Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 56-2 (2023) 1101–1106

Integration of Robust Control with Reinforcement Learning for Safe Autonomous Vehicle Motion ⋆

Attila Lelkó, Balázs Németh, Dániel Fényes, Péter Gáspár

Institute for Computer Science and Control (SZTAKI), Eötvös Loránd Research Network (ELKH), 13-17. Kende u., H-1111 Budapest, Hungary.

Department of Control for Transportation and Vehicle Systems, Budapest University of Technology and Economics 2 Stoczek u., H-1111 Budapest, Hungary

E-mail: [attila.lelko;balazs.nemeth;peter.gaspar]@sztaki.hu

Abstract: This paper presents a control design framework for the integration of robust control and reinforcement learning-based (RL) control agent. The proposed integration method is applied for motion control of autonomous road vehicles, providing safe motion. In the integrated motion control, longitudinal and lateral dynamics are incorporated. The high-performance motion of the vehicle, e.g., high-velocity motion, path following, and reduction of lateral acceleration, through the RL-based control agent is achieved. The training through Proximal Policy Optimization during episodes is performed. Safe motion with guaranteed performances, i.e., keeping limits on lateral error, through the robust control and the supervisor is achieved. The robust control is designed through the $\mathcal{H}_\infty$ method, and in the supervisor, a constrained quadratic programming task is performed. As a result, lateral and longitudinal control inputs of the vehicle are calculated by the integrated control system. The effectiveness of the proposed control method using simulation scenarios and test scenarios on a small-scaled test vehicle is illustrated.

Keywords: Adaptive and robust control of automotive systems, Autonomous vehicles,

1. INTRODUCTION AND MOTIVATION

Nowadays, due to the appearance of fast hardware tools for solving learning problems, data-based methods are becoming more popular and efficient in the solution of complex control problems. One of the typical examples is autonomous vehicle control, in which sensing, perception, decision, and control problems must be solved in continuously varying traffic environments. Various performance specifications can be defined concerning autonomous vehicle control systems. Usually, there are primary performance specifications, which due to safety reasons must be kept, such as guaranteeing stable vehicle motion, and reliability, or keeping different traffic regulations. Moreover, several further non-safety performance requirements can be defined, which have lower priority, e.g., providing passenger comfort, achieving economic motion, minimization of traveling times, etc. Lots of performance criteria together with the complex vehicle environment lead to challenging problems for robust and optimal control design methods.

The limitations of classical and modern learning-based have led to developing integrated methods, in which classical model-based control techniques and learning-based approaches are incorporated simultaneously. The integration aims to combine these two solutions to achieve the high performance of learning-based methods, and also the robustness and reliability of classical techniques (Németh and Gáspár (2021)). The integration on various levels of autonomous vehicle control can be achieved. Advanced vehicle modeling frameworks have been developed, which involve data processing on the step of model formulation, e.g., through closed-loop matching (Pedro et al. (2018); Hegedűs et al. (2021)). Focusing on the step of control design, it has been provided design frameworks, in which classical and learning-based control solutions jointly have been involved, e.g., robust (Varga et al. (2022)) and LPV control with neural networks (Németh and Gáspár (2021)), or Bao and Velni (2022) has proposed a safe model-based reinforcement learning approach to control LPV systems. Furthermore, data can be incorporated in the coordinated control design of multiple unmanned vehicles, i.e., on the step of coordination Kumaravel et al. (2022); Németh et al. (2022). The benefits of data-aided control on this level are reduced energy consumption or transportation network load Sun et al. (2020); Han et al. (2022).

The goal of this paper is to present an integrated vehicle control strategy with which longitudinal and lateral controls are designed. For achieving high-performance motion of the vehicle, reinforcement learning (RL) is used. For guaranteeing safe motion, robust control based on the $\mathcal{H}_\infty$ method and a supervisor based on quadratic programming are used. As a result, both the lateral and the longitudinal control inputs of the vehicle are calculated by the control

⋆ The research was supported by the European Union within the framework of the National Laboratory for Autonomous Systems (RRF-2.3.1-21-2022-00002). The paper was partially funded by the National Research, Development and Innovation Office (NKFIH) under OTKA Grant Agreement No. K 135512.

