Design of learning-based control with guarantees for autonomous vehicles in intersections

Balázs Németh *, Péter Gáspár *

* Systems and Control Laboratory, Institute for Computer Science and
Control (SZTAKI), Eötvös Loránd Research Network (ELKH)
Kende u. 13-17, H-1111 Budapest, Hungary.
E-mail: [balazs.nemeth;peter.gaspar]@sztaki.hu

Abstract: This paper proposes a design method for the coordination of velocity profiles of autonomous vehicles in non-signalized intersection scenarios. The coordination is motivated by the avoidance of vehicle collision and the minimization of their energy loss resulted by stop and go maneuvers in the intersection. Therefore, the coordinated design is formed as an optimal control problem, which is solved through two optimization tasks. A quadratic optimization task with online solution is formed, which provides guarantees on the avoidance of the collision. Moreover, a reinforcement-learning-based optimization task with offline solution is formed, which is able to improve the economy performances of the autonomous vehicles. The optimization tasks are interconnected, i.e. the quadratic optimization with the vehicle model is used as an environment during the training process. The effectiveness of the proposed coordinated control through simulation examples with three number of autonomous vehicles is illustrated.

Keywords: automated vehicles, intersection, reinforcement learning, collision avoidance

1. INTRODUCTION AND MOTIVATION

The coordinated design for the motions autonomous vehicles in intersections is a recent hot topic of the vehicle control problems due to their high number of challenges, e.g. sensing, communication and optimal coordination problems. This paper focuses on the coordinated design of the velocity profile of the vehicles, i.e. on their longitudinal control. It poses several control problems, as follows.

- The objective can contain multiple criteria, the most important are the minimization of the traveling time, the energy consumption and the maximization of the comfort performances. The variables of the optimization task are the control inputs of the vehicles, e.g. longitudinal acceleration command or traction/braking forces.
- The dynamics of each vehicles provides constraint in the optimization problem.
- Furthermore, the safe motion of the vehicles, i.e. collision avoidance must be guaranteed, which leads to constraints for the vehicles with intercrossing route.

- The velocity of the vehicles must be kept in a pre-defined range, considering the speed limits of the vehicles.
- The control input of the vehicles must also be limited due to the physical limits of the driveline, braking system and tyre-road contact.

The solution of the optimization problem above have several challenges. Achieving a global optimal solution requires the computation of the control inputs along the entire intersection scenario. Due to the uncertainties in dynamics of the vehicles and the time delays in the control intervention the computation of the control inputs during the vehicle motion must be continuously performed. A solution is to apply Model Predictive Control methods with optimization on finite rolling horizon Kim and Kumar [2014], Riegger et al. [2016], Bichiou and Rakha [2019], Hult et al. [2019]. Although it can provide appropriate results, the increase of vehicle numbers can make the real-time computation difficult. A possible solution on the problem of increasing computation effort is the approximate the optimal solution with neural networks, see e.g. Németh et al. [2018].

Another approach for the solution of the vehicle motion problem is to use learning-based control solutions, especially reinforcement learning-based methods Isele et al. [2018], Wu et al. [2020], Chen et al. [2020], Zhou et al. [2020]. The advantage of it is that some of these methods are model-free, which can provide solution on the problem of constraint formulation. Moreover, the training of the

control agent through high number of episodes is carried out, which can lead to improved performances. In spite of the promising achievements, the resulted neural-network-based agents cannot provide guarantee on the collision avoidance of the vehicles.

In this paper the design of the motion profile for autonomous vehicles in non-signalized intersections is proposed in a novel way. The optimization problem of the vehicle motion is separated, depending on the performances. First, it is proposed a quadratic optimization problem, whose role is to guarantee the collision avoidance and the limitation of the velocities of the vehicles. The constraint of collision avoidance is formed through a linear approximation of the quadratic constraint, which leads to high efficiency in the reduction of the computation requirements. The quadratic optimization in every time step during the motion of the vehicles is solved. Second, the advantages of reinforcement learning in the improvement of the economy performances, e.g. minimization of the control input are exploited. In the training process the previously formed quadratic optimization task is applied as a part of the environment for learning. Similarly, during the operation of the control system the trained neural network and the quadratic optimization task operate together.

