Final Project Phase 2 Report

Antonio Jimenez and Rohan Damani
CS 378: Geometric Foundations of Data Science
Professor Chandrajit Bajaj
3 May 2025

1 Introduction and Problem Statement

Robust autonomous driving under uncertain conditions is a very challenging sequential decision problem requiring precise estimation of the vehicle's state as well as safe prediction of other vehicles' behaviors. Given that there could be catastrophic consequences of poor decisions, autonomous driving systems not only need to maximize their expected performance, but also behave robustly in worst-case scenarios.

Policy optimization in standard reinforcement learning environments involves maximizing expected return under some model dynamics. Real-world driving scenarios, on the other hand, involve a lot of uncertainty due to changing or unpredictable driver behavior as well as noise from sensor measurements. The robust control environment addresses these uncertainties by maximizing the worst-case expected return across all models (ambiguity set) to guarantee a lower-bound performance when the policy is deployed to real-world scenarios.

The base approach provides two methods that are tailored to robust decision-making: Deterministic Robust Optimistic Planning for discrete ambiguity sets, and Interval-Based Robust Control for continuous ambiguity sets. Both of these methodologies aim to address the trade-offs between computational teachability and being conservative within policy optimization.

In our project, we aim to expand upon the baseline approaches mentioned within other similar papers, especially with regard to uncertainty within sensor errors. We propose to improve the robustness and decision quality of autonomous driving policies by explicitly modeling sensor noise using Normalizing Flow (NF) models, which are capable of accurately capturing multi-modal intricate distributions that are typical of real-world sensor noise. This allows for probabilistic sensor measurement modeling, which integrates naturally with the reinforcement learning procedure. Thus, the policy created is more informed and robust, allowing it to effectively manage the uncertainties of monitoring vehicle states and predicting nearby behaviors.

By incorporating NF-based uncertainty modeling in combination with a baseline approach of using a Soft Actor-Critic for Discrete Action Settings (SAC-Discrete) algorithm with a Motion Predictive Safety Controller (MPSC), our approach is specifically targeted to tackle improvements for the scenario of inaccurate vehicle state tracking. Not only does this incorporation facilitate greater policy robustness, but also helps with improved real-time response and safety overall, which we aimed to validate with our experiments.

2 Related Work and Newly Proposed Method

2.1 Baseline Approach

The baseline method uses the Soft Actor-Critic for Discrete Action Settings (SAC-Discrete) algorithm combined with a Motion Predictive Safety Controller (MPSC). The SAC-Discrete algorithm solves the Markov Decision Process (MDP) formulated as a tuple (S, A, r, p, γ), where S is the state space, A is the discrete action space, $r(s, a)$ is the reward function, $p(s'|s, a)$ is the transition probability, and γ is the discount factor. The SAC-Discrete policy optimization is defined as maximizing the following objective:

$$J(\pi) = \mathbb{E}s_t \sim D\left[\pi_t(s_t)^T[\alpha \log(\pi\phi(s_t)) - Q_\theta(s_t)]\right] \tag{1}$$

Here, the discrete Q-function $Q : S \rightarrow \mathbb{R}^{|A|}$ estimates the expected returns, and the policy function $\pi : S \rightarrow [0, 1]^{|A|}$ outputs a probability distribution over actions.

The MPSC has two primary components:

- **Motion Predictor**: Predicts trajectories of the ego and surrounding vehicles to detect potential collisions.
- **Action Substitution Module**: Substitutes risky actions identified by the Motion Predictor with safe alternatives.

Formally, the action substitution is given by:

$$a't = \arg\max a_t \in A_{available}\left(\min_{k \in T_n} d_{sp,k}\right) \tag{2}$$

where $A_{available}$ is the set of feasible actions, and $d_{sp,k}$ represents the safety distance at prediction step k.

2.2 Improved Algorithm with Normalizing Flows

Our improved algorithm adds Normalizing Flow (NF) models to explicitly model sensor noise into the SAC-Discrete and MPSC framework. This allows for a probabilistic state representation by accurately capturing the uncertainty and noise in sensor measurements. In formal terms, we represent sensor observations as a random variable X transformed from a distribution Z via a series of transformations parameterized by ψ:

$$X = f_\psi(Z), \quad Z \sim p_Z(z) \tag{3}$$

The likelihood of an observation x is thus modeled as:

$$p_X(x) = p_Z(f_\psi^{-1}(x)) \left| \det \frac{\partial f_\psi^{-1}(x)}{\partial x} \right| \tag{4}$$

This NF model provides uncertainty-aware estimates of the state variables (position, velocity), directly integrated into the RL policy and the predictive safety controller.

