MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

# Safe Interactive Motion Planning by Differentiable Optimal Control and Online Preference Learning

Chavez Armijos, Andres; Berntorp, Karl; Di Cairano, Stefano

TR2025-104 July 09, 2025

**Abstract**

We present an interactive motion planner that integrates online learning of human driver preferences with parametric control barrier functions. Using stochastic models with Gaussian disturbances to capture human-driven vehicle behavior uncertainty, we update parameters in real-time parameter by Kalman filtering while ensuring safety by control barrier functions. A case study on highway lane-changing tasks demonstrates improved traffic flow, reduced disruptions, and lighter actuation effort compared to non-adaptive algorithms.

*American Control Conference (ACC) 2025*



Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

# Safe Interactive Motion Planning by Differentiable Optimal Control and Online Preference Learning

Andres S. Chavez Armijos, Karl Berntorp, and Stefano Di Cairano,

***Abstract*— We present an interactive motion planner that integrates online learning of human driver preferences with parametric control barrier functions. Using stochastic models with Gaussian disturbances to capture human-driven vehicle behavior uncertainty, we update parameters in real-time parameter by Kalman filtering while ensuring safety by control barrier functions. A case study on highway lane-changing tasks demonstrates improved traffic flow, reduced disruptions, and lighter actuation effort compared to non-adaptive algorithms.**

## I. Introduction

The development of autonomous vehicles (AVs) has gained significant interest due to their potential to improve safety and optimize traffic flow. However, the successful deployment of AVs in real-world environments requires robust motion planners that can manage complex interactions with human-driven vehicles (HDVs) exhibiting diverse driving behaviors [1]. Traditional approaches assume HDVs follow fixed rules or rely on simplified models, which may result in continuous plan adjustments when interacting with HDVs due to varying driving styles [2].

Recently, learning-based methods have shown promise for learning policies from interactions with other agents [3], [4], at the price of extensive training data and possible challenges in generalization. Other approaches are based on modeling vehicle interactions as multi-agent games [5], which can be computationally intensive and require accurate models of other agents' behaviors, which are hard to obtain. Combining learning and optimization methods allows for maximizing both strengths [6], and also for ensuring safety through control barrier functions (CBFs) designed from data [7], [8]. However, this may still require a significant amount of data.

In this paper we propose an approach that leverages automated controller calibration by uses an Extended Kalman filter (EKF) [9] to learn stochastic models of HDV behaviors and CBFs for safety, which allows the ego vehicle to adapt as it interacts with HDVs. We demonstrate the effectiveness of our approach in highway lane-changing scenarios, where the proposed adaptive algorithm show reduced traffic flow disruption compared to non-adaptive methods.

Next, Section II defines the problem, Section III describes the control design and Section IV describes the proposed learning algorithm. Section V present a case study on lane-changing maneuvers and Section VI present the conclusions.

S. Di Cairano, and K. Berntorp are with the Mitsubishi Electric Research Labs (MERL), Cambridge, MA. (e-mail: {dicairano, berntorp}@merl.com)

A. S. Chavez Armijos is with Boston University, Brookline, MA. He was an intern at MERL during this work. (e-mail: aschavez@bu.edu)

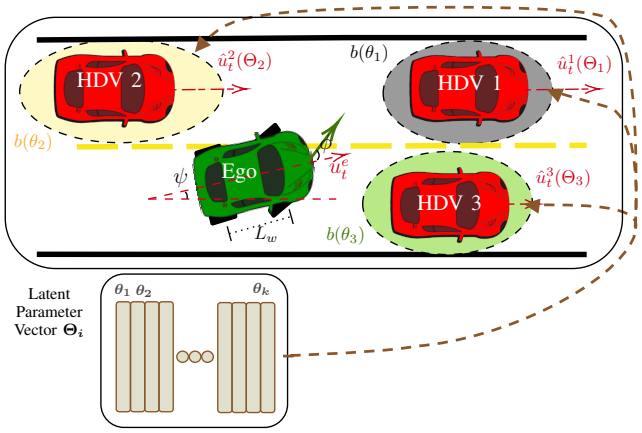


Fig. 1: Interactive scenario with ego vehicle and HDVs. The latent parameter vector $\Theta$ affects each vehicle's trajectory as well as its minimum ellipsoidal safety distances.

*Preliminaries.* Consider the control-affine system

$$\dot{\boldsymbol{x}} = f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u}, \tag{1}$$

where $x \in \mathcal{X} \subset \mathbb{R}^{n_x}$ and $u \in \mathcal{U} \subset \mathbb{R}^{n_u}$ are the state and control input vectors, respectively, $\mathcal{X}$ and $\mathcal{U}$ are the corresponding admissible sets, and $f : \mathbb{R}^n \to \mathbb{R}^n$ $g : \mathbb{R}^m \to \mathbb{R}^{n\times m}$ are locally Lipschitz continuous.

*Definition 1.1:* A continuously differentiable function $b : \mathbb{R}^n \to \mathbb{R}$ is a control barrier function [10] (CBF) for (1) if there exists a $\mathcal{K}_\infty$ function [11] $\alpha(\cdot)$ such that $\sup_{\boldsymbol{u}\in\mathcal{U}} [\mathcal{L}_f b(\boldsymbol{x}) + \mathcal{L}_g b(\boldsymbol{x})\boldsymbol{u}] \geq -\alpha\left(b(\boldsymbol{x})\right)$ for all $\boldsymbol{x} \in \mathcal{C}$, $\mathcal{C} := \{\boldsymbol{x} \in \mathbb{R}^n : b(\boldsymbol{x}) \geq 0\}$, where $\mathcal{L}_h$ denotes the Lie derivative along $h$.

*Definition 1.2:* A function $V : \mathbb{R}^n \to \mathbb{R}$ such that there exist $c_1, c_2, c_3 > 0$, $c_1|\boldsymbol{x}|^2 \leq V(\boldsymbol{x}) \leq c_2|\boldsymbol{x}|^2$, $\inf_{\boldsymbol{u}\in U} [\mathcal{L}_f V(\boldsymbol{x}) + \mathcal{L}_g V(\boldsymbol{x})\boldsymbol{u} + c_3 V(\boldsymbol{x})] \leq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$, is a control Lyapunov function (CLF) for (1).