Thus, the contributions of the paper is as follows. The control design problem for autonomous vehicles in intersections is formed in a novel way through the separation of the problem. It leads to reduced complexity of the control problem in real-time computation. Moreover, it is created an environment model for reinforcement learning, with which guarantees on the collision avoidance can be provided. Although the proposed method is proposed for n number of vehicles, its effectiveness is illustrated through an example with three vehicles, see Figure 1. In the example $Vehicle\ 1$-$Vehicle\ 2$ and $Vehicle\ 1$-$Vehicle\ 3$ are in conflicts, which means that their collision must be avoided.
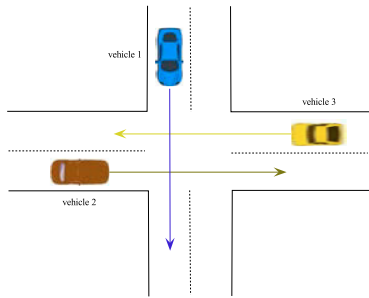

Fig. 1. Example on intersection scenario

The paper is organized as follows. In Section 2 the model formulation for handling intersection scenarios is proposed. The application of reinforcement learning for the improvement of economy performances is presented in Section 3. In Section 4 the effectiveness of the proposed method through simulation examples is illustrated. Finally, the conclusions and the future challenges are provided in Section 5.

2. FORMULATION OF VEHICLE MODELS FOR THE MOTION IN INTERSECTIONS

The goal of this section is to formulate the collision-free motion of the vehicles in intersection scenarios. It is based on the simple longitudinal kinematic model of the vehicles:

$$v_i(k+1) = v_i(k) + Ta_i(k), \tag{1a}$$

$$s_i(k+1) = s_i(k) + Tv_i(k) + \frac{T^2}{2}a_i(k), \tag{1b}$$

where $_i$ index represents the number of the vehicle, n is the number of vehicles, v_i is longitudinal velocity, s_i is longitudinal displacement. a_i represents longitudinal acceleration of the vehicle, which is handled as a control input command and T is time step of the discrete motion model. The longitudinal displacement is related to the center point of the intersection and thus, it is defined as $s_i = 0$ for all i in the center point. The longitudinal displacement of the approaching vehicle has negative value and the displacement of the vehicle moving away has positive value. The control input of the system is separated into two elements, such as

$$a_i(k) = a_{K,i}(k) + \Delta_i(k), \tag{2}$$

where $a_{K,i}$ is the control input command of the robust controller for the ith vehicle and $\Delta_i(k)$ is the additional input from the supervisor in the model.

The goal of the supervisor in the collision-free motion model is to select $\Delta_i(k)$ for all vehicles. The aim of the selection is to minimize the difference between $a_i(k)$ and $a_{L,i}(k)$ to preserve the performance level of the learning-based controller. Nevertheless, it is constrained as follows. Since the routes of some of the vehicles may be crossed, a constraint for avoiding collision is formed. The actuation $a_i(k)$ must provide motion for vehicle i, with which the safe distance s_{safe} between vehicle i and the further vehicles can be guaranteed. Moreover, the velocity must be inside of a bounded range. The upper bound is determined by the speed limit v_{max} and the lower bound is represented by the stopping of the vehicle. Thus, it is necessary to select Δ_i for all vehicles to keep velocities inside of the range.

The selection process of Δ_i for all vehicles is formed as an optimization problem as follows:

$$\min_{\Delta_1(k)...\Delta_n(k)} \sum_{i=1}^{n} \big(a_i(k) - a_{L,i}(k)\big)^2 \tag{3a}$$

subject to

$$(s_i(k+1) - s_j(k+1))^2 \geq s_{safe}, \qquad \forall i,j \in n, \tag{3b}$$

$$0 \leq v_i(k+1) \leq v_{max}, \qquad \forall i \in n, \tag{3c}$$

$$\Delta_i \in \mathbf{\Delta}_i, \qquad \forall i \in n, \tag{3d}$$

where i,j represent the pair of vehicles, whose motion can be in conflict, i.e. their routes are intercrossed. $\mathbf{\Delta}_i$ represents the domain of the optimization variable. In the optimization problem the kinematics of the vehicle motion (1) is considered through the formulation of the constraints, the separation of the control input (2) is involved in the objective function.