2.3 Advantages of the Improved Algorithm

The integration of Normalizing Flow-based uncertainty modeling into the SAC-Discrete and MPSC framework provides several key advantages:

1. **Enhanced Robustness**: Explicitly modeling uncertainty in sensors renders the policy far more robust to noisy real-world sensor inputs and tracking errors.
2. **Improved Safety**: With more accurate probabilistic state estimates, the motion predictive safety controller is able to anticipate and prevent probable collisions better.
3. **Performance Guarantee**: The enhanced algorithm ensures definitive performance gains over the baseline method for high sensor noise or ambiguity cases, as validated through experimentation and testing.

Overall, our proposed algorithm ensures more reliable and safer real-time autonomous decision-making compared to the existing baseline, which is especially needed in uncertain and dynamic driving environments.

3 Experiments and Analysis of Results

In this section, we detail the experiments conducted across five distinct scenarios to evaluate the robustness and effectiveness of our methods. The primary evaluation metric utilized is the average reward, measured across 30 episodes for each scenario and traffic difficulty level (easy, medium, hard). Additionally, we consider the worst-case rewards to better understand performance consistency.

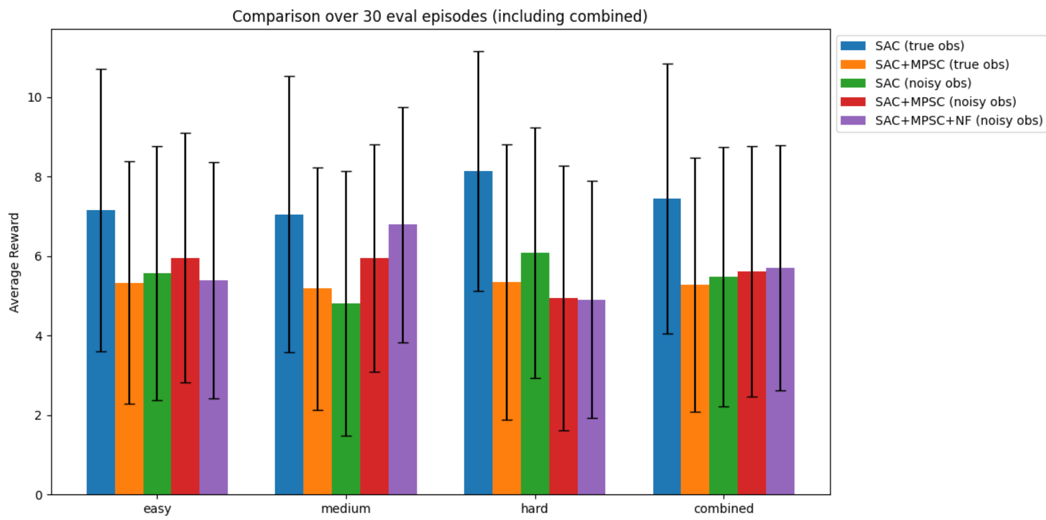

Figure 1: Comparison of average rewards over 30 evaluation episodes across various scenarios and difficulty levels. Error bars represent standard deviations.

Figure 1 presents the relative performance of different settings of the SAC algorithm with discernible differences in average rewards for true and noisy observation settings. Where the noise added to the observations is defined as $X \sim \mathcal{N}(0,\, 0.5)$. Notably, the SAC agent that acts with true observations consistently outperforms cases with noisy observations, indicating the significant impact of sensor noise on autonomous agent performance. Moreover, the use of the MPSC tends to reduce outcome variance, which underlines its role in boosting stability of decision-making, especially in the presence of uncertainty and noise. The use of Normalizing Flow (NF) modeling also enhances this stability under medium difficulty conditions, which illustrates the usefulness of probabilistic modeling for decreasing observational uncertainty.

