Reinforcement Learning of Dynamic Collaborative Driving II: Lateral Adaptive Control

Luke Ng
Department of Mechanical and Mechatronics Engineering, University of Waterloo, 200 University Ave W., Waterloo, Ontario, N2L 3G1, Canada Fax: (519) 888-4333 Email: l4ng@engmail.uwaterloo.ca

Christopher M. Clark
Computer Science Department, California Polytechnic State University San Luis Obispo, CA, 93407, USA Fax: (805) 756-2956 Email: cmclark@calpoly.edu

Jan P. Huissoon
Department of Mechanical and Mechatronics Engineering, University of Waterloo, 200 University Ave W., Waterloo, Ontario, N2L 3G1, Canada Fax: (519) 888-4333 Email: jph@uwaterloo.ca

Abstract: In dynamic collaborative driving, multiple vehicles coordinate their motion to optimize road usage using shared information. The basic prerequisites for a vehicle participating in dynamic collaborative driving are longitudinal and lateral control. This paper focuses on the lateral vehicle control on which higher-level maneuvers such as entering or exiting a formation are based. The plant, a conventional automobile, is a composite nonlinear system powered by an internal combustion engine, equipped with automatic transmission, rolling on rubber tires with hydraulic braking systems and steering system. A vehicle model is introduced which serves as the control system design platform. A lateral adaptive preview control system which uses Monte Carlo Reinforcement Learning is introduced. The results of the reinforcement learning phase and the performance of the adaptive preview control system for a single automobile as well as the performance in a multi-vehicle platoon is presented.

Keywords: autonomous robotics; mobile robots; motion control; collaborative driving; lane-keeping; vehicle dynamics; vehicle simulation; artificial intelligence; machine learning; reinforcement learning; adaptive control


Copyright © 2008 Inderscience Enterprises Ltd.
Biographical notes: Luke Ng is a doctoral candidate at the Lab for Autonomous and Intelligent Robotics, Dept. of Mechanical and Mechatronics Engineering, University of Waterloo. Chris M. Clark is an Assistant Professor at the Computer Science Dept., California Polytechnic State University, San Luis Obispo, CA. Jan P. Huissoon is a Professor at the Dept. of Mechanical and Mechatronics Engineering, University of Waterloo. He is the currently the Department Deputy Chair and is a Professional Engineer.

1. Introduction

Throughout the world in both developed and developing countries the existing roadways are reaching their capacity as vehicle traffic flow increases. The construction of more roadways is being outpaced by the proliferation of motor vehicles. The adverse effects of increased vehicle traffic flow include traffic congestion, driving stress, increased vehicle collisions, pollution, and logistical delays. One promising solution is to automate the driving so as to optimize both road usage and driver safety. Dynamic Collaborative Driving is one approach to automated driving where multiple vehicles dynamically form groups and networks, sharing information in order to build a dynamic representation of the road to coordinate efficient road travel while maintaining safety.

The ultimate goal of our research is to create a decentralized control system capable of performing dynamic collaborative driving which is scalable to a large number of vehicles and can be implemented in any environment. However, basic control of the vehicle must be achieved before we can deal with the issue of coordination. In a previous paper (Ng et al 2008) the problem of longitudinal vehicle modelling and control was addressed. The focus of this paper is the study of the basic vehicle control problem of lateral control or steering, also referred to as lane-keeping and lane-changing, given that longitudinal control or adaptive cruise control has been established.

Early research in automated driving began with lateral vehicle control for lane following on passenger vehicles conducted by General Motors and RCA in the late 1950s (Kargels 1960). These studies included vehicle dynamics modelling, lane sensor development and the design of classical lateral controllers. Approaches to lateral control or automatic steering can be grouped into look-down and look-ahead systems, in terms of the measurement of lateral displacement. Look-down systems measure lateral displacement down from the front bumper using electric wire (Fenton et al 1976) or magnetic markers (Zhang and Parsons 1990), these systems deal with the immediate lateral position of the vehicle, that is they do not rely on preview information (Guldner et al 1994; 1997). Thus, the performance is more speed dependent, leads to more sluggish performance and larger lateral errors. When implemented in hardware, look-down systems are practical only at low speed of less than 20 m/s. Fenton and Selim (1984; 1988) employed an optimal control approach to design a velocity-adaptive, lateral controller for the design goals of lateral-position tracking accuracy, robustness, and ride comfort. Guldner (1994) and Pham et al (1996) used sliding mode control without preview to achieve lateral control. Later, the coupling of the longitudinal control with lateral control led to sliding surface control (Pham et al 1994; 1997). Rajamani et al (2000) implemented in hardware for Demo '97, a lateral control system incorporating...
both *look-down* and *look-ahead* control systems for integrated lane-keeping and lane-changing.