## II. Problem Formulation

Consider a scenario $\mathcal{S}$ where an autonomous ego vehicle $e$ drives around $N$ HDVs $\mathcal{H} = \{1, 2, \ldots, N\}$. The $i^{th}$ HDV's behavior, such as safety distances and actuation effort preferences, is influenced by the latent variable $\Theta_i \in \mathbb{R}^d$. Figure 1 shows an example scenario where the preferences are the minimum safety distances for each HDV represented by ellipsoids of different sizes. The predicted trajectory $\hat{\xi}_i(\Theta_i)$ for each vehicle depends on latent parameters $\Theta_i$. Thus, for predicting the HDVs control actions for an ego vehicle trajectory, we need to learn the HDVs latent variables.

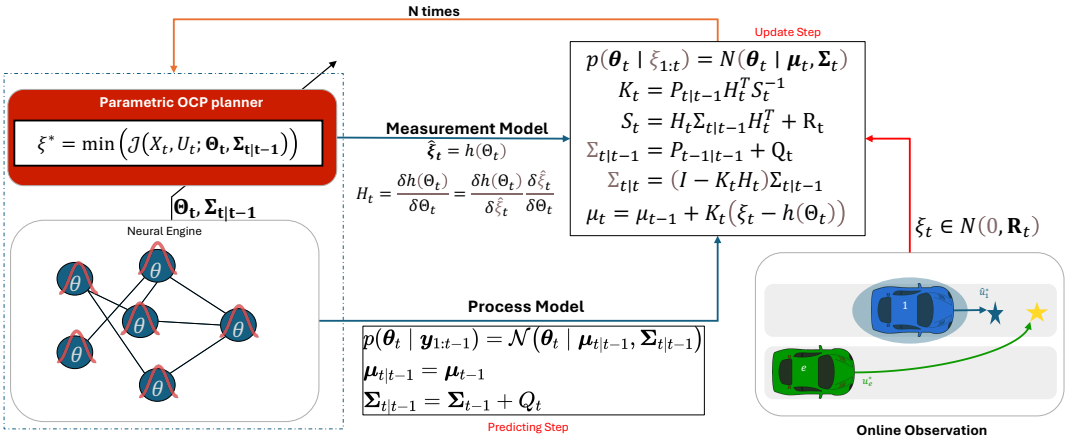


Fig. 2: Online adaptation of motion planning using EKF to update HDVs latent parameters (e.g., safety preferences).

Our objective is to design a control policy that solves that controls the ego vehicle towards its goal while learning the HDV behavior and ensuring avoidance of collision to a sufficient degrees of confidence. To this end we: $(i)$ Design an optimal control problem that models the interactions between the ego vehicle and surrounding HDVs, considering dynamic constraints, safety requirements, and vehicle observations; $(ii)$ construct an estimate of the safety function $b_i(\cdot)$ for each vehicle $i$ to model its reactions and behaviors; $(iii)$ learn the maximum likelihood set of HDV latent parameter vectors $\Theta_i$, $\boldsymbol{\Theta}$, based on the dataset $\mathcal{D}$ of past interactions; $(iv)$ solve the optimal control problem to determine the action of the ego vehicle and the expected reactions of the HDVs. The schematic of the overall method is shown in Figure 2.

## III. Interaction-based Parametric Optimal Control

First, we design a Parametric Optimal Control Problem (POCP) that operates both as a planner for the ego vehicle and as a trajectory estimator for HDV.

### A. Dynamic Constraints

We model vehicles $i \in \{\mathcal{H}, e\}$ by the kinematic single-track model [12]

$$\underbrace{\begin{bmatrix} x_i \\ y_i \\ \psi_i \\ v_i \end{bmatrix}}_{\dot{\boldsymbol{x}}_i} = \underbrace{\begin{bmatrix} v_i \cos(\psi_i) \\ v_i \sin(\psi_i) \\ 0 \\ 0 \end{bmatrix}}_{f(\boldsymbol{x}_i(t))} + \underbrace{\begin{bmatrix} 0 & -v \sin(\psi_i) \\ 0 & v_i \cos(\psi_i) \\ 0 & v_i / L_w \\ 1 & 0 \end{bmatrix}}_{g(\boldsymbol{x}_i(t), \boldsymbol{u}_i(t))} \begin{bmatrix} u_i \\ \phi_i \end{bmatrix}, \tag{2}$$

where $x_i$, $y_i$, $\psi_i$, $v_i$ are the longitudinal and lateral position, heading angle and speed, $u_i$ is the acceleration, $\phi_i$ is the steering angle, and $L_w$ is the wheelbase length.

Model (2) is subject to state and input constraints $\boldsymbol{x}_i \in \mathcal{X}_i$ and $\boldsymbol{u}_i \in \mathcal{U}_i$, which include speed limits, road boundaries, lane restrictions, and vehicle limitations. The combined traffic model is obtained by concatenating states and inputs for each vehicle, $\boldsymbol{X}_t \in \mathbb{R}^{n_x \times (N+1)}$, $\boldsymbol{U}_t \in \mathbb{R}^{n_u \times (N+1)}$.

### B. Parametric Cost Function

The cost function for the interacting vehicles is based on deviation from an intended trajectory. Here, an intention is defined as a predicted control input or goal.

*Assumption 1 (Availability of Intentions):* The ego vehicle $e$ has access to a "stand-alone" trajectory predictor $\mathcal{F}(\boldsymbol{X}(t_0), \mathcal{S}(t_0)) : \mathbb{R}^{4\times N} \to \mathbb{R}^{2\times N\times T}$ that maps the traffic state $\boldsymbol{X}(t_0)$ and scenario $\mathcal{S}(t_0)$ to the "stand-alone" intentions $\{\hat{\boldsymbol{x}}_t^i, \hat{\boldsymbol{u}}_t^i\}$ for $t \in (t_0, T]$ for each vehicle $i \in \mathcal{H}$.