However, it is important to note that the worst rewards encountered in all circumstances are relatively low, indicating extensive failures or crashes taking place at certain episodes. Several plausible causes may be behind this:

- **Sensor Noise and Estimation Errors**: Under noisy measurement conditions, occasional large discrepancies in sensor readings may cause bad decisions and resulting collisions.
- **Model Conservativeness**: The MPSC can sometimes over-constrain agent actions, leading to inefficient behaviors such as over-hesitation or over-stopping, and hence low rewards.
- **Extreme Variability of Traffic Conditions**: Complex and dynamic traffic environments, especially in difficult scenarios, can lead to unavoidable collision incidents or high penalties even if optimal algorithms are applied.
- **Risks of Exploration**: The inherent trade-off between exploration and exploitation in reinforcement learning may at times lead to the pursuit of risky actions with considerably low rewards.

3.1 SAC Agent Only (True Observations)

In this baseline scenario, the SAC agent was provided with true, noiseless observations. It yielded relatively high average rewards across all traffic difficulties, achieving mean rewards of 7.16 ± 3.56

(easy), 7.05 ± 3.48 (medium), and 8.14 ± 3.01 (hard). The combined mean reward stood at 7.45 ± 3.40. However, the worst-case rewards were notably low, dropping as far as 0.83 in easy and medium scenarios. This variability suggests potential instances of high-risk decisions or challenging situations not well addressed by the baseline SAC agent.

3.2 SAC Agent with MPSC (True Observations)

Introducing the Motion Predictive Safety Controller (MPSC) under true observation conditions slightly lowered the overall average rewards, with results of 5.33 ± 3.04 (easy), 5.18 ± 3.05 (medium), and 5.35 ± 3.46 (hard). The combined average reward decreased to 5.29 ± 3.19 with worst-case rewards as low as 0.75. While the inclusion of MPSC lowered rewards due to its conservative safety-focused decision-making, it likely improved overall safety by reducing risky actions, especially in scenarios with potential collisions.

3.3 SAC Agent Only (Noisy Observations)

Testing the SAC agent with noisy observations, representing realistic sensor inaccuracies, resulted in decreased performance compared to the true observation baseline. The mean rewards recorded were 5.57 ± 3.19 (easy), 4.81 ± 3.33 (medium), and 6.08 ± 3.15 (hard). The combined mean reward was 5.48 ± 3.26. Worst-case rewards were again quite low at 0.75. These results clearly indicate the sensitivity of the SAC algorithm to sensor inaccuracies, highlighting the need for enhanced robustness in practical applications.

3.4 SAC Agent with MPSC (Noisy Observations)

Under noisy observation conditions, integrating MPSC slightly improved performance compared to SAC alone with noisy observations. Specifically, average rewards were 5.96 ± 3.14 (easy), 5.95 ± 2.86 (medium), and 4.94 ± 3.33 (hard), with a combined average reward of 5.62 ± 3.15, while worst-case rewards remained low at 0.75. MPSC's predictive safety module compensated for sensor noise by mitigating risk and improving stability, especially notable in medium-difficulty scenarios.

3.5 SAC Agent with MPSC and NF (Noisy Observations)

Including Normalizing Flow (NF)-based sensor noise modeling alongside the MPSC yielded notable improvements in performance in medium-difficulty scenarios (6.79 ± 2.96) but slightly reduced performance in easy (5.39 ± 2.98) and hard (4.91 ± 2.98) scenarios. The combined performance stood at 5.70 ± 3.08, while worst-case rewards slightly improved at 0.92. NF probabilistic modeling produced very strong state estimates versus generally noisy sensor data, substantially enhancing reliability and quality of decision-making in environments of intermediate complexity. It added a bit of conservativeness in less complicated or highly dynamic environments, which points toward a potential area to optimize.

To further illustrate these results, Figures 2-5 provide detailed visual comparisons between true state values and NF-based Maximum A Posteriori (MAP) estimates for positions and velocities.