The objective of (3) is reformulated through (2):

$$\sum_{i=1}^{n} \left(a_i(k) - a_{L,i}(k)\right)^2 = \Delta(k)^T I \Delta(k) + 2f^T \Delta(k), \quad (4)$$

where $\Delta(k) = [\Delta_1(k)\ldots\Delta_n(k)]^T$, I is an $n \times n$ identity matrix and f vector is formed as $f = [a_{K,1} - a_{l,1}\ldots a_{K,n} - a_{l,n}]$.

The constraint for collision avoidance (3b) is formed to achieve keeping s_{safe} between the vehicles. The distance is measured in the sense of the longitudinal displacement of the vehicles on their route. Geometrically, the quadratic constraints (3b) represent that the trajectories of each related vehicles must be out of a circle. The radius of the circle is defined by s_{safe}, see Figure 2(a).
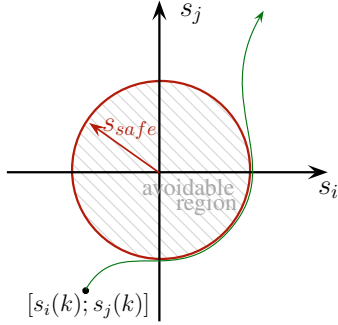

Fig. 2. Geometrical illustration of the quadratic constraints

Although the circle determined the avoidable region of the state-space, formally, it results in a quadratic constraint in (3). Nevertheless, the optimization problem must be solved in each k time step during the motion of the vehicles, and thus, the use of quadratic constraints can make difficulties for achieving a real-time solution. Consequently, it recommended to find an alternative formulation, e.g. the approximation of the quadratic constraints with linear constraints. First, the tangent lines to the circle from
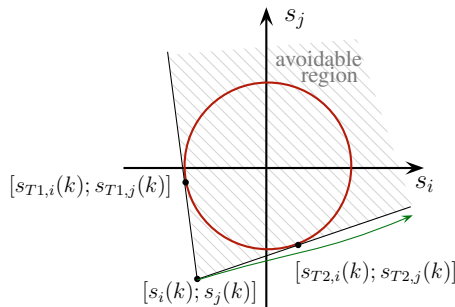

Fig. 3. Illustration of constraint approximation

the actual state $[s_i(k) \ s_j(k)]^T$ are assigned. The avoidable half-plane is determined by the region between the tangent lines, i.e. the trajectories must be out of it, see Figure 3. Second, two linear inequality constraints are specified, which represent that the trajectory of the state must be out of the avoidable half-plane, such as

$$\begin{bmatrix} s_{T1,i}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} s_i(k) \\ s_j(k) \end{bmatrix} \leq \begin{bmatrix} s_{T1,i}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} s_i(k+1) \\ s_j(k+1) \end{bmatrix}, \quad (5a)$$

$$\begin{bmatrix} s_{T2,i}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} s_i(k) \\ s_j(k) \end{bmatrix} \geq \begin{bmatrix} s_{T2,i}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} s_i(k+1) \\ s_j(k+1) \end{bmatrix}, \quad (5b)$$

where $[s_{T1,i}(k) \ s_{T1,j}(k)]$, $[s_{T2,i}(k) \ s_{T2,j}(k)]$ are the tangent points on the circle in time step k.

Longitudinal displacement values at $k+1$ in (5) are transformed to express the linear constraints in term of Δ. The transformation is based on the motion equation (1) and the relation of actuation separation (2), such as

$$s_i(k+1) = s_i(k) + Tv_i(k) + \frac{T^2}{2}a_{K,i}(k) + \frac{T^2}{2}\Delta_i(k), \quad (6a)$$

or

$$s_j(k+1) = s_j(k) + Tv_j(k) + \frac{T^2}{2}a_{K,j}(k) + \frac{T^2}{2}\Delta_j(k), \quad (6b)$$