Based on Assumption 1, the ego vehicle can predict each HDV's intended path $\hat{\xi}_i = \{\hat{\boldsymbol{x}}_t^i, \hat{\boldsymbol{u}}_t^i\}$ in isolation, with no interactions. Any predictor, e.g., [13], [14], can be used. Thus, the cost function is

$$\mathcal{J}(\boldsymbol{U}_k; \boldsymbol{\Theta}) = \underbrace{\|\boldsymbol{u}_t^e\|_{\boldsymbol{Q}_e}}_{\mathcal{J}_e(\boldsymbol{u}_t^e)} + \sum_{i=1}^{N} \underbrace{\left(\|\boldsymbol{u}_t^i - \hat{\boldsymbol{u}}_t^i\|_{\boldsymbol{Q}_i^{\theta_{\text{dev}}}} + \|\boldsymbol{u}_t^i\|_{\boldsymbol{Q}_i^{\theta_u}}\right)}_{\mathcal{J}_i(\boldsymbol{u}_t^i, \hat{u}_t^i; \boldsymbol{\Theta}^i)} \tag{3}$$

where $\boldsymbol{Q}_e, \boldsymbol{Q}_i^{\theta_{\text{dev}}}, \boldsymbol{Q}_i^{\theta_u} \geq 0$ are nonnegative weight matrices. For each HDV $i \in \mathcal{H}$, $\mathcal{J}_i(\cdot)$ penalizes deviation from the predicted control input $\hat{\boldsymbol{u}}_t^i$ and minimizes the actuation effort using the parametric matrices, while $\mathcal{J}_e(\cdot)$ minimizes the actuation efforts of the ego vehicle.

### C. Parametric Safety Description

We consider *joint safety* where both vehicles contribute to maintaining safety.

*Definition 3.1 (Joint Safety):* Let $S_S^i := \{\boldsymbol{x}_i \mid g_S(\boldsymbol{x}_i) \geq 0\}$ and $S_S^j := \{\boldsymbol{x}_j \mid g_S(\boldsymbol{x}_j) \geq 0\}$ denote the safety sets for vehicles $i$ and $j$, respectively. The joint safety set is

$$S^S := S_S^i \cap S_S^j = \{\boldsymbol{x} \mid \Psi_S^{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0\}, \tag{4}$$

where $\Psi_S^{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ characterizes the intersection.
Based on Definition 3.1, we make the following assumption.

*Assumption 2 (Rational Drivers):* Every HDV on the road is considered rational and safe, and thus avoids situations where a collision cannot be averted.

Under Assumption 2 and Definition 3.1, we model the safety set as a belief dependent on $\boldsymbol{\Theta}$. Then, safety is defined

using the ellipsoidal CBF constraint

$$\Psi_S^{ij}(\boldsymbol{x}_t^i, \boldsymbol{x}_t^j) := \frac{\left(x_t^i - x_t^j\right)^2}{r_x^2} + \frac{\left(y_t^i - y_t^j\right)^2}{b^2} - 1 \geq 0, \quad (5)$$

where $r_x = \frac{\left(v_t^j - v_t^i\right)^2}{d_{\max}} + a$ depends on the velocity difference between the vehicles $v_t^i$ and $v_t^j$, the maximum deceleration $d_{\max}$, and a baseline semiaxis $a$, while $b$ remains constant. We derive the CBF constraint as

$$\mathcal{L}_{f_i}\Psi_S^{ij} + \mathcal{L}_{g_i}\Psi_S^{ij}\boldsymbol{u}_i + \mathcal{L}_{f_j}\Psi_S^{ij} + \mathcal{L}_{g_j}\Psi_S^{ij}\boldsymbol{u}_j \geq -\alpha\left(\Psi_S^{ij}\right). \quad (6)$$

The function $\alpha\left(\Psi_S^{ij}\right)$ represents the safety margin for constraint violation. The safety preference of a vehicle pair $(i, j)$ is parameterized by the safety margin function $\alpha\left(\Psi_S^{ij}\right)$. A larger $\alpha(\cdot)$ indicates a more cautious preference, while a smaller value represents a more aggressive preference.

Thus, the safety margin $\alpha\left(\Psi_S^{ij}\right)$ is parameterized by $\boldsymbol{\Theta}^{ij}$. We define $\alpha$ as an odd polynomial

$$\alpha_{ij}(\boldsymbol{x}_t^i, \boldsymbol{x}_t^j; \boldsymbol{\Theta}^{ij}) = \sum_{k=0}^{d} \theta_{2k+1} \Psi_S^{ij}(\boldsymbol{x}_t^i, \boldsymbol{x}_t^j)^{2k+1}, \quad (7)$$

where $d$ is the degree, and $\theta_{2k+1}$ are the parameters. Then, $\alpha$ in (6) is parametrized by (7) resulting in the constraint

$$G_s^{ij}(\boldsymbol{x}_t^i, \boldsymbol{x}_t^j) \geq -\alpha_{ij}(\boldsymbol{x}_t^i, \boldsymbol{x}_t^j; \boldsymbol{\Theta}^{ij}), \quad (8)$$

where $\Theta^{ij} \in \mathbb{R}^m$ contains the parameters describing the safety margin and $G_s^{ij}(\boldsymbol{x}_t^i, \boldsymbol{x}_t^j)$ describes the LHS of (6).

### D. Safety Robustification

The *joint safety* parameter vector $\boldsymbol{\Theta}^{ei}$ for the ego vehicle $e$ and an adjacent vehicle $i$ contains uncertainties. Thus, we robustify the safety constraints as chance constraints

$$\Pr\left(G_s^{ej}(\boldsymbol{x}_t^e, \boldsymbol{x}_t^i) \geq -\alpha_{ei}(\boldsymbol{x}_t^e, \boldsymbol{x}_t^i; \boldsymbol{\Theta}^{ei})\right) \leq 1 - \epsilon, \ \ \forall i \in \mathcal{H}, \quad (9)$$

ensuring the probability of unsafe interactions between $e$ and any other vehicle $i$ remains below a risk threshold $\epsilon$ while accounting for the uncertainty. The constraint is parameterized by $\boldsymbol{\Theta}^{ei}$, which captures the interaction between the ego vehicle and each $i$. It is assumed that every parameter $\theta_i$ is corrupted by noise $\nu \sim \mathcal{N}(0, \omega)$ with covariance matrix $C_{\Theta^{ei}} = \mathrm{diag}([\omega_1, \ldots, \omega_m])$. We reformulate (6), (7) as