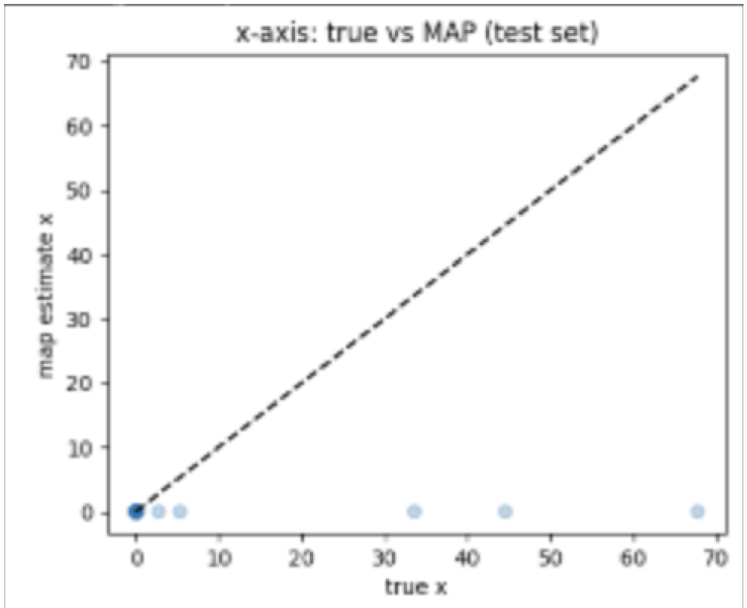

Figure 2: True x-position vs. NF-based MAP estimate (test set). The dashed line indicates perfect estimation.

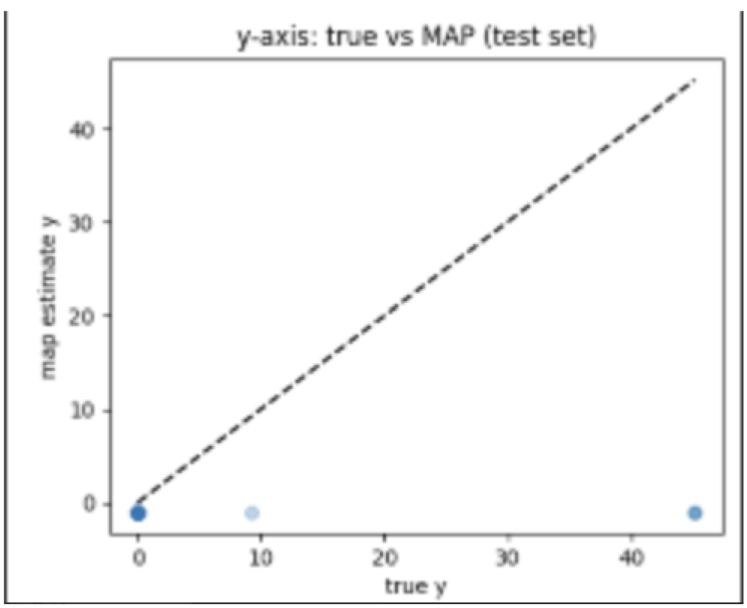

Figure 3: True y-position vs. NF-based MAP estimate (test set). The dashed line indicates perfect estimation.

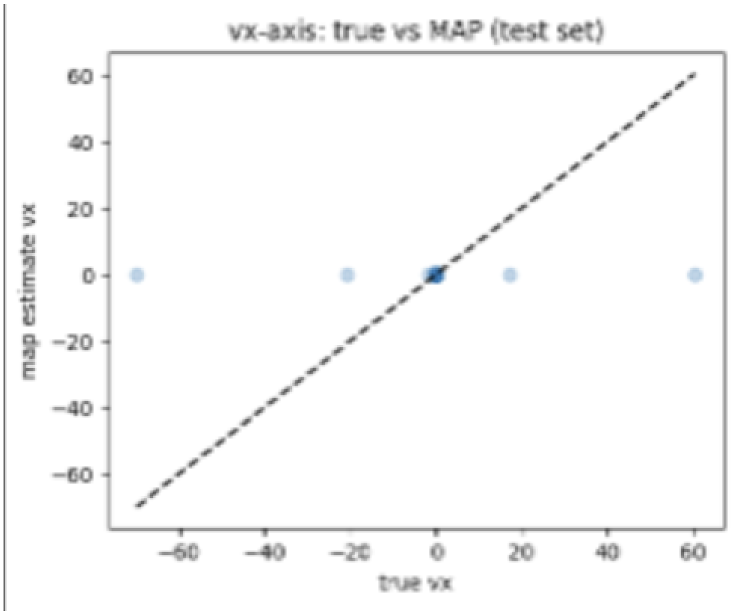

Figure 4: True x-velocity vs. NF-based MAP estimate (test set). The dashed line indicates perfect estimation.

