

Speed Control in the Presence of Road Obstacles: A Comparison of Model Predictive Control and Reinforcement Learning

P. Mandl^{1(\boxtimes)}, F. Jaumann^{1,2}, M. Unterreiner², T. Gräber², F. Klinger¹, J. Edelmann¹, and M. Plöchl¹

¹ TU Wien, Institute of Mechanics and Mechatronics, Vienna, Austria philipp.mandl@tuwien.ac.at
² CARIAD SE, Vehicle Motion & Energy, Wolfsburg, Germany https://www.mec.tuwien.ac.at/vsd, https://cariad.technology/

Abstract. The paper compares two optimal control methods — Reinforcement Learning and Model Predictive Control — for adaptive speed control in the presence of road obstacles to enhance ride comfort. Both methods use a model for training or prediction and a reward or cost function to achieve a desired control objective. Using the same quarter-car model and objective function for both methods, differences in planned speed profiles, optimality of the control objective, and differences in computational time are analysed through simulations over a series of cosine-shaped road bumps.

Keywords: Model Predictive Control \cdot Reinforcement Learning \cdot Longitudinal Control \cdot Ride Comfort

1 Introduction

The automotive industry is moving towards fully self-driving vehicles by automating both lateral and longitudinal driving tasks. To achieve this, vehicles have to respond to road obstacles using preceding road information. Two key methods for the control task are Reinforcement Learning (RL) and Model Predictive Control (MPC). RL has gained significant interest for its ability to learn optimal policies directly from environmental interactions, enabling robust control of complex systems. Although training is computationally expensive, evaluating the trained models is fast. MPC is an established optimal control method that, like RL, uses model information to predict future system behaviour and optimise actions over a defined horizon. While MPC is fast to deploy, its online computational requirements increase significantly with system complexity [1].

This paper presents a comparative study of RL and MPC on a novel control problem. It introduces a speed planner for the coupled problem of vertical and longitudinal dynamics when traversing road obstacles to improve ride comfort, specifically road bumps by controlling the vehicle's longitudinal motion. Improving ride comfort through suspension control using classical control methods [2] and RL [3,4] has been extensively studied in the literature. However, optimising ride comfort via speed planning is an emerging topic [5].

2 Problem Description and Methods

To maximise ride comfort over a given road segment within the preview distance l_{prev} , it is crucial to select the optimal vehicle speed v. This decision takes into account the current vehicle state \boldsymbol{x} , the speed limits v_{max} and lower v_{min} , and the acceleration limits a_{max} and a_{min} . The control architecture is illustrated in Fig. 1a, while the quarter-car model is shown in Fig. 1b.



Fig. 1. Optimal longitudinal motion control using either MPC or RL on the left. Both methods are based on the quarter-car model shown on the right.

2.1 Vehicle Model

RL and MPC share similarities in their approach, utilising the same quarter-car model in Fig. 1b for prediction or training. The governing equations of motion are taken from [6]. The spring force $F_{c,s}$ is modelled by an air suspension model based on [7]. The damper force $F_{k,s}$ is represented by piecewise linear damper characteristics with distinct high and low-speed damping for compression and rebound. Additional end-stops for rebound and compression are included. The tyre load is modelled by a linear spring c_t and damping coefficient k_t . The quarter-car state is $\boldsymbol{x} = [\zeta - z_W, \dot{z}_W, z_W - z_B, \dot{z}_B, v]^T$ with road elevation ζ , wheel travel z_W , sprung-mass travel z_B and vehicle speed $v = \dot{s}$. The nonlinear continuous-time equations are transformed into the space domain, similar to [6].

2.2 Model Predictive Control

The Optimal Control Problem (OCP) for the MPC is formulated as a nonlinear static optimisation with CasADi¹ and solved with IPOPT². The continuous-space dynamics are discretised using an implicit Euler integration scheme. The

¹ web.casadi.org.

² github.com/coin-or/Ipopt.

road is represented by the change in road elevation $\zeta' = \frac{d\zeta}{ds}$ at discrete points along $l_{\text{prev}} = 40 \text{ m}$ with step size $\Delta s = 5 \text{ cm}$. The OCP is expressed as the following multishooting problem:

$$\min_{\boldsymbol{X},a} \sum_{k=1}^{N} \underbrace{Q_{\ddot{z}_{B}}\left(\frac{\ddot{z}_{B,k}}{g}\right)^{2} + Q_{\ddot{z}_{W}}\left(\frac{\ddot{z}_{W,k}}{g}\right)^{2}}_{J_{\text{heave},k}} + \underbrace{Q_{a}\left(\frac{a_{k}}{g}\right)^{2}}_{J_{\text{long},k}} + \underbrace{Q_{v}\left|\frac{v_{k} - v_{\text{ref}}}{v_{\text{ref}}}\right|}_{J_{\text{speed},k}} \quad (1a)$$