$$\alpha_{ej}(\boldsymbol{x}_t^e, \boldsymbol{x}_t^i; \boldsymbol{\Theta}^{ei}) = \sum_{k=0}^{d} \left(\theta_{2k+1} + \omega_{2k+1}\right) \Psi_S^{ei}(\boldsymbol{x}_t^e, \boldsymbol{x}_t^i)^{2k+1}, \quad (10)$$

$$\Pr\left[\mathcal{L}_{f_e}\Psi_S^{ei} + \mathcal{L}_{g_e}\Psi_S^{ei}\boldsymbol{u}_i + \mathcal{L}_{f_i}\Psi_S^{ei} + \mathcal{L}_{g_i}\Psi_S^{ei}\boldsymbol{u}_j \geq -\alpha_{ei}(\boldsymbol{x}_t^e, \boldsymbol{x}_t^i; \boldsymbol{\Theta}^{ei})\right] \geq 1 - \epsilon, . \quad (11)$$

Letting $p = 1 - \epsilon$ and $\Phi^{-1}(p)$ denote the $p$-th quantile of the standard normal distribution. The safety constraint using a radial basis function is

$$\begin{aligned} &\forall i \in \mathcal{H}, \quad G_r(\boldsymbol{x}_t^e, \boldsymbol{x}_t^i; \boldsymbol{\Theta}^{ei}) := \\ &\mathcal{L}_{f_i}\Psi_S^{ei} + \mathcal{L}_{g_i}\Psi_S^{ei}\boldsymbol{u}_i + \mathcal{L}_{f_j}\Psi_S^{ei} + \mathcal{L}_{g_j}\Psi_S^{ei}\boldsymbol{u}_j \geq \\ &-\alpha_{ei}(\boldsymbol{x}_t^e, \boldsymbol{x}_t^i; \boldsymbol{\Theta}^{ei}) - \Phi^{-1}(p)\sqrt{\mathrm{tr}\left(\Sigma_\Omega\right)}, \end{aligned} \quad (12)$$

where $\Phi^{-1}(p)\sqrt{\mathrm{tr}\left(\Sigma_\Omega\right)}$ adjusts the safety margin to account for the uncertainty. The robustified constraint ensures that the probability of violating the safety condition remains below $\epsilon$, while accounting for uncertainty in the learned parameters.

### E. Parametric Goal Description

To model the ego vehicle's goal-reaching objectives (e.g., reaching a merging point or a specific lane), we construct a deterministic CLF that minimizes the distance between the current state $\boldsymbol{x}_t^e$ and the goal $\boldsymbol{x}_{\text{goal}}^e$

$$\begin{aligned} \mathcal{V}_e(\boldsymbol{x}_t^e, \boldsymbol{x}_{\text{goal}}^e) &= ||\boldsymbol{x}_t^e - \boldsymbol{x}_{\text{goal}}^e||^2, \\ \mathcal{L}_{g_e}\mathcal{V}_e(\boldsymbol{x}_t^e, \boldsymbol{x}_{\text{goal}}^e)\boldsymbol{u} + \mathcal{L}_{f_e}\mathcal{V}_e(\boldsymbol{x}_t^e, \boldsymbol{x}_{\text{goal}}^e) + p_2\mathcal{V}_e(\boldsymbol{x}_t^e, \boldsymbol{x}_{\text{goal}}^e) &\leq \boldsymbol{\varepsilon}_e, \end{aligned} \quad (13)$$

where $p_2$ is a linear class-$\mathcal{K}$ function parameter and the slack $\boldsymbol{\varepsilon}_e$ allows for relaxing the constraint to avoid infeasibility.

For each vehicle $i \in \mathcal{H}$, we construct a CLF $\mathcal{V}_i(\boldsymbol{x}_t^i)$ to account for deviations from the predicted intention $\hat{\boldsymbol{x}}_{\boldsymbol{t}}^{\boldsymbol{i}}$

$$\begin{aligned} \mathcal{V}_i(\boldsymbol{x}_t^i, \hat{\boldsymbol{x}}_t^i) &= ||\boldsymbol{x}_t^i - \hat{\boldsymbol{x}}_t^i||^2, \qquad \forall i \in \mathcal{H}, \\ \mathcal{L}_{g_i}\mathcal{V}_i(x_t^i, \hat{x}_t^i)\boldsymbol{u} + \mathcal{L}_{f_i}\mathcal{V}_i(\boldsymbol{x}_t^i, \hat{\boldsymbol{x}}_t^i) + \theta_g^i\mathcal{V}_i(\boldsymbol{x}_t^i, \hat{\boldsymbol{x}}_t^i) &\leq \boldsymbol{\varepsilon}_i, \end{aligned} \quad (14)$$

where $\theta_g^i$ is a linear parametric class-$\mathcal{K}$ function that ensures $\mathcal{V}_i$ decreases over time, driving $\boldsymbol{x}_t^i$ toward $\hat{\boldsymbol{x}}_t^i$. Unlike (9), this constraint does not require robustification since we cannot control the assumed risk for vehicle $i$.