10.1016/j.ifacol.2023.10.1711

system. Although some preliminary results in the paper (Németh (2021)) can be found, three main new contributions have been achieved. First, in this paper the training of learning-based and design of model-based controllers are independent. Second, in this paper two dynamics, such as longitudinal and lateral in the control system are handled. Third, the effectiveness of the integrated control through simulations and demonstrations on small-scaled indoor test vehicles is presented.

The paper is organized as follows. The concept of the integrated control, together with the applied control-oriented vehicle model is found in Section 2. The reinforcement-learning-based control design is found in Section 3, and Section 4 presents details on robust control design. Operation principles of the supervisor can be found in 5. The effectiveness of the proposed control system through simulations and demonstrations in Section 6 is presented. Finally, the work is summarized and the future challenges are proposed in 7.

2. OVERVIEW ON THE CONCEPT OF INTEGRATED CONTROL DESIGN

In this paper, the design of the control system is based on the following concept of integration. The structure of the control loop with its control and supervisor elements can be seen in Figure 1. During the control design, the RL-based control agent and the robust controller are designed independently, and both control elements use measured signals on the system independently and provide candidate control inputs for the supervisor. The supervisor calculates the control input based on the candidate control inputs (Németh and Gáspár (2021)).

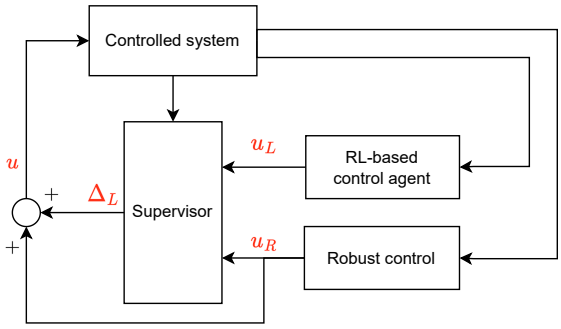


Fig. 1. Scheme of the integrated control concept

The relationships behind the concept in the case of a single-input system are shown below. Nevertheless, in the application of the method for the autonomous vehicle control problem, it is extended for multiple-input cases. The control input u for the system, i.e., the output of the supervisor, is formed as follows

$$u = u_R + \Delta_L, \tag{1}$$

where u_R is the output of the robust controller. Δ_L is a peak-bounded signal, which is computed by the supervisor:

$$\Delta_L \in [\Delta_{L,min}; \Delta_{L,max}], \tag{2}$$

where $\Delta_{L,min}, \Delta_{L,max}$ denote bounds on Δ_L.

Behind the rule of selection Δ_L, the idea of integration is found. It is considered that the output of the RL-based control agent, such as u_L, can provide high-performance operation for the vehicle. Nevertheless, the training of the RL-based agent is not necessarily able to result in guarantees on the minimum performance level. Although the minimization of $u - u_L$ can be beneficial, i.e., $\Delta_L = u_L - u_R$, the resulting u_L signal from the viewpoint of safety performance requirements must be evaluated. This evaluation leads to a constrained optimization problem, whose objective is the minimization of $\|u(\Delta_L) - u_L\|_2^2$, and the constraints through containment condition (2) and further vehicle-safety-oriented conditions are formed. These latter conditions can require additional measurements on the system for the supervisor, see Figure 1. In this integrated control concept, the robust controller works as a baseline control: if the candidate u_L is close to u_R, then it is considered safe to use as u. But, if the difference is larger than a predefined value (i.e., out of the bounds), the supervisor limits the deviation by choosing an appropriate Δ_L value.

The proposed integrated control framework has two main beneficial contributions to the operation of the system. First, the supervisor through the selection of Δ_L provides guarantees on the minimum performance level on safety performances, which is equal to the performance level of the robust controller. Nevertheless, an improved maximum level of performance can be achieved through the operation of the RL-based control agent, when candidate u_L is acceptable. Second, limitations on the learning process (e.g., model structure, measurements, etc.) for achieving an RL-based control agent are not posed, it is only through its output is evaluated. Moreover, in the robust control design, only the bounds of Δ_L are involved, as requested preliminary selections. Further details on the theoretical background can be found in (Németh and Gáspár (2021)), and the application of the method in the given autonomous vehicle control problem in the rest of the paper is detailed.