which can be substituted into (5). It leads to the linear constraints

$$\begin{bmatrix} s_{T1,i}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} -Tv_i(k) - \frac{T^2}{2}a_{K,i}(k) \\ -Tv_j(k) - \frac{T^2}{2}a_{K,j}(k) \end{bmatrix} \leq$$
$$\frac{T^2}{2} \begin{bmatrix} s_{T1,i}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} \Delta_i(k) \\ \Delta_j(k) \end{bmatrix}, \quad (7a)$$

or

$$\begin{bmatrix} s_{T2,i}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} -Tv_i(k) - \frac{T^2}{2}a_{K,i}(k) \\ -Tv_j(k) - \frac{T^2}{2}a_{K,j}(k) \end{bmatrix} \geq$$
$$\frac{T^2}{2} \begin{bmatrix} s_{T2,i}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} \Delta_i(k) \\ \Delta_j(k) \end{bmatrix}. \quad (7b)$$

Figure 3 illustrates that the reachable set for the state $[s_i(k+1) \ s_j(k+1)]^T$ is non-convex, which means that (7) formulates disjunctive inequalities. However, each constraints in (7) lead to convex reachable sets, which means that the optimization problem can be separated, as it is proposed below.

Another constraint in the optimization (3) is on the velocity of the vehicles, see (3c). In case of this constraint $v_i(k+1)$ is expressed in term of Δ using the motion equation (1) and the relation of actuation separation (2). The linear inequality constraints are formed as

$$0 \leq v_i(k) + Ta_{K,i}(k) + T\Delta_i(k), \quad \forall i \in n, \quad (8a)$$
$$v_{max} \geq v_i(k) + Ta_{K,i}(k) + T\Delta_i(k), \quad \forall i \in n. \quad (8b)$$

which leads to

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} \Delta_i(k) \leq \begin{bmatrix} \dfrac{v_i(k)}{T} + a_{K,i}(k) \\ \dfrac{v_{max} - v_i(k)}{T} - a_{K,i}(k) \end{bmatrix}, \quad \forall i \in n. \quad (9)$$

The last constraint in the optimization problem (3) is the limitation of the resulted optimization variable, see (3d). The value of Δ_i is limited by the bounds of $a_i(k)$, such as $a_{min,i}$, $a_{max,i}$, which represent full braking and throttle. Since $a_i(k)$ is also influenced by $a_{K,i}(k)$, the constraints on $\Delta_i(k)$ are formed as

$$a_{min,i} - a_{K,i}(k) \leq \Delta_i(k), \quad \forall i \in n, \quad (10a)$$
$$a_{max,i} - a_{K,i}(k) \geq \Delta_i(k), \quad \forall i \in n, \quad (10b)$$

which leads to the constraint

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} \Delta_i(k) \leq \begin{bmatrix} a_{K,i}(k) - a_{min,i} \\ a_{max,i} - a_{K,i}(k) \end{bmatrix}, \quad \forall i \in n. \quad (11)$$

The optimization task (3) using (4), (7), (9) and (11) is reformulated as

$$\min_{\Delta(k)} \Delta(k)^T I_{n \times n} \Delta(k) + 2f^T \Delta(k) \qquad (12a)$$

subject to

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} \Delta_i(k) \le \begin{bmatrix} \dfrac{v_i(k)}{T} + a_{K,i}(k) \\ \dfrac{v_{max} - v_i(k)}{T} - a_{K,i}(k) \end{bmatrix}, \quad \forall i \in n,$$
$$(12b)$$

and

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} \Delta_i(k) \le \begin{bmatrix} a_{K,i}(k) - a_{min,i} \\ a_{max,i} - a_{K,i}(k) \end{bmatrix}, \quad \forall i \in n, \qquad (12c)$$

and

$$\begin{bmatrix} s_{T1,i}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} -Tv_i(k) - \dfrac{T^2}{2} a_{K,i}(k) \\ -Tv_j(k) - \dfrac{T^2}{2} a_{K,j}(k) \end{bmatrix} \le$$
$$\dfrac{T^2}{2} \begin{bmatrix} s_{T1,i}(k) \\ s_{T1,j}(k) \end{bmatrix}^T \begin{bmatrix} \Delta_i(k) \\ \Delta_j(k) \end{bmatrix}, \quad \forall i,j \in n, \quad (12d)$$

or

$$\begin{bmatrix} s_{T2,i}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} -Tv_i(k) - \dfrac{T^2}{2} a_{K,i}(k) \\ -Tv_j(k) - \dfrac{T^2}{2} a_{K,j}(k) \end{bmatrix} \ge$$
$$\dfrac{T^2}{2} \begin{bmatrix} s_{T2,i}(k) \\ s_{T2,j}(k) \end{bmatrix}^T \begin{bmatrix} \Delta_i(k) \\ \Delta_j(k) \end{bmatrix}, \quad \forall i,j \in n. \quad (12e)$$