### F. Robust Interactive CBF-CLF QP

We discretize the time interval $[0, T]$ into equal steps of duration $\Delta t$ with a zero-order hold. To satisfy Definition 1.1, we formulate the Quadratic Program (QP) as

$$\begin{aligned} \min_{\boldsymbol{U}_t, \mathcal{E}_t} \quad & \underbrace{\|\boldsymbol{u}_t^e\|_{Q_e}}_{\mathcal{J}_e(\boldsymbol{u}_t^e)} + \sum_{i=1}^{N} \underbrace{\mathbb{E}\left(\|\boldsymbol{u}_t^i - \hat{\boldsymbol{u}}_t^i\|_{\boldsymbol{Q}_i^{\theta_{\text{dev}}}} + \|\boldsymbol{u}_t^i\|_{\boldsymbol{R}_i^{\theta_{\boldsymbol{u}}}}\right)}_{\mathcal{J}_i(\boldsymbol{u}_t^i, \hat{\boldsymbol{u}}_t^i; \boldsymbol{\Theta}^i)} + \mathcal{E}_t P \\ \text{s.t.} \quad & (2), (12), (13), (14) \\ & \boldsymbol{x}_t^i \in \mathcal{X}, \ \boldsymbol{u}_t^i \in \mathcal{U} \qquad \forall i \in \{e, \mathcal{H}\} \end{aligned} \quad (15)$$

where the objective function is the expected value of the cost $\mathcal{J}(\boldsymbol{X}_t, \boldsymbol{U}_t; \boldsymbol{\Theta})$, accounting for the uncertainty in $\boldsymbol{\Theta}$, and $\mathcal{E}_t$ contains all slack variables in the CLF constraints, with slack penalty vector $P$.

## IV. Online Parameter Learning

In the online learning process, we leverage POCP (15) to predict trajectories and update the latent parameters based on the observed interactions. Given a dataset $\mathcal{D}$ of previous interactions, we learn $\boldsymbol{\Theta}$ by solving

$$\arg\max_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\Theta}; \mathcal{D}), \quad (16)$$

where $\mathcal{L}(\boldsymbol{\Theta}; \mathcal{D})$ is the likelihood of $\boldsymbol{\Theta}$ given data in $\mathcal{D}$. We use an EKF to solve (16) and update the latent parameters that encode the preferences on safety, actuation effort, and deviation from the original intents for each HDV in the scene. A neural network models these latent parameters and maps them into the parametric constraints and objectives for each HDV in the scene. At the next step, POCP (15) uses the updated latent parameters and covariances from the EKF, and updates the HDVs prediction.

### A. Observer-Based Parameter Adaptation

We define a possibly nonlinear function $f_\Theta(\boldsymbol{X}_t) : \mathbb{R}^{n_x \times (N+1)} \to \mathbb{R}^{n_p}$ that maps the current scenario state $\boldsymbol{X}_t$ to the set of parameters $\Theta_t \in \mathbb{R}^{n_p}$ that form the objectives and constraints within the POCP (15). In addition, we introduce the function $\hat{\xi}_{t-T|t} = h(f_\Theta(\boldsymbol{X}_t), \boldsymbol{x}_{t-T|t})$. With a little abuse of notation, we denote the right-hand side as $\hat{h}(\boldsymbol{\Theta_t})$. The actual trajectory $\xi_t$ of the surrounding vehicles is

$$\xi_t = \hat{h}(\boldsymbol{\Theta_t}) + \nu_t, \tag{17}$$

where $\nu_t$ is a noise term that accounts for the discrepancy between the predicted and actual trajectories.

The prediction equations for the observer-based parameter adaptation scheme are

$$\Theta_{t+1} = \boldsymbol{\Theta_t} + \Delta\boldsymbol{\Theta_t} \tag{18a}$$

$$\xi_t = \hat{h}(\boldsymbol{\Theta_t}) + \Delta\nu_t \tag{18b}$$

where $\Delta\boldsymbol{\Theta_t}$ and $\Delta\nu_t$ are zero-mean Gaussian random variables with covariances $C_\Theta$ and $C_v$, respectively, such that $\Delta\boldsymbol{\Theta_t} \sim \mathcal{N}(0, C_\Theta)$ and $\Delta\nu_t \sim \mathcal{N}(0, C_v)$. The parameter adaptation law is derived from the posterior distribution $\Pr(\Theta_{k+1} \mid \Theta_{0:t}, y_{0:t}) = \prod_{i=0}^{t} \Pr(\Theta_{i+1} \mid \Theta_i, \xi_i) \Pr(\Theta_0)$, where $\xi_{0:t}$ is the observed trajectory up to time $t$.

If $\hat{h}(\boldsymbol{\Theta_t})$ is differentiable, an EKF can be employed to update the latent parameter vector

$$\Delta\boldsymbol{\Theta_t} = K_t \left(\xi_t - \hat{h}(\boldsymbol{\Theta_t})\right), \tag{19}$$

where $K_t$ denotes the Kalman gain computed by

$$\begin{aligned} K_t &= P_{t|t-1} H_t^T S_t^{-1} \\ S_t &= H_t P_{t|t-1} H_t^T + C_v \\ P_{t|t-1} &= P_{t-1|t-1} + C_\Theta \\ P_{t|t} &= (I - K_t H_t) P_{t|t-1}, \end{aligned} \tag{20}$$

$H_t$ is the Jacobian of $\hat{h}(\boldsymbol{\Theta_t})$ with respect to $\boldsymbol{\Theta_t}$, $P_{t|t-1}$ and $P_{t|t}$ are the prediction and corrected covariance matrices.

To compute the Jacobian $H_t$, we construct the linearization based on the chain rule,

$$H_t = \frac{\delta\hat{h}(\boldsymbol{\Theta_t})}{\delta\boldsymbol{\Theta_t}} = \frac{\delta\hat{h}(\boldsymbol{\Theta_t})}{\delta\xi_t} \frac{\delta\xi_t}{\delta\boldsymbol{\Theta_t}} \tag{21}$$

where $\frac{\delta\xi_t}{\delta\boldsymbol{\Theta_t}}$ is computed by differentiating the KKT conditions of the optimal control problem [15].

### B. Parameter Mapping Functions

The latent parameters represent the HDV's preferences in the POCP $\hat{h}(\boldsymbol{\Theta_t})$. We consider two mapping functions $f_\Theta(\boldsymbol{X}_t)$ from the current scenario state $\boldsymbol{X}_t$ to the latent parameter vector $\boldsymbol{\Theta_t}$.

*Direct mapping:* The simplest mapping is the identity function

$$\boldsymbol{\Theta_t} = f_\Theta(\boldsymbol{X}_t) = \boldsymbol{X}_t. \tag{22}$$

Here, $\boldsymbol{\Theta_t}$ directly corresponds to the cost and safety constraints, offering a straightforward interpretation. However, this method may not capture complex relationships between the system state and the optimal parameters.