Vehicle model for integrated control purposes

The design of motion control for autonomous vehicles requests the formulation of their dynamic models. The model has high importance in integrated vehicle control since it is used in the RL training, and the robust control design and it can be built in the supervisor to form vehicle-safety-oriented conditions in the constrained optimization problem. Despite its importance, a complex nonlinear formulation of the vehicle model is not useful, because it can lead to slow training process, numerical difficulties in the robust control design, and finally, slow real-time running performance in the solution of the supervisory optimization process. Therefore, it is recommended to form a simplified dynamical two-wheel (bicycle) motion model, such as:

$$m\dot{v}_x = F_{drive} - bv_x + mv_y\dot{\psi}, \tag{3a}$$

$$m\dot{v}_y = -C(\alpha_F + \alpha_R) - mv_x\dot{\psi}, \tag{3b}$$

$$\theta_z\ddot{\psi} = -C(\alpha_F + \alpha_R)\frac{L}{2}, \tag{3c}$$

where F_{drive} is the driving force, b is a coefficient of friction in the longitudinal velocity, C is the cornering stiffness, α_F and α_R are tire side-slips at the front and rear tires respectively, and L is the length of the wheelbase. Tire side-slip angles and v_y lateral velocity can be expressed

as the functions of yaw rate $\dot{\psi}$, steering angle δ and v_x (Németh (2021)). Relations in (3) can be used within the environment, resulting in local velocities in the frame of vehicle reference and the yaw angle ψ. The global velocities can be calculated as

$$V_x = v_x \cos\psi - v_y \sin\psi, \tag{4a}$$
$$V_y = v_x \sin\psi + v_y \cos\psi, \tag{4b}$$

which can be used for computing the position of the vehicle through the integration of V_x, V_y.

3. DESIGN OF RL-BASED AGENT FOR AUTONOMOUS VEHICLE CONTROL

In this section, the design method of the learning-based control controller is detailed for handling longitudinal and lateral dynamics. The lateral and longitudinal controllers are designed in the form of two separate neural networks to increase the modularity of the system and to achieve faster convergence during training, but the networks are trained in an iterative method.

In the design process of an RL-based agent Proximal Policy Optimization is used (Schulman et al. (2017)). The reason behind its selection is its fast training capability, compared to Trust Region or Policy Gradient methods. The aim of this method is that for training purposes it uses a clipped surrogate objective function, which limits the variation of actions between two steps, i.e., a penalty for having too large of a policy update is applied. The training process is performed through simulation episodes, in which the vehicle must move on given tracks. The tracks are generated through linear and arc segment primitives with a predetermined width. During training, if the vehicle left the track the current scenario is interrupted and a large punishment through reward functions is applied. The output of the lateral agent is the steering angle, in the case of the longitudinal agent, its output is the reference longitudinal velocity.

The improvement of the performance level by the RL-based agents can be achieved through reward functions. At every step of the environment, the reward is calculated, based on the unique reward functions of the agents. In the case of the lateral agent, the parametric reward function is formed as

$$R_{Lat}(s,a) = -Ax_{\text{Lat,err}}^2 - B\psi_{\text{err}}^2 - D\Delta_\delta^4 + f(s,a), \tag{5}$$

where $x_{\text{Lat,err}}$ is the lateral path tracking error, ψ_{err} is the orientation error, Δ_δ is the difference between the actual and previous step steering angles and it is used to limit the rate of change of the control input, decreasing unwanted oscillations in the resulting control, A, B and D are the corresponding weights and

$$f(s,a) = \begin{cases} 1 & \text{if the next checkpoint is closer than the last} \\ 0 & \text{otherwise.} \end{cases}$$

The checkpoints are designated points on the centerlines of the track, and the progression along the track, and the lateral and orientations error are estimated based on these points. The driving behavior of the agents is greatly influenced by the weights of reward functions. The most typical example is if one chooses weights A and B large, then the result will be an agent that follows the center of the track accurately. But, if these weights are small compared to $f(s,a)$, then faster progress will be more important and the agent will tend to cut corners aggressively to reduce lap time. The result of a high D value is the reduction of steering angle oscillation.