Alternatively, *look-ahead* systems replicate human driving by measuring lateral displacement ahead of the vehicle. With *look-ahead* system modalities such as machine vision (Jochem et al 1995) or radar (Unyelioglu et al 1996), it is possible to predict or anticipate where the vehicle is heading and provides a means of feed-forward control. Variya (1993) suggests that automatic steering systems should have some anticipatory capabilities. Prior to this, Peng and Tomizuka (1991) introduced a lateral feed-forward control algorithm which utilized preview information pertaining to road curvature as well as super-elevation angle. More recently, Netto et al (2004) conducted a simulation study to design a self-tuning linear regulator for lateral control. The controller is based on a simplified linear model of lateral vehicle dynamics.

Due to the high costs associated with procuring large numbers of vehicles and the safety issues involved, full-scale vehicle studies can only be conducted through large scale research projects in association with governments and automobile manufacturers such as Demo '97 (Thorpe et al 1997; Tan et al 1998; Rajamani et al 2000) and in Japan during Demo 2000 (Tsugawa et al 2000; Kato et al 2002). In Canada, smaller projects have used mobile robots to model cars (Michaud et al 2006), however the cost and complexity associated with these mobile robot studies can also be quite high. In addition the vehicle dynamics of a mobile robot platform are significantly different from those of full-sized automobiles thereby limiting the applicability of those results.

An alternative methodology is the use of simulation. Simulation studies have the benefit of faster development, flexibility, more cost effective, have better repeatability and explore situations not easily achieved in reality. For these reasons, the National Highway Traffic Safety Administration (NHTSA) began researching the use and construction of a new state-of-the-art driving simulator in 1989, the project was called the National Advanced Driving Simulator (NADS) (Haug 1990). NADS has been used as a substitute for actual vehicle testing. The NHTSA’s Vehicle Research and Test Center (VRTC) published vehicle data and matching NADS data for several vehicles such as the 1997 Jeep Cherokee (Salaani and Heydinger 2000) which can be used to validate new simulators. With the adoption of high fidelity simulation on modern computers, simulation has become the dominant method for study in this field.

The methodology used in our research involves first creating an accurate vehicle model to be used both in the process of design and validation of the control system. Our methodology can be considered a Computer Aided Engineering (CAE) approach to control design since it uses a computer model which allows the designer to assess the performance of the control system and predict its limits. We proceed with a description of the vehicle dynamics model followed by an explanation of the longitudinal control system's design, then, the results of the learning process and the evaluation of the system’s performance are presented.

### 2. Vehicle Dynamics Modelling

The vehicle dynamics model developed in this research has its roots going back to the late 1980’s. From the 1980’s to the late 1990’s, a significant amount of research was conducted at the Vehicle Dynamics Laboratory at the University of California at Berkeley by Hedrick under the PATH project which resulted in the development of a
complex numerical automobile model. The model was used to design and evaluate the performance of various controllers (McMahon and Hedrick 1989; Peng and Tomizuka 1991; Pham et al 1994) and went on to become a key tool in the development of more complex nonlinear controllers (Pham et al 1997).

Our vehicle dynamics simulation adopts many of the models used by Hedrick’s group for key subsystems such as the engine, transmission, suspension, tires and rigid body dynamics. However, in order to have a simulation which can be subjected to reinforcement learning, these separate models have to be integrated to provide system performance throughout its operating range. Figure 1 illustrates how each subsystem model is interconnected into a coherent model of an automobile.

In our previous paper (Ng et al 2008), the details of the longitudinal portions of our vehicle dynamics simulation are explained. The simulation used in this paper is the same with the addition of the Steering Actuator Model and the Tire Model.

2.1. Steering Model

In this simulation, a steering input signal of [-1, 1] is applied directly to steering actuator system. The steering actuator is modelled as a first order system with a time constant of $\tau = 0.125\ ms$. The range of the output, which is the tire angle $\delta_{\text{steer}}$, is [-15°, +15°]. The mapping from the steering input to output tire angle is linear with a first order response.

2.2. Tire Model

Pham et al (1997) describe a simplified tire model referred to as the Bakker-Pacejka model adopted from the work of Peng (1992). This model calculates the traction force resulting from the road-tire interaction based on empirical curve-fitting with experimental data for a Yokohama P205/60R1487H tire (Peng 1992). In this model, tire pressure, tire camber angle, and the road and tire physical parameters are fixed, but the forces generated at the tire are the functions of slip ratio $\lambda$, slip angle $\nu$, and the tire normal force $F_N$.

The surfaces shown in Figure 2 are plots of longitudinal and lateral force as function of normal force and slip ratio or slip angle respectively. These surfaces are based on test data of the Yokohama tires under laboratory conditions for the ideal friction coefficient of $\mu = 1.0$. To determine the longitudinal tire force the slip ratio is computed for both traction or braking using the following equations

\[
\lambda_{\text{Traction}} : \frac{r_\omega}{r_\nu} - \frac{V_t}{V_e} \geq 1 \quad \text{Braking:} \quad \lambda_{\text{Braking}} : \frac{r_\omega}{r_\nu} - \frac{V_t}{V_e} < 1
\]

where $r_\omega$ is the rotational speed of each wheel determined in the drive-train subsystem model (Section 3.1.4) and the radius of the tire is $r_\nu = 0.304\ m$.