*Nonlinear mapping:* To represent more complex relationships, we use a Multilayer Perceptron (MLP)

$$f_\Theta(\boldsymbol{X}_t; \Theta) = \mathbf{W}_L (\cdots \varphi(\mathbf{W}_1 \boldsymbol{X}_t + \mathbf{b}_1)) + \mathbf{b}_L \tag{23}$$

where $L$ is the number layers, $\mathbf{W}_\ell$ and $\mathbf{b}_\ell$ are the weights and biases for each layer $\ell$, and $\varphi(\cdot)$ is the activation function.

The variance propagation differs for the two mappings. Direct mapping uses the variance from the EKF's covariance estimation in (20). For the MLP mapping, we propagate the variance through each MLP layer, using Monte Carlo sampling. Future work will explore variance approximations as suggested in [16].

*Remark 1:* The EKF assumes latent parameter observability, which requires sufficient system excitation. Conservative initial parameters can limit this, affecting observability. Thus, careful selection of initial conditions and system inputs is crucial to balance safety and excitation for EKF estimation.

## V. Case Study

We tested the proposed algorithm in a lane-changing scenario to evaluate its effectiveness in ensuring safety and efficiency during interactions with HDVs. All simulations were conducted in Python using an AMD Ryzen 9 5900X, 3.70 GHz 12-core CPU without a GPU, utilizing JAX and the OSQP solver [17].

### A. Interactive Lane-Changing Problem Description

Consider an ego vehicle $e$ encountering a slower HDV (HDV 3) and two other HDVs in the adjacent lane. The goal for $e$ is to merge safely between HDV 1 and HDV 2, minimizing sudden deceleration that could disrupt traffic. The HDVs aim to maintain their lane and desired velocity.

The scenario includes constraints on lateral space due to lane lines. The road width is $l$ and the center of the slow lane is the $y = 0$ axis. The ego vehicle's initial lateral position at $t_0$ is $y_e(t_0) = 0$. The lateral positions of all vehicles are constrained as $-\frac{l}{2} \le y_t^i \le \frac{3}{2}l$ for all $i \in \{\mathcal{H}, e\}$.

The maneuver starts at $t_0$ and ends at $t_f$ when the ego vehicle has fully transitioned to the target lane. The control input and speed for all vehicles are constrained by $\boldsymbol{u}_{\min}^i \le \boldsymbol{u}_t^i \le$ and $\boldsymbol{u}_{\max}^i$, $v_{\min}^i \le v_t^i \le v_{\max}^i$, where $\boldsymbol{u}_{\min}^i$ and $\boldsymbol{u}_{\max}^i$ are the control bounds, and $v_{\min}^i, v_{\max}^i$ are the speed limits.

Each HDV aims at maintaining its lane, $y_t^i \in \{y_{\text{left}}, y_{\text{right}}\}$, and desired speed $v_{\text{des}}^i$, hence using as CLFs

$$\mathcal{V}_{v_{\text{des}}}^i = (v_t^i - v_{\text{des}}^i)^2, \tag{24a}$$

$$\mathcal{V}_{y_{\text{des}}}^i = (y_t^i - y_{\text{des}}^i)^2. \tag{24b}$$

For the ego vehicle traveling on $y_{\text{right}}$ at $t_0$, the goal is lane-changing, thus $\mathcal{V}_{y_{\text{des}}}^e = (y_t^e - y_{\text{left}})^2$, with the desired speed given by $\mathcal{V}_{v_{\text{des}}}^e = (v_t^e - v_{\text{des}}^e)^2$. This ensures that the lateral position constraint influences the control inputs, adapting to varying conditions.

To address lateral constraints, we define CBFs

$$\begin{aligned} \Psi_{y_{\min}}(\boldsymbol{x}_t^i) &= v_t^i \sin(\psi_t^i) - y_{\min} + p_{y_{\min}}(y_t^i - y_{\min}), \\ \Psi_{y_{\max}}(\boldsymbol{x}_t^i) &= -v_t^i \sin(\psi_t^i) - p_{y_{\max}}(y_{\max} - y_t^i), \end{aligned} \tag{25}$$

where $p_{y_{\min}}$ and $p_{y_{\max}}$ are positive constants. Additionally, for speed constraints we implement

$$\Psi_{v_{\min}}\left(\boldsymbol{x}_t^i\right) = v_t^i - v_{\min}, \quad \Psi_{v_{\max}}\left(\boldsymbol{x}_t^i\right) = v_{\max} - v_t^i. \tag{26}$$

State CBF constraints are not parameterized as they are determined by traffic rules and road topology. Instead, we parameterize the ego vehicle's safety constraints to facilitate tracking. This may affect the accuracy of estimating other vehicles' trajectories but will not compromise safety.

### B. Simulation Performance and Analysis

The speed and acceleration ranges used in the simulations were $v \in [15, 33]$ m/s and $u \in [-7, 3.3]$ m/s. The target speed for all vehicles was set to $v_{\text{des}} = 30$m/s, and the highway had a lane width of $l = 4$m. Each vehicle followed the dynamics in (2) with $L_w = 5$m.

For the collision avoidance CBFs in (5), we used $a = 1$m, $b = 3$ m, and $d_{max} = 5\ m/s^2$. For the CBFs from (26), (25), and the deterministic control Lyapunov functions (CLFs) in (24b), we defined linear class-$\mathcal{K}$ functions with gain $p = 1$. The time step length was $\Delta t = 0.05\,s$. The initial states for each vehicle were $X_e(t_0) = [20\text{m}, 0\text{m}, 0\text{rad}, 25\text{m/s}]$, $X_1(t_0) = [62\text{m}, 4\text{m}, 0\text{rad}, 29\text{m/s}]$, $X_2(t_0) = [19.5\text{m}, 4\text{m}, 0\text{rad}, 25\text{m/s}]$, $X_3(t_0) = [62\text{m}, 4\text{m}, 0\text{rad}, 29\text{m/s}]$.

We defined two baseline scenarios: one without robustness-promoting designs and another with robust safety constraints. For adaptive cases, we consider use of direct mapping (22) and of nonlinear mapping (23). The MLP for nonlinear mapping had two hidden layers of size 32, initialized with $\Theta = 0.01$ and covariance $C_\Theta = 0.1I$. The risk parameter was $\epsilon = 0.25$.