In the case of the longitudinal control agent, the reward function for the training process is

$$R_{Long}(s,a) = -Ex_{\text{Lat,err}}^2 - F\psi_{\text{err}}^2 - Ga_y^{10} + f(s,a), \tag{6}$$

where a_y is used to punish the lateral acceleration to limit the traction force needed and to increase passenger comfort, it is at the 10^{th} power resulting in low punishments for small acceleration which increases rapidly beyond a value determined by the corresponding weight. E, F, and G are design parameters to achieve a balance between the different terms in the longitudinal reward function. Thus, the selection of $R_{Long}(s,a)$ expresses that longitudinal control has a high impact on the lateral motion of the vehicle, e.g., through the appropriate selection of velocity profile a_y can be limited.

The inputs of the networks consist of measurements on the track in the neighborhood of the vehicle, which is the relative position of N number of checkpoints ahead of the vehicle, and also the actual difference in vehicle heading and track orientation:

$$x_{NN} = [X_{r,1}\ Y_{r,1}\ X_{r,2}\ Y_{r,2}\ \dots\ X_{r,N}\ Y_{r,N}\ \Delta\psi]\,. \tag{7}$$

This input vector is used for both control agents. Although the selection of N for a long horizon can result in high-efficient efficient control intervention due to the lots of information on the track, it can overcomplicate the neural network and consequently, training time is significantly increased. Therefore, N is recommended to select depending on the characteristics of the track, i.e., the requested minimum look-ahead distance, which determined the actual selection of δ and v_x.

The training of the neural networks is performed iteratively through the following steps

(1) The lateral agent is initialized and trained in an environment, where the reference longitudinal velocity is set to constant until the agent is able to navigate without leaving the track.
(2) The longitudinal agent is initialized and trained in an environment, where the lateral control is performed by the previously trained longitudinal agent, until $R_{Long}(s,a)$ is converged.
(3) The lateral agent is trained while the longitudinal control is performed by the longitudinal agent, and then, the longitudinal agent is trained with the help of the lateral agent until $R_{Lat}(s,a)$ is converged.
(4) Step (3) repeats until the increase of the expected rewards in the episodes is larger than a threshold (ΔR_{min}) in the training scenarios. If there is no significant increase in the rewards the training stops and the agents are considered to be trained.

4. DESIGN OF THE ROBUST ELEMENT OF THE AUTONOMOUS VEHICLE CONTROL SYSTEM

The design of the robust control is based on the method presented in this section. In the control design, it is necessary to consider that u may differ from u_R, due to the additional value of Δ_L. Therefore, Δ_L can be handled as

an additive input disturbance to u_R, and thus, robustness against Δ_L must be guaranteed. The block diagram of the system in Figure 1 can be restructured to form a simple control loop with additive input disturbance, see Figure 2. Thus, from the viewpoint of the robust controller, the supervisor and the reinforcement learning-based control agent are not considered, only the Δ_L.

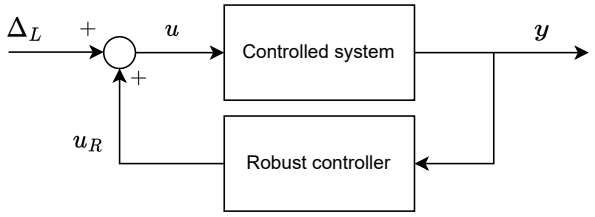


Fig. 2. Schematic view of the loop for robust control design

In the robust control design process below, the worst-case scenario is considered, i.e., Δ_L in the robust design process through its bound is involved, see (2). In the design is assumed that $L_\Delta \triangleq |\Delta_{L,max}| = |\Delta_{L,min}|$, i.e., the measure of additive input disturbance is symmetric. Since L_Δ has an impact on the robust control design, it influences the values of $\Delta_{L,max}, \Delta_{L,min}$ in the optimization process of the supervisor. Choosing L_Δ to have a large interval allows increased differences in the two control signals, resulting in $\Delta_L = u_L - u_R$ often can be selected. Nevertheless, it results in a more conservative robust controller to provide robust stability even in case of larger disturbances. But, if L_Δ is tight, u is close to u_R, and thus, performance level increase from the RL-based control agent can be lost.