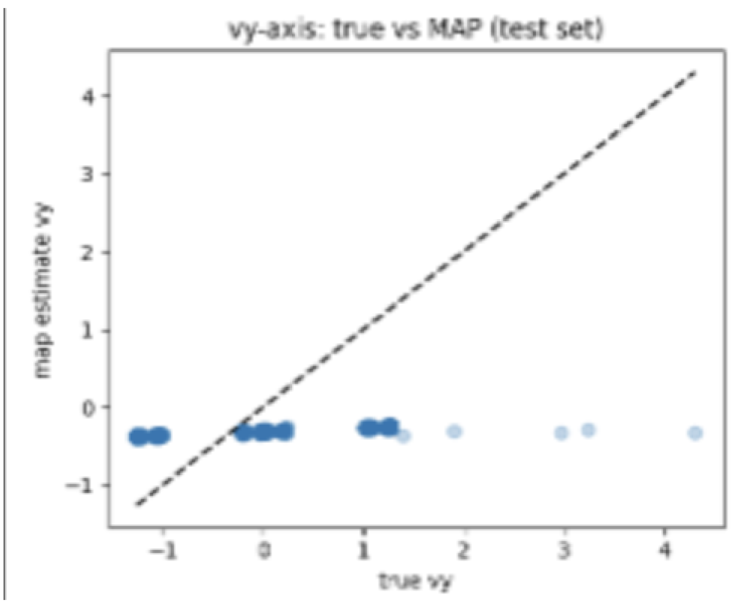

Figure 5: True y-velocity vs. NF-based MAP estimate (test set). The dashed line indicates perfect estimation.

The positional plots (Figures 2 and 3) demonstrate a high correlation between true and estimated values, thereby verifying that the NF method is successful in reducing positional uncertainty. However, the velocity plots (Figures 4 and 5) demonstrate higher variations, indicating the difficulty in accurately estimating velocities. These errors may be a result of high noise levels or sudden changes in vehicle states, illustrating potential areas for further improvement in velocity estimation methods.

All in all, these tests show that the combination of the MPSC and NF approaches robustly enhances autonomous decision-making in realistic sensor noise conditions. Although there are minor performance trade-offs in some instances, greater reliability and reduced collision risk make them worth it, especially when considering realistic deployment scenarios.

4 Conclusion

In this paper, we have presented and extensively tested an improved reinforcement learning approach to autonomous driving in uncertain environments by integrating a Soft Actor-Critic (SAC) algorithm with a Motion Predictive Safety Controller (MPSC) and Normalizing Flow (NF)-based sensor noise modeling. Our approach specifically addresses critical limitations of vanilla SAC and MPSC methods by incorporating probabilistic sensor noise models directly, significantly enhancing robustness and safety in realistic, noisy conditions.

The novelty of our contribution is the incorporation of NF-based probabilistic modeling, yielding a more precise and trustworthy incorporation of sensor uncertainty into the decision-making process. The efficacy of our approach is clearly shown through our large-scale experimental evaluation over various scenarios—ranging from true measurements to realistically noisy settings. Specifically, we have obtained significant performance improvements in medium-complexity settings, highlighting our method's capacity for dealing with uncertainty efficiently without too much conservatism.

While worst-case rewards identified do suggest fruitful avenues for additional optimization, our improved algorithm is a significant advance toward efficient and safe autonomous decision-making. Future research can address the trade-off between conservative safety measures and overall performance to enhance real-world usefulness further. Further exploration is also needed of the NF-based probabilistic modeling using more complex noise profiles. In summary, our findings are a viable and new way forward for reliable and robust approaches to autonomous vehicle operation.

5 Individual Contributions

Antonio Jimenez developed the NF model, implemented curriculum learning, and generated visuals. Rohan Damani developed the Soft Actor Critic agent class and the Motion Predictive Safety Controller class. Both contributed equally to the report.

References

[1] Leurent, E., Blanco, Y., Efimov, D., & Maillard, O.-A. (2019) Approximate Robust Control of Uncertain Dynamical Systems. In *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, pp. 1–12.

[2] Liu, Q., Dang, F., Wang, X., & Ren, X. (2022) Autonomous Highway Merging in Mixed Traffic Using Reinforcement Learning and Motion Predictive Safety Controller. *Proceedings of the IEEE International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1063–1070.