Table I presents performance metrics including total actuation effort, speed disruption, and computation times. Total speed disruption, $\sum_{t=0}^{T}(v_t^i - v_{\text{des}}^i)^2\Delta t$, serves as an indicator of smooth interactions [18]. The adaptive methods generally showed lower total speed disruption for HDV 2 compared to baseline cases. The robust baseline case was overly conservative, avoiding lane changes, while the non-robust baseline resulted in aggressive maneuvers and significant speed disruptions.

For a safety degree of 1, direct mapping exhibited conservative behavior with smoother accelerations and longer maneuver times (14.5s), resulting in the lowest total speed disruption for both the ego vehicle (539.196) and HDV 2 (922.523). The nonlinear mapping case showed more conservative maneuvers (16.3s) but with lower actuation effort (9.0 for the ego). Robustness improved safety in the nonlinear mapping at the cost of slightly higher speed disruptions.

Figures 3 to 6 provide further analysis. State history plots (Figure 3) show that the baseline case without robustness led to significant speed fluctuations for HDV 2, indicating poor adaptation. The robust baseline avoided lane changes due to its conservative nature. Direct and nonlinear mapping cases achieved a balanced approach, with nonlinear mapping showing better actuation effort and adaptation.

The 2D trajectory plot (Figure 4) shows that adaptive cases maintained smoother trajectories and respected safety ellipsoids, whereas the baseline case violated safety constraints, causing abrupt deceleration of HDV 2. RMS error analysis (Figure 5) indicates that nonlinear mapping achieved lower prediction errors, reflecting better adaptation. The parameter learning plot (Figure 6) shows faster convergence in the nonlinear mapping case.

In summary, while including robustness in adaptive methods like nonlinear mapping enhances safety, it introduces a trade-off between speed disruption and maneuverability. These findings emphasize the need to balance safety and efficiency in adaptive control systems for autonomous driving.

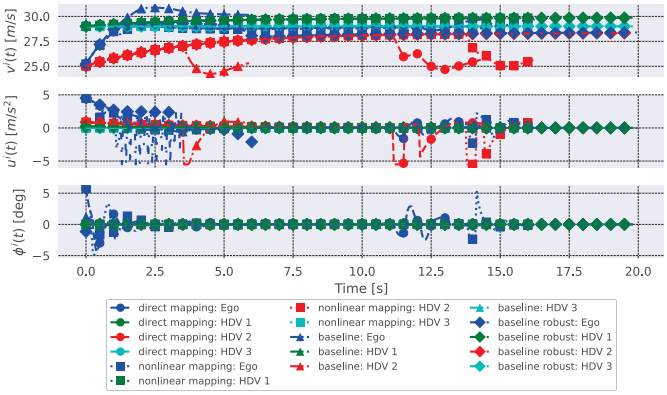


Fig. 3: State history sample of all vehicles for Direct Mapping, Nonlinear Mapping, and Baseline cases with a safety degree of 1. Adaptive approaches result in smoother traffic flow, reducing abrupt speed changes.

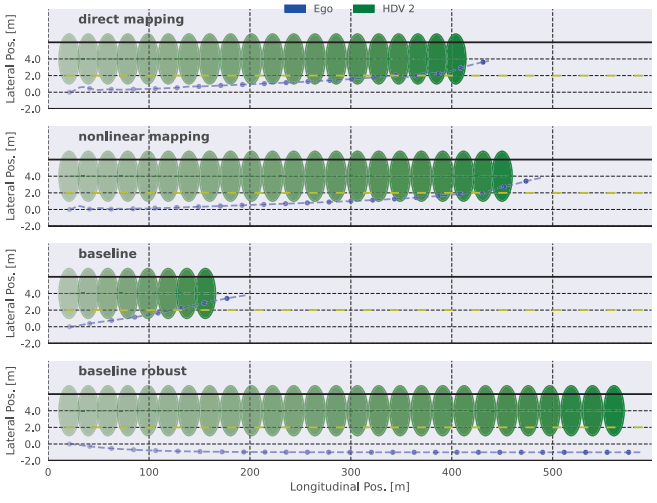


Fig. 4: Sample 2D trajectories of HDV 2 and Ego for Direct Mapping, Nonlinear Mapping, and Baseline cases. Adaptive cases maintain smoother trajectories and respect safety ellipsoids.

## VI. Conclusions and Future Work

We introduced an interactive motion planning framework that combines online learning of human driver preferences

TABLE I: Performance Metrics Comparison Table

| Case | Safety Degree | QP Avg. Computation Time $[s]$ | Belief update Computation Time $[s]$ | Maneuver Time $[s]$ | Total Actuation $\frac{1}{2}\sum_{t=0}^{T}(u_t^e)^2\,\Delta t$ | | Total Speed Disruption $\sum_{t=0}^{T}(v_t^i - v_{des}^i)^2$ | | Avg. Speed Disruption $\frac{1}{2}\sum_{t=0}^{T}(v_t^i - v_{des}^i)^2$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Ego | HDV 2 | Ego | HDV 2 | Ego | HDV 2 |
| direct mapping | 1 | 0.056 | 0.278 | 14.500 | 7.597 | 9.397 | 539.196 | 922.523 | 1.859 | 3.871 |
| nonlinear mapping | 1 | 0.129 | 0.347 | 16.300 | 9.000 | 8.331 | 610.398 | 821.368 | 1.872 | 2.826 |
| baseline | 1 | 0.075 | NA | 5.950 | 9.057 | 6.866 | 228.291 | 1058.341 | 1.918 | 8.894 |
| baseline robust | 1 | 0.077 | NA | 19.950 | 20.854 | 0.734 | 1139.645 | 509.305 | 2.849 | 1.273 |

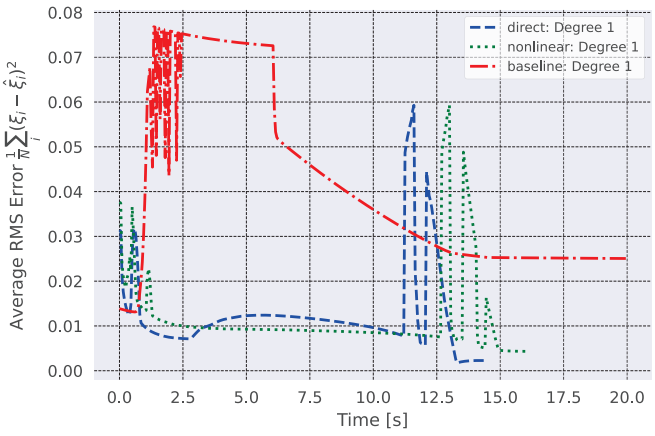


Fig. 5: Average RMS prediction error for all vehicles over time. Lower RMS error in the nonlinear mapping case indicates better adaptation and accurate prediction of vehicle behaviors.