Computation of safe steering angle

In designing a robust control system, several methods are available (e.g., $\mathcal{H}_\infty$ or robust LPV methods), with which theoretical guarantees against bounded disturbances can be guaranteed. In the rest of this section a robust $\mathcal{H}_\infty$ control design method on the lateral dynamics, considering Δ_L is proposed. The design is based on the vehicle motion model (3)(b),(c), which can be transformed into the following state-space representation:

$$\dot{x} = Ax + B_2 u = Ax + B_2\Delta_L + B_2 u_K, \tag{8}$$

where A, B_2 are matrices of the system, $x = \begin{bmatrix}\dot{y} & \dot{\psi}\end{bmatrix}^T$ represents state vector and $u = \delta$ steering angle.

The primary, i.e., the safety performance of the system is to guarantee the limitation of the lateral error of the vehicle from the centerline of the road:

$$z_1 = y_{ref} - y; \quad |z_1| \to min, \tag{9}$$

where y_{ref} is the reference lateral position for the vehicle. Moreover, the limitation of the steering angle is requested to avoid the unwanted effect of actuator saturation, which leads to further performance:

$$z_2 = u = u_K + \Delta_L; \quad |z_2| \to min. \tag{10}$$

In the design process of the $\mathcal{H}_\infty$ controller weighting function for scaling disturbances and for finding balance between different performances must be applied, see Sename et al. (2013) and Németh and Gáspár (2021) for details on selecting weighting functions and the formulation of the $\mathcal{H}_\infty$ design problem.

Computation of safe velocity profile angle

The computation of the safe velocity profile, i.e, the actual reference velocity is determined by the local curvature of the track. Since the reference path (e.g., centerline) of the track is a known two-dimensional parametric curve, its curvature can be calculated as

$$\kappa(s) = \frac{|\dot{C}_{track}(s) \times \ddot{C}_{track}(s)|}{|\dot{C}_{track}(s)|^3}, \tag{11}$$

where s is the parameter of the curve (e.g., the distance traveled along the centerline). The reference velocity is calculated in a way to limit the required lateral acceleration of the car for traction reasons:

$$v_{ref}^2 \kappa \leq a_{y,max}, \tag{12}$$

where $a_{y,max}$ denotes the maximum achievable lateral acceleration, based on the maximal tire force available. Its value can be the result of tire characteristics or estimation, see e.g., (Villagra et al. (2011)). The resulting reference velocity as a function of s is

$$v_{ref}(s) \leq \sqrt{\frac{a_{y,max}}{\kappa(s)}}. \tag{13}$$

In straight sections, the curvature of the track is 0, which results in infinite reference velocity. Thus, v_{ref} must be limited, especially in case of low curvature values, such as $v_{ref}(s) \leq v_{max}$, where v_{max} is the maximum velocity limit on the given road section.

5. DESIGN OF THE SUPERVISOR FOR THE INTEGRATION OF DIFFERENT CONTROLLERS

The supervisor results in Δ_L signal, which is used for the computation of control input u, see (1). The goal of the supervisor is to achieve a control signal, which results in a high-performance level for the vehicle through the minimization of the following objective:

$$\|u(\Delta_L) - u_L\|_2^2 = \|u_R + \Delta_L - u_L\|_2^2 \to \min, \tag{14}$$

where $\Delta_L \in [\Delta_{L,min}; \Delta_{L,max}]$ is considered (2), and $\Delta_L = \begin{bmatrix}\Delta_{L,Lat} & \Delta_{L,Long}\end{bmatrix}^T$ contains the additional disturbance value respect to lateral and longitudinal controls.

Additionally, in the case of the vehicle path tracking control, the lateral distance from the centerline can be limited using a model-based prediction layer in the supervisor. The trajectory of the vehicle can be predicted (3)-(4), see Németh and Gáspár (2021), which results in the vector of predicted $P_{pred,T}$ vehicle positions. The predicted lateral error can be estimated using the known checkpoints of the track $C_{track}(s)$:

$$e_{Lat,T} = \min_s \|P_{pred,T} - C_{track}(s)\|_2^2. \tag{15}$$

Thus, the constrained optimization task in the supervisor is formed through (14),(2), and (15) as

$$\underset{\Delta_L}{\operatorname{argmin}} \|u_R + \Delta_L - u_L\|_2^2 \tag{16a}$$

subject to

$$\Delta_L \in [\Delta_{L,min}; \Delta_{L,max}] \tag{16b}$$

$$e_{Lat,T} \leq e_{max}. \tag{16c}$$

In case of the infeasibility of (16), e.g., the vehicle is out of the track, $\Delta_L = 0$ is selected, and a command on maximum reduction of velocity for the vehicle is sent.