The quadratic optimization in (12) contains disjunctive inequalities. It means that the optimization task for the solution can be reformulated to a mixed-integer optimization problem Belotti et al. [2011]. Nevertheless, in the given optimization problem alternative solution method can be found, because the distinct quadratic scenarios are related to each convex constraints. For example, two intercrossing vehicles results in two quadratic optimization tasks with the same objective function, but with different constraints. In case of the presented scenario in Figure 1 four quadratic optimization tasks are resulted with the same objective function, but with different constraints. Since the objective functions have the same formula in each optimization tasks, the results of the optimization can be evaluated through the comparison of their objectives. Therefore in practice, the optimization problem (12) through the solution of various quadratic optimization tasks in each steps is found. The solution is resulted by $\Delta(k)$, which leads to the minimum value of the objective, considering all of the optimization tasks.

3. DESIGN OF MOTION PROFILE USING REINFORCEMENT LEARNING

In this section the consideration of the economy aspects, such as the minimization of the longitudinal acceleration command in the design of the vehicle motion are presented. It is achieved through a reinforcement learning-based process, whose goal is to train neural-network-based agents, with which the predefined performance requirements are maintained.

The model for the learning process contains the optimization task (12). The model guarantees the avoidance of the collision in the intersection for every $a_{L,i}(k)$ signals. Thus, during the training process of the agent in every episodes the safety performances are guaranteed and similarly, the economy performance is improved. The output of the motion model is reward $r(k)$, which is composed by $a_i(k)$ and $v_i(k)$ as follows

$$r(k) = -Q_1 \sum_{i=1}^{n} a_i^2(k) + Q_2 \sum_{i=1}^{n} v_i(k), \qquad (13)$$

where Q_1 and Q_2 positive values are design parameters, which scale the importance of each terms in $r(k)$. The reward contains the control inputs in the vehicles $a_i(k)$, which represent the economy performance of the vehicles. If the reward contains only $a_i(k)$, it can result in unacceptable slow motion for the vehicles, because $a_i(k) = 0$ is the best choice tor the maximization of the reward. Thus, in the reward the velocity of the vehicles is also incorporated. The observation for the agent contains the positions of the vehicles $s_i(k)$ and their velocities $v_i(k)$. The goal of the reinforcement learning process is to maximize reward (13) during episodes. In this paper the training process through deep deterministic policy gradient (DDPG) is carried out, which is a model-free, online, off-policy reinforcement learning method Lillicrap et al. [2016].

In the example of Figure 1 the outputs of the agent are $a_{L,1}(k)$, $a_{L,2}(k)$, $a_{L,3}(k)$ and the observations contains the signals $s_1(k)$, $s_2(k)$, $s_3(k)$, $v_1(k)$, $v_2(k)$, $v_3(k)$. The initial values of the vehicles $(s_i(0), v_i(0))$ for the intersection scenarios in each episodes are generated randomly: $s_i(0)$ can vary between $-10 \ldots - 20 \ m$ and $v_i(0)$ is between $0 \ldots 50 \ km/h$. The actor network has 6 neurons in the input layer, 3 fully connected layers with 48 neurons and ReLu functions in each layers and 3 neurons with hyperbolic tangent functions in the output layer. The critic network has the same structure, but it also contains the actions as an input. The sampling time in each episodes is selected to $T = 0.05s$ and 500 episodes are carried out. The terms in the reward function are considered with the same design parameters, such as $Q_1 = Q_2 = 0.1$. The achieved value of the reward at the end of the training process is above 400. The result of the training process is an agent, whose outputs are $a_{L,i}(k)$. In the control process of the autonomous vehicles the agents works together with the control strategy (12).