s.t.
$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{f}_{\text{quarter-car}} \left(\boldsymbol{x}_{k+1}, a_k, \zeta'_k \right), \quad \boldsymbol{x}_1 = \boldsymbol{x}(t),$$
 (1b)

$$v_{\min} \le v_k \le v_{\max}, \quad a_{\min} \le a_k \le a_{\max},$$
 (1c)

where $k \in \{1, 2, ..., N\}$ with $N = \frac{l_{\text{prev}}}{\Delta s}$. $J_{\text{heave},k}$ compromises ride comfort through $\ddot{z}_{B,k}$ and dynamic wheel load through $\ddot{z}_{W,k}$. The sprung mass m_B is 567 kg. The unsprung mass m_W is 60 kg. Longitudinal comfort and control input a_k are considered via $J_{\text{long},k}$. Reference speed v_{ref} tracking is managed by $J_{\text{speed},k}$. By suitably weighing these criteria through $Q_v = 1$, $Q_a = 1$, $Q_{\ddot{z}_B} = 50$ and $Q_{\ddot{z}_W} = 0.5$, the ride comfort is improved while maintaining swift passage of the obstacle. g is the gravitational acceleration.

2.3 Reinforcement Learning

Assuming a Markov decision process (MDP) that, starting from an initial state x_0 , forms a trajectory τ of states, actions and rewards. RL aims for the optimal control policy $\pi^*(a|\mathbf{x})$ that solves the optimization problem

$$\pi^* = \arg\max_{\pi_{\theta}} \mathop{\mathbb{E}}_{\tau \sim \pi} \left[-\sum_{k=0}^{\infty} \gamma^k R_k(\boldsymbol{o}_k) \right]$$
(2)

with discount factor $\gamma \in [0, 1)$, step reward R_k and observations o_k .

Observation Space and Action Space. The list of observations visible to the agent comprises the necessary information to learn an optimal policy and pose a subset of the vehicle state and road. The observation space o_k is defined by

$$\boldsymbol{o}_{k} = \left[v_{k}, a_{k}, \ddot{z}_{B,k}, \ddot{z}_{W,k}, z_{W,k} - z_{B,k}, d_{\mathrm{sb},k}, h_{\mathrm{sb},k}, l_{\mathrm{sb},k}, v_{\mathrm{ref},k} \right]^{T}.$$
 (3)

While v_k , a_k describe the longitudinal motion of the vehicle, the vertical movement is observed by $\ddot{z}_{B,k}$, $\ddot{z}_{W,k}$ and $z_{W,k} - z_{B,k}$. The agent sees the upcoming road obstacle via the longitudinal distance between the current vehicle position and the peak position of the obstacle $d_{sb,k}$, the obstacle's maximum height $h_{sb,k}$, and the obstacle length $l_{sb,k}$. With v_{ref} , the agent is aware of the current reference speed. The agent controls the longitudinal motion of the vehicle by setting a_k . The choice of the interval of possible acceleration values is motivated by the system limitations of a real-world adaptive cruise control system.

Reward Function. The reward function encourages or punishes an agent's behaviour by defining favourable environment states. To ensure comparability, the step reward function R_{step} is based on (1a):

$$R_k = J_{\text{heave},k} + J_{\text{long},k} + J_{\text{speed},k} + J_{v_{\min},k} + J_{\text{step},k}.$$
(4)

To enforce the lower speed limit, an additional speed cost $J_{v_{\min},k}$ is added when the agent drops below v_{\min} . For numerical reasons, a step reward $J_{\text{step},k} = -0.05$ is added to encourage progress along the road. Additionally, to penalise premature termination of an episode, such as when the vehicle speed drops below 1 km/h, a large cost of $J_{\text{termination}} = 5000$ is added.

Training and Network Architecture. The RL agent is trained using Stable Baselines3's implementation of the Proximal Policy Optimization (PPO) algorithm³. It utilises a multilayer perceptron (MLP) with two hidden layers of 128 neurons each and is optimised with the Adam optimiser using a learning rate of 3×10^{-4} and a discount factor of 0.999. Each training episode begins with a randomly initialised road, with all training roads having a length of 100 m. The obstacle's dimensions and position vary for each road, with the obstacle height $h_{\rm sb}$ and length $l_{\rm sb}$ ranging between $[0.03 \,\mathrm{m}, 0.08 \,\mathrm{m}]$ and $[0.65 \,\mathrm{m}, 2 \,\mathrm{m}]$, respectively. The obstacle is positioned between $[40 \,\mathrm{m}, 80 \,\mathrm{m}]$. The vehicle's initial v_0 and reference speeds $v_{\rm ref}$ are set between $[10 \,\mathrm{km/h}, 50 \,\mathrm{km/h}]$ and $[25 \,\mathrm{km/h}, 50 \,\mathrm{km/h}]$, respectively. During training, all values are sampled from a uniform distribution within specified bounds. To ensure robust training, there is a ten percent chance that no obstacle will be present, which enforces the training of reference speed tracking. Each training episode consists of 10,000 steps. The policy is evaluated based on a predefined set of roads and velocities.