In most of the operation of the vehicle, it can be considered that the vehicle moves under normal vehicle dynamic conditions, i.e., it is assumed that u_L can result in keeping the lateral error under e_{max}. Consequently, in the first computation step of the supervisory process, the assumption is checked, such as the ensuring of condition $e_{Lat,T} \leq e_{max}$ with u_L. If the assumption in the checking process is validated to be true, an equivalent solution of the optimization process (16) is

$$\Delta_{L,i} = \begin{cases} u_{L,i} - u_{R,i} & \text{if } (u_{L,i} - u_{R,i}) \in [\Delta_{L,min}; \Delta_{L,max}] \\ \Delta_{L,max} & \text{if } (u_{L,i} - u_{R,i}) > \Delta_{L,max} \\ \Delta_{L,min} & \text{if } (u_{L,i} - u_{R,i}) < \Delta_{L,min}, \end{cases} \tag{17}$$

where the indices denote the i^{th} control input. But, if the assumption in the checking process is validated to be false, the optimization process (16) must be performed.

In the case of a practical application, the process above can significantly reduce the computation effort, i.e., the assumption in most of the vehicle operation is verified to be true.

6. ILLUSTRATION OF THE DESIGNED INTEGRATED CONTROL SYSTEM

Finally, in this section, the effectiveness of the integrated control system is illustrated. In the examples, the goal of the integrated control system is to navigate the car on a racetrack. First, the design of the RL-based control agent is detailed, and second, simulations of the operation are shown.

The training of the RL-based control agent based on the motion of the vehicle on various tracks has been performed. The environment using route primitives, e.g., linear or arc segments, has been generated. The neural networks were structured similarly in both the lateral and longitudinal agents. They had 3 hidden layers containing 5, 15, and 25 neurons each and ReLU activation layers. The output layers use hyperbolic tangent activations to appropriately limit the control outputs. The output of the network is fed back to the input leading to a recurrent neural network structure. For the input measurements of the networks, $N = 5$ with 0.5 m equidistant segments is selected, with which information on the track 2.5 m distance ahead of the vehicle is considered. For this type of track $v_{max} = 2$ m/s is selected. Using (5) and (19), the reward functions are selected as

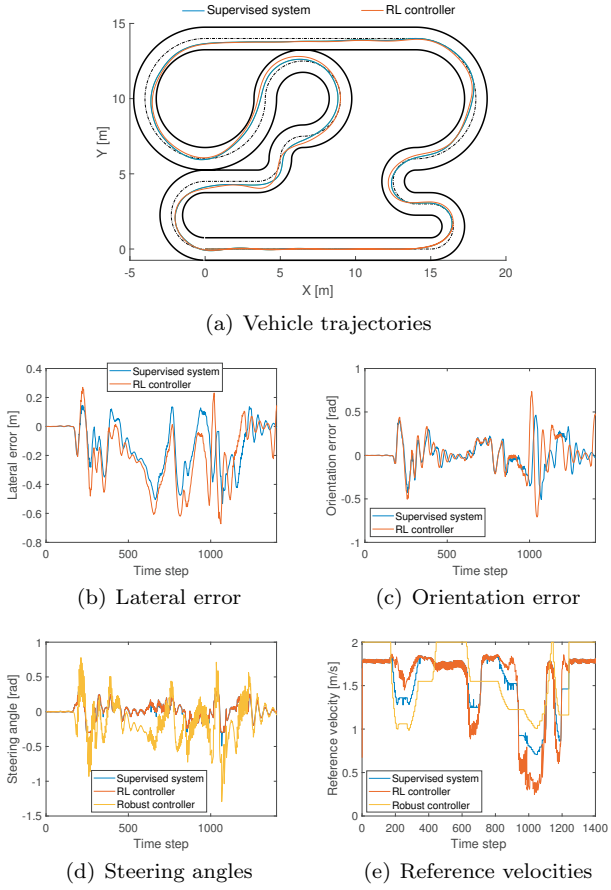
$$R_{Lat} = -0.5x^2_{\text{Lat,err}} - 0.5\psi^2_{\text{err}} - 10\Delta^4_{\delta} + f(s,a), \tag{18}$$

$$R_{Long} = -0.5x^2_{\text{Lat,err}} - 0.5\psi^2_{\text{err}} - \frac{a_y^{10}}{1024} + f(s,a). \tag{19}$$