4. SIMULATION RESULTS

In this section the effectiveness of the proposed method through simulation examples is illustrated. The example is the same as it has been presented in Figure 1, which contains three vehicles.

The results of two scenarios are presented. The initial positions of the vehicles are the same in both scenarios, such as $s_1(0) = -12m$, $s_2(0) = -5m$, $s_3(0) = -18m$. The safety distance is selected to $s_{safe} = 8m$, the input constraints are $a_{min,i} = -4m/s^2$ and $a_{max,i} = 3m/s^2$ for all vehicles. Samplig time is selected to $T = 0.05s$. In the first scenario the initial velocities of the vehicles are $v_1(0) = 5m/s$, $v_2(0) = 4m/s$, $v_3(0) = 4m/s$, which is

modified to $v_1(0) = 5m/s$, $v_2(0) = 4m/s$, $v_3(0) = 6m/s$ in the second scenario.

The scenarios with different initial values result in different control input and ordering in the intersection. The motions of the vehicles for each scenarios are illustrated in Figure 4. In both scenarios $Vehicle$ 2 reaches the intersection, but the ordering of the further vehicles are not the same in the two scenarios. In the first scenario $Vehicle$ 1 is the second (see Figure 4(b)), while in the second scenario $Vehicle$ 3 is the second in the ordering (see Figure 4(d)).

(a) Initial positions (Scenario 1)   (b) Conflict situation (Scen. 1)

(c) Initial positions (Scenario 2)   (d) Conflict situation (Scen. 2)
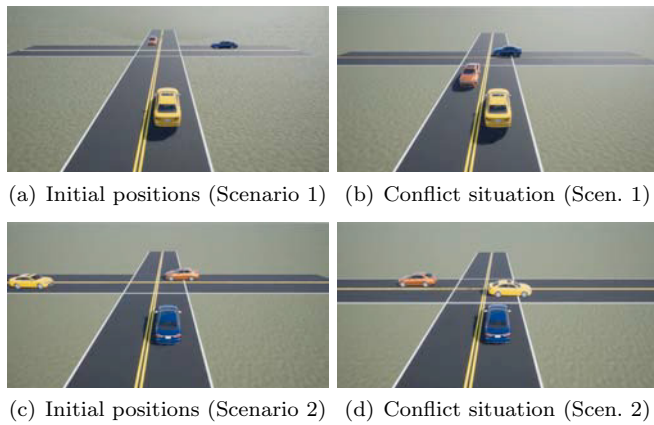
Fig. 4. Illustration of the intersection scenarios (from the viewpoint of the vehicle with last ordering)

The keeping of s_{safe} is illustrated in Figure 5 for each constraints and scenarios. Figure 5(a) shows that the trajectories of the vehicles are close to the border of the avoidable region, but in case of the constraint on s_1 and s_3 increased distance between the vehicles are achieved, see Figure 5(b). But, in the second scenario both trajectories are close to the borders, see Figure 5(c)-(d). These differences in the scenarios can also be seen in Figure 4(b) and Figure 4(d). In Scenario 1 $Vehicle$ 3 is far from $Vehicle$ 1 in the conflict situation (Figure 4(b)), while in Scenario 2 the vehicles are close to each other (Figure 4(b)). Moreover, the differences in the directions of the trajectories show the different ordering of $Vehicle$ 2 and $Vehicle$ 3. In case of Scenario 2 s_3 is significantly increased, which means that $Vehicle$ 3 reached the intersection earlier as $Vehicle$ 1.