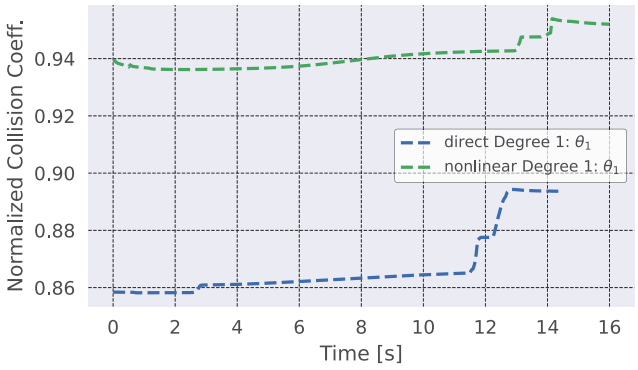


Fig. 6: Collision parameters learned for HDV 2 as a function of time. Nonlinear mapping shows faster adaptation and parameter convergence, indicating more efficient learning.

with parametric optimal control and parametric control barrier functions, robustified by chance constraints for the uncertainty from parameter learning.

Simulations in a lane-changing scenario show the effectiveness of the methods in improving traffic flow and reducing disruptions to human-driven vehicles. Among the proposed parameter mapping, direct mapping results in smoother maneuvers, while nonlinear mapping effectively maintains traffic flow at the expense of more conservative maneuvers, emphasizing the need for adaptability in mixed-traffic settings.

While our framework updates uncertainties online, it uses a predefined basis for safety constraints. Future work will focus on learning the basis from large-scale datasets and extending the prediction horizon to further increase robustness.

## REFERENCES


[1] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions." in *Robotics: Science and systems*, vol. 2, 2016, pp. 1–9.

[2] Y. Chen, S. Veer, P. Karkus, and M. Pavone, "Interactive joint planning for autonomous vehicles," *IEEE Rob. Autom. Lett.*, vol. 9, no. 2, pp. 987–994, 2024.

[3] J. Lubars, H. Gupta, S. Chinchali, L. Li, A. Raja, R. Srikant, and X. Wu, "Combining reinforcement learning with model predictive control for on-ramp merging," in *IEEE Int. Transp. Sys. Conf.*, 2021.

[4] R. Zhang, C. Yu, J. Chen, C. Fan, and S. Gao, "Learning-based motion planning in dynamic environments using gnns and temporal encoding," in *Adv. Neural Inform. Proc. Sys.*, vol. 35, 2022.

[5] R. Chandra, M. Wang, M. Schwager, and D. Manocha, "Game-theoretic planning for autonomous driving among risk-aware human drivers," in *IEEE Int. Conf. Robotics and Autom.*, 2022.

[6] L. Yang, C. Lu, G. Xiong, Y. Xing, and J. Gong, "A hybrid motion planning framework for autonomous driving in mixed traffic flow," *Green Energy and Intell. Transp.*, vol. 1, no. 3, 2022.

[7] K. Leung, S. Veer, E. Schmerling, and M. Pavone, "Learning autonomous vehicle safety concepts from demonstrations," in *American Control Conf.*, 2023.

[8] W. Xiao, T.-H. Wang, R. Hasani, M. Chahine, A. Amini, X. Li, and D. Rus, "BarrierNet: Differentiable control barrier functions for learning of safe robot control," *IEEE Trans. Robotics*, vol. 39, 2023.

[9] M. Menner, K. Berntorp, and S. Di Cairano, "Automated controller calibration by kalman filtering," *IEEE Trans. Control Systems Tech.*, vol. 31, no. 6, pp. 2350–2364, 2023.

[10] A. D. Ames, K. Galloway, and J. W. Grizzle, "Control lyapunov functions and hybrid zero dynamics," in *51st IEEE Conf. Decision and Control*, 2012.

[11] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Prentice Hall, 2002.

[12] S. He, J. Zeng, B. Zhang, and K. Sreenath, "Rule-based safety-critical control design using control barrier functions with application to autonomous lane change," in *American Control Conf.*, 2021.

[13] Y. Yuan, X. Weng, Y. Ou, and K. Kitani, "Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting," in *IEEE Int. Conf. Computer Vision*, 2021.

[14] N. Nayakanti, R. Al-Rfou, A. Zhou, K. Goel, K. S. Refaat, and B. Sapp, "Wayformer: Motion forecasting via simple & efficient attention networks," in *Int. Conf. Robotics and Autom.*, 2023.

[15] B. Amos and J. Z. Kolter, "Optnet: differentiable optimization as a layer in neural networks," in *34th Int. Conf. Machine Learning*, 2017.

[16] P. Wagner, X. Wu, and M. F. Huber, "Kalman bayesian neural networks for closed-form online learning," in *Proc. AAAI Conf. Artificial Intell.*, vol. 37, 2023.

[17] M. Blondel, Q. Berthet, M. Cuturi, R. Frostig, S. Hoyer, F. Llinares-López, F. Pedregosa, and J.-P. Vert, "Efficient and modular implicit differentiation," *arXiv preprint arXiv:2105.15183*, 2021.

[18] A. S. C. Armijos, A. Li, and C. G. Cassandras, "Maximizing safety and efficiency for cooperative lane-changing: A minimally disruptive approach," in *26th IEEE Int. Conf. Intell. Transp. Sys.*, 2023.