For performing supervisory optimization, the regions of $\Delta_{L,Lat} = [-0.25, 0.25]$ and $\Delta_{L,Long} = [-0.3, 0.3]$ are selected.

6.1 Simulation results

First, the evaluation of the proposed integrated control method through simulations is performed. A comparison can be seen in Figure 3 where the operation of the integrated control (blue) and of the RL-based controller (orange) are found. Both controllers provide accurate motion (see Figure 3(a)), however, in the case of the integrated control the maximum values of the lateral error are reduced, due to the safety constraint in the optimization. In some cases, the controller without the supervisor tends to cut the corners more sharply resulting in possibly dangerous situations, but in case of the integrated control, the supervisor avoids these situations. In Figure 3(b) the lateral error of the vehicles is shown and it can be seen that in three cases around time steps 650, 800, and 1100 the controller without the supervisor violates the maximal lateral error constraint e_{max}, while with the integrated control the vehicle is within the specified interval at all the time. In Figure 3(d)-3(e) the candidate control interven-



(a) Vehicle trajectories

(b) Lateral error

(c) Orientation error

(d) Steering angles

(e) Reference velocities

Fig. 3. Comparison of vehicle motion signals and control interventions

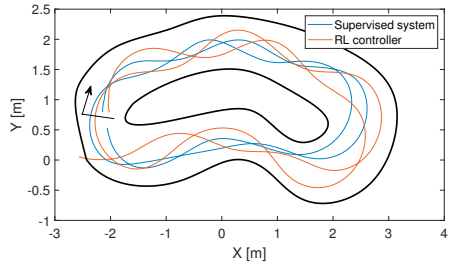
tions of the RL-based control agent, the robust controller, and the integrated control are shown. It can be seen that the signals of the integrated control and the RL-based control agent are close to each other, but at critical situations, e.g., at sharp corners (time steps 200, 650, 800, 1100), δ and v_{ref} are modified. Numerical results on lap time using the two controllers are 59.95 s in case of the RL-based agent and 62.55 s in case of the integrated controller. It shows that lap time with the RL-based controller is smaller than with the integrated controller, which is resulted by the reduction of v_x in the latter case, i.e., 4.2% slower lap time has resulted. Nevertheless, the smaller lap time of the

RL-based controller is not achievable, due to the avoidance of the constraints. Thus, the integrated control results in an appropriate minimum increase in the lap time.

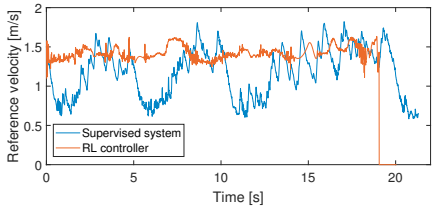
6.2 Results with small-scaled test vehicle implementation

The effectiveness of the control has been investigated in real-life experiments using an F1TENTH small-scaled test vehicle with a LiDAR-based localization algorithm, which has provided estimated information about centerline (Ghallabi et al. (2018)). A track has been set up using buoys, whose path is independent of the tracks of the training set. The motion of the test vehicle has been monitored by the OptiTrack motion capture system, but this data has been used only for evaluation.

A comparison of vehicle trajectories with the integrated and the RL-based controls can be seen in Figure 4(a). It can be seen that without the supervisor the RL-based controller is only able to navigate one full lap and failed on the second by leaving the track. It can be avoided through lateral error prediction of the supervisor, with which the vehicle is able to complete both laps safely.



(a) Measured trajectories of F1TENTH vehicle



(b) Reference velocity signals of the vehicle

Fig. 4. Measurements on the F1TENTH vehicle

The possible cause of leaving path can be found by investigating reference velocities chosen by the two controllers. Figure 4(b) shows that the supervisor consistently decreases the reference velocity in corners, which results in s 9.5% increase in lap time, but in this case, the leaving of the track is avoided.