The control input signals a_i and the output of the agents $a_{L,i}$ are shown in Figure 6. it can be seen that all a_i signal keep the input constraints, which underlines the effectiveness of the robust longitudinal controller. Moreover, it can be seen that a_i and $a_{L,i}$ have the same values in most of the simulations. The signals of $a_{L,i}$ are generally overwritten at the beginning of the simulations. For example, in Scenario 2 a_3 is slightly increased and a_2 is reduced to generate braking, with which the avoidance of the collision between $Vehicle$ 2 and $Vehicle$ 3 is achieved. Since $a_{L,i}$ signals are overwritten rarely, the improved economy performance of the learning agent is preserved.
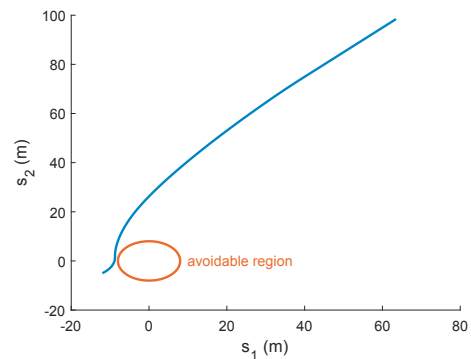
5. CONCLUSIONS

The paper has proposed the design of a learning-based control with guarantees for autonomous vehicles in intersection scenarios. The effectiveness of the method has been
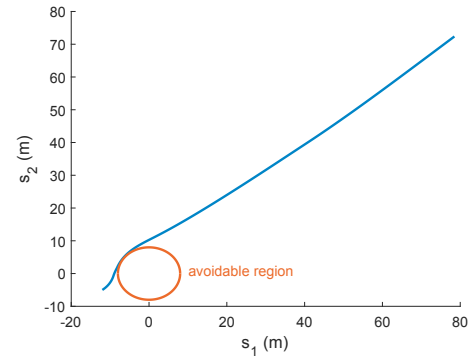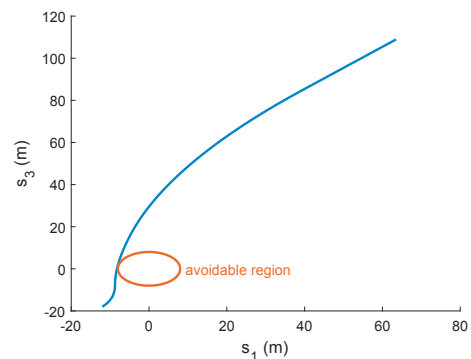

(a) $s_1 - s_2$ (Scenario 1)
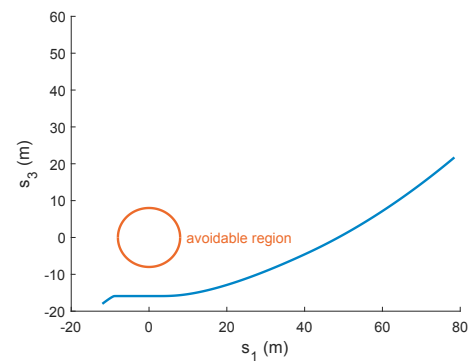
(b) $s_1 - s_3$ (Scenario 1)

(c) $s_1 - s_2$ (Scenario 2)

(d) $s_1 - s_3$ (Scenario 2)

Fig. 5. Illustration of the positions of the vehicles with the avoidable regions

illustrated through simulation examples on three vehicles, in which the avoidance of the collision has been carried

(a) Control inputs (Scenario 1)

(b) Agent outputs (Scenario 1)

