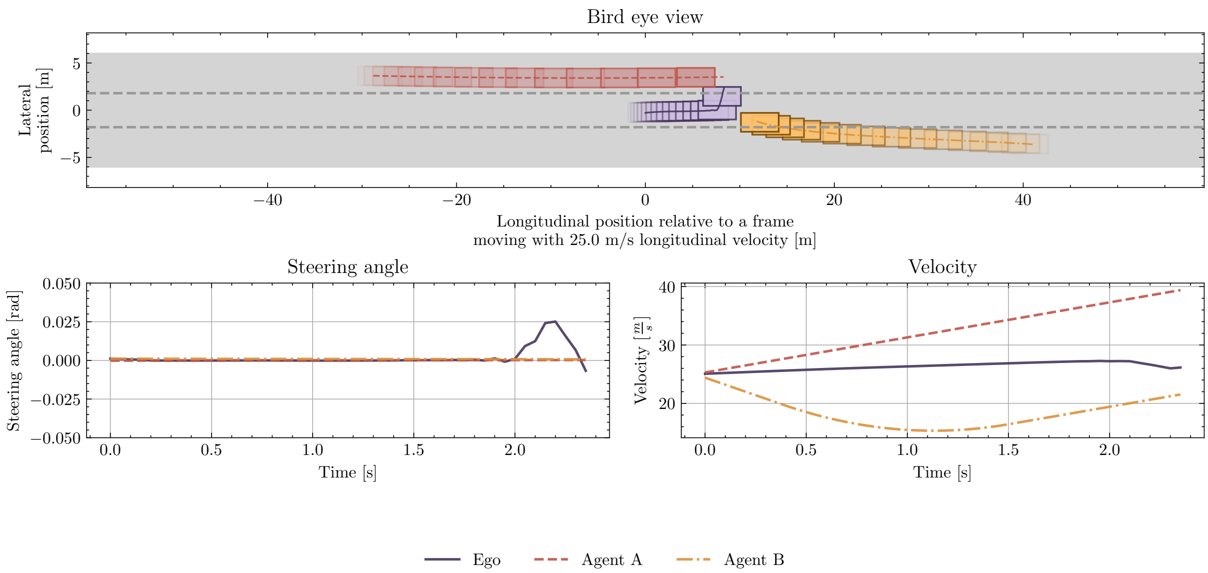


**Figure 4.7.** Example of an adversarial scenario generated by the developed method. In the scenario, the leading vehicle (Agent B) initiates a drastic braking action, followed by an acceleration and a sudden lane change to the left. In response to the abrupt cut-in maneuver, the ego vehicle controlled by the tested algorithm executed an aggressive steering maneuver, resulting in a collision with Agent A, which had been accelerating in the left lane.

as late cut-ins, sudden braking, and erratic acceleration and braking patterns that resulted in erroneous behaviors of the evaluated policy.

It can be noted, that the proposed method can be used to evaluate arbitrary Autonomous Driving and Advanced Driver Assistance Systems, not necessarily based on Machine Learning. Produced scenarios can also be used as a base for further training of evaluated RL-based policies, that would increase their robustness to difficult situations.

# 5. Summary and contributions

Interest in Autonomous Driving and Advanced Driver Assistance Systems has been growing at an exponential rate in recent years, accelerated by revolutionary advancements in the Machine Learning field. While supervised learning is already widely used in perception systems, unsupervised learning methods, such as Reinforcement Learning (RL), seem to be a promising research direction from the perspective of vehicle behavior planning systems. RL-based driving policies can fulfill various Autonomous Driving tasks, being capable of long-term strategical planning and exhibiting situational awareness, behavior prediction abilities, and generalization skills.

To enable the use of such methods in Autonomous Driving applications, however, several challenges must be solved. Transparency of Machine-Learning-based solutions is limited, and thus a special effort is needed to ensure that the resulting system will be operating correctly in difficult situations, such as in the presence of uncertainties, perception systems errors, or atypical situations such as erratic driving of other road users in the vicinity of the controlled vehicle.

In this thesis, several issues regarding the robustness and safety of RL-based driving policies are addressed with novel algorithm proposals. In particular, three main areas are explored.

- Optimization-based trajectory generation for uncertain situations. A novel patent-protected method is introduced to generate vehicle control trajectories in a way, that allows finding several trajectories based on different hypotheses regarding the current and future state of the vehicle's environment, preceded by a common short-term trajectory. The method can be used in conjunction with RL-based solutions to execute maneuvers planned by a high-level driving policy in a safe manner, e.g., in a Fail-Safe Planning setup, that ensures the existence of a collision-free emergency maneuver.

- Models of perception systems intended for training and evaluation of RL-based driving policies. High-level models are proposed for dynamic object detection, as well as lane markers detection systems. Proposed models were used to train a driving policy, that exhibited significantly better robustness compared to policies trained with simple domain randomization techniques.

- Automatic adversarial scenario generation. A novel patent-pending method is introduced for the generation of a database of test scenarios, that pose a particular challenge to the evaluated system. The proposed method can be used to generate challenging scenarios, as well as adversarial perception error patterns, that may expose issues in the tested Machine-Learning-based driving policies. Generated scenarios can be used for evaluation purposes, as well as in training of robust RL-based policies.

Described methods answer the stated research hypotheses, offering a way to address uncertainties in the environment, increasing the robustness of driving policies to the perception errors, as well as introducing an automated testing method intended for active exploration of the automated driving systems' deficiencies.