7. CONCLUSIONS

The paper has proposed an integrated control design framework for motion control of autonomous vehicles. Using simulations and implementation on a small-scaled test vehicle the effectiveness of the control system is illustrated. In the framework, the robust controller and the reinforcement learning-based control agent are designed independently, but a supervisory algorithm guarantees the safe and efficient operation of the closed loop. The illustrations have shown that the control interventions of the RL-based agent, which can result in unsafe motion, can be avoided through the supervisory structure, i.e., a minimum performance level on safety requirements is guaranteed. Nevertheless, the high-performance operation of the RL-based control agent is only slightly reduced.

REFERENCES

Babu, V.S. and Behl, M. (2020). f1tenth.dev - An Open-source ROS based F1/10 Autonomous Racing Simulator. In *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, 1614–1620.

Bao, Y. and Velni, J.M. (2022). Safe control of nonlinear systems in LPV framework using model-based reinforcement learning. *International Journal of Control*, 0(0), 1–12.

Boyd, S., Ghaoui, L.E., Feron, E., and Balakrishnan, V. (1997). *Linear Matrix Inequalities in System and Control Theory*. Society for Industrial and Applied Mathematics, Philadelphia.

Ghallabi, F., Nashashibi, F., El-Haj-Shhade, G., and Mittet, M.A. (2018). Lidar-based lane marking detection for vehicle positioning in an hd map. 2209–2214.

Han, K., Li, N., Tseng, E., Filev, D., Kolmanovsky, I., and Girard, A. (2022). Improving autonomous vehicle in-traffic safety using learning-based action governor. *Advanced Control for Applications*, 4(2), e101.

Hegedűs, T., Fényes, D., Németh, B., and Gáspár, P. (2021). Improving sustainable safe transport via automated vehicle control with closed-loop matching. *Sustainability*, 13(20), 11264.

Kumaravel, S.D., Malikopoulos, A.A., and Ayyagari, R. (2022). Optimal coordination of platoons of connected and automated vehicles at signal-free intersections. *IEEE Transactions on Intelligent Vehicles*, 7(2), 186–197.

Németh, B. and Gáspár, P. (2021). *Guaranteed Performances for Learning-Based Control Systems Using Robust Control Theory*, 109–142. Springer International Publishing, Cham.

Németh, B. (2021). Coordination of lateral vehicle control systems using learning-based strategies. *Energies*, 14(5).

Németh, B. and Gáspár, P. (2021). Ensuring performance requirements for semiactive suspension with nonconventional control systems via robust linear parameter varying framework. *International Journal of Robust and Nonlinear Control*, 31(17), 8165–8182.

Németh, B., Lelkó, A., Antal, Z., and Csaba, A. (2022). Collision-free trajectory design for dance choreography of virtual drones in hierarchical structure. *IFAC-PapersOnLine*, 55(6), 502–507. 11th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2022.

Pedro, J.O., Dangor, M., Dahunsi, O.A., and Ali, M.M. (2018). Dynamic neural network-based feedback linearization control of full-car suspensions using PSO. *Applied Soft Computing*, 70, 723–736.

Scherer, C. and Weiland, S. (2000). *Lecture Notes DISC Course on Linear Matrix Inequalities in Control*. Delft University of Technology, Delft, Netherlands.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms.

Sename, O., Gáspár, P., and Bokor, J. (2013). *Robust Control and Linear Parameter Varying Approaches*. Springer Verlag, Berlin.

Sun, C., Guanetti, J., Borrelli, F., and Moura, S.J. (2020). Optimal eco-driving control of connected and autonomous vehicles through signalized intersections. *IEEE Internet of Things Journal*, 7(5), 3759–3773.

Varga, B., Kulcsár, B., and Chehreghani, M.H. (2022). Deep Q-learning: A robust control approach. *International Journal of Robust and Nonlinear Control*, n/a(n/a).

Villagra, J., d'Andréa Novel, B., Fliess, M., and Mounier, H. (2011). A diagnosis-based approach for tire-road forces and maximum friction estimation. *Control Engineering Practice*, 19(2), 174–184.