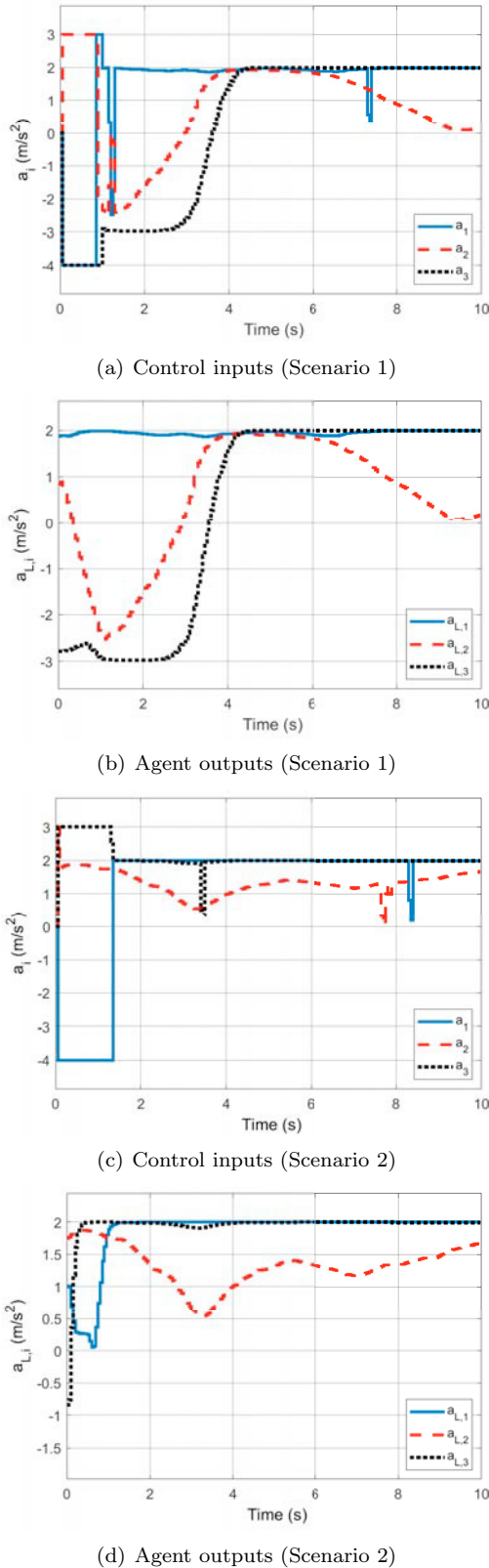(c) Control inputs (Scenario 2)

(d) Agent outputs (Scenario 2)

Fig. 6. Control inputs on the vehicles

out, and the economy performance of the vehicles has been improved. The method has resulted a centralized controller for the vehicles, with which their acceleration profiles have been influenced.

The future challenge of the method is to reformulate the reinforcement learning problem for decentralized controller. In this concept each vehicles have their own agent, which are the same, but they are independent from each other in their operation. The learning problem is to use the train the same agents parallel, using all of the rewards in the training process. The achievement of the concept is that the time requirement of the learning process can be significantly reduced.

REFERENCES

Pietro Belotti, Leo Liberti, Andrea Lodi, Giacomo Nannicini, and Andrea Tramontani. *Disjunctive Inequalities: Applications And Extensions.* American Cancer Society, 2011.

Youssef Bichiou and Hesham A. Rakha. Real-time optimal intersection control system for automated/cooperative vehicles. *International Journal of Transportation Science and Technology*, 8(1):1 – 12, 2019.

X. Chen, B. Xu, X. Qin, Y. Bian, M. Hu, and N. Sun. Non-signalized intersection network management with connected and automated vehicles. *IEEE Access*, 8: 122065–122077, 2020.

R. Hult, M. Zanon, S. Gros, and P. Falcone. Optimal coordination of automated vehicles at intersections: Theory and experiments. *IEEE Transactions on Control Systems Technology*, 27(6):2510–2525, 2019.

D. Isele, R. Rahimi, A. Cosgun, K. Subramanian, and K. Fujimura. Navigating occluded intersections with autonomous vehicles using deep reinforcement learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2034–2039, 2018.

K. Kim and P. R. Kumar. An mpc-based approach to provable system-wide safety and liveness of autonomous ground traffic. *IEEE Transactions on Automatic Control*, 59(12):3341–3356, 2014.

T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. Continuous control with deep reinforcement learning. *International Conference on Learning Representations*, 2016.

Balázs Németh, Péter Gáspár, Dávid Szőcs, and András Mihály. Design of the optimal motions of autonomous vehicles in intersections through neural networks. *IFAC-PapersOnLine*, 51(9):19 – 24, 2018. 15th IFAC Symposium on Control in Transportation Systems CTS 2018.

L. Riegger, M. Carlander, N. Lidander, N. Murgovski, and J. Sjöberg. Centralized mpc for autonomous intersection crossing. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1372–1377, 2016.

Tianhao Wu, Mingzhi Jiang, and Lin Zhang. Cooperative multiagent deep deterministic policy gradient (*comaddpg*) for intelligent connected transportation with unsignalized intersection. *Mathematical Problems in Engineering*, 2020.

M. Zhou, Y. Yu, and X. Qu. Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: A reinforcement learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 21(1):433–443, 2020.