3 Comparison Between MPC and RL

Both approaches are compared by simulation when crossing over three consecutive cosine-shaped bumps of varying heights and lengths. The first bump is at 50 m with a length of 1 m and height of 5 cm, the second bump at 90 m with a length of 0.75 m and height of 3.5 cm, and the third bump at 100 m with a length of 0.65 m and height of 7.5 cm. The preview distance l_{prev} for both methods is 40 m. The admissible speed range is 5 to 50 km/h, with a reference speed of v_{ref} of 50 km/h. The longitudinal acceleration limits are $a_{\text{max}} = 2.5 \text{ m/s}^2$ and $a_{\text{min}} = -3.7 \text{ m/s}^2$. Note, that this scenario exceeds the training dataset of the RL agent.

Planned Speed and Acceleration Profile. Figure 2 illustrates the planned speed and acceleration profiles for both MPC and RL. RL is represented in red, while MPC in blue. For the first bump, the MPC reduces the speed to approximately 20 km/h, whereas the RL slows down to about 6 km/h. As bump heights increase and length decreases, the MPC approaches the lower speed limit as well. Acceleration profiles show the MPC with a linear increase in braking and acceleration, while the RL prefers constant braking and acceleration. The MPC utilizes the entire available acceleration band, while the RL only uses the maximum acceleration. Reference speed tracking is achieved at the start and end for both methods.

³ stable-baselines3.readthedocs.io.



Fig. 2. Speed profiles v for MPC and RL running over consecutive cosine-shaped bumps with road elevation ζ on the left. Longitudinal acceleration a on the right.

Optimality. Figure 3 provides a detailed breakdown of the planned speed profiles' differences. The top row displays J_{heave} for each bump, followed by the speed cost J_{speed} in the second row, and the longitudinal cost J_{long} in the last row. When observing J_{heave} for the three bumps in the top row, it becomes evident that the RL approach enhances the ride comfort criterion more notably on the first and second bump due to its lower transition speed than the MPC. This improvement comes with the drawback of ocurring larger costs in J_{speed} . Overall, the total cost is primarily influenced by J_{speed} . While the RL approach significantly outperforms the MPC w.r.t J_{heave} , its cummulative cost or optimality w.r.t. the cost function is worse, with a score of 7045 for the RL approach compared to 5762 for the MPC.



Fig. 3. Cost terms for the simulation running over three consecutive cosine-shape bumps. Total accumulated cost: $J_{\text{MPC}} = 5762$, $J_{\text{RL}} = 7045$.

Computational Demand. The calculations were performed on consumergrade laptops, with several runs averaged. The average computation time for the MPC was around 380 ms, with peaks of 1700 ms, compared to an average time of 0.15 ms with peaks of less than 1 ms for the RL approach.

4 Summary and Outlook

This study compared RL and MPC in speed control to improve ride comfort when crossing road obstacles. Both methods utilised the same quarter-car model and cost function for control decisions. While RL learnt optimal policies directly from interactions, MPC used model-based predictions to optimise upcoming behaviour. Through simulations of running over cosine-shape road bumps, the study compared their performance in planned speed profiles, optimality, and computational efficiency. Results showed that the RL outperformed the MPC regarding improved ride comfort, albeit with increased speed costs, resulting in a less optimal solution. The computational demands varied significantly, raising concerns about MPC's suitability for vehicle application in this case. RL demonstrated potential in chassis control application, particularly in planning tasks, but further exploration is needed. Future research should focus on optimising hyperparameters and exploring alternative learning algorithms. The road embedding method used in this study should be extended to a more generic approach. For MPC, computational efficiency can be enhanced by adopting a different road embedding method and employing variable space discretisation to reduce the number of free variables in the OCP.

References

- D. P. Bertsekas. Reinforcement learning and optimal control. Belmont, Massachusetts: Athena Scientific, 2019. isbn: 978-1-886529-39-7
- 2. Tseng, H.E., Hrovat, D.: State of the art survey: active and semi-active suspension control. In: Vehicle System Dynamics (2015)
- 3. Dridi, I., Hamza, A., Ben Yahia, N.: A new approach to controlling an active suspension system based on reinforcement learning. Adv. Mech. Eng. (2023)
- 4. Ming, L. et al.: Semi-active suspension control based on deep reinforcement learning. In: IEEE Access (2020)
- 5. Wu, J. et al.: Ride comfort optimization via speed planning and preview semi-active suspension control for autonomous vehicles on uneven roads. In: IEEE Trans. Vehic. Technol. (2020)
- Mandl, P.: Predictive driver model for speed control in the presence of road obstacles. MA thesis. Vienna, Austria: TU Wien (2021)
- Savaresi, S.M. et al.: CHAPTER 2 Semi-active suspension technologies and models. In: semi-active suspension control design for vehicles. Ed. by S. M. Savaresi et al. Boston: Butterworth-Heinemann, isbn: 978- 0-08-096678-6 (2010)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