To determine the lateral tire force, a moving reference frame is placed on the automobile with the $x$-axis aligned to the longitudinal axis of the car and the $y$-axis aligned with the lateral axis, the slip angle $\nu$ is simply the difference in angle between the tire’s orientation and its relative velocity vector angle $\zeta$ and can be determined with the following equation,

\[
\nu = \zeta - \dot{\nu} \quad \text{rad}
\]
Reinforcement Learning of Dynamic Collaborative Driving II: Lateral Adaptive Control

where $\delta$ is the steering angle of the tire and $i = 1, 2, 3, 4$ refers to front left, front right, rear left and rear right tires respectively. Note that for the rear tires ($i = 3, 4$) the steering angle $\delta = 0$. The relative velocity angle $\zeta$ is calculated using the following equations

$$
\zeta_1 = \arctan \left( \frac{V_x + l_f \dot{\psi}}{V_y - s_f \dot{\psi}} \right) \text{ rad} \quad (3)
$$

$$
\zeta_2 = \arctan \left( \frac{V_x + l_r \dot{\psi}}{V_y + s_r \dot{\psi}} \right) \text{ rad} \quad (4)
$$

where $V_x, V_y$ are the $x$ and $y$ components of the vehicle relative velocity vector, $\dot{\psi}$ is the vehicle’s yaw rate, $l_f = 1.034$ m and $l_r = 1.491$ m is the longitudinal distance from the center of gravity to the front and rear axles respectively, and $s_f = 1.450$ m and $s_r = 1.450$ m are the front and rear axles respectively.

According to Bakker et al (1987), road-tire interaction under non-ideal conditions can be extrapolated from the ideal curve by multiplying the ideal tire forces by the coefficient of friction $\mu$. Typically for average freeway operation, $\mu = 0.8$, for wet road conditions $\mu = 0.6$, and for icy road conditions $\mu = 0.2$.

2.3. Vehicle Model Response

The model is subjected to input signals for the steering command and the velocity response is charted, to demonstrate the performance and the validity of the vehicle dynamics simulation. The simulation is a composite of various subsystems that does not correlate to a standard vehicle (i.e. Ford V6 3.0L engine, Yokohama 15” radial tires, Toyota Camry chassis dimensions). Salaani and Heydinger (2000) work at the NHTSA’s Vehicle Research and Test Center (VRTC) provides some vehicle response data for a 1997 Jeep Cherokee. The data supplied concerning lateral response is limited, Figures 3 and 4 show the vehicle’s lateral acceleration and yaw rate response to a step and ramp steering input for the actual vehicle and its simulation. The results are collected with the vehicle under cruise control for 12 m/s and 11 m/s respectively.

Our simulation does not include a driver model which performs cruise control. In addition, it would be more desirable to obtain plots of lateral position to visualize the path taken by the vehicle when subject to various inputs. For these reasons, a commercial mechanical simulation called Adams Car (MSC Software Corp.) is used to validate our simulation instead. With Adams Car, one can obtain vehicle responses with or without cruise control. The vehicle modelled in Adams Car is a high performance sports car, however a comparison with the 1997 Jeep Cherokee shows the same behaviour. Figure 5 and 6 show the Adams Car simulation lateral acceleration and yaw rate responses for step and ramp inputs to steering. The Adams Car simulation matches the experimental vehicle data in terms of behaviour.

To demonstrate the validity of our simulation in modelling lateral vehicle dynamics, our simulation and the Adams Car simulation are subjected to a number of open-loop steering tests and their results compared. These tests are accomplished without the use of cruise control. Figure 7 through 10 show the results of each of these tests. The Adams Car results are shown first followed by the results of our simulation for the various inputs. For each test, the Adams Car simulation and our simulation match well. Note that for the ISO lane change test, the Adams Car simulation employs a driver model to execute the manoeuvre but our simulation does not have a driver model. Therefore a
piecewise combination of sinusoidal functions is used to emulate the manoeuvre. However, the responses do match.

3. Controller Design

Our research objective is to demonstrate that reinforcement learning algorithms can be applied effectively to the decentralized control of dynamic collaborative driving. In this paper, an adaptive lateral vehicle control system is introduced whereby the controller is learned through reinforcement learning using a complex nonlinear vehicle dynamics model. The output of the lateral control problem is a normalized steering angle of the front wheels while the inputs are i) lateral displacement and ii) the vehicle speed. The control problem therefore is to determine the correct steering angle trajectory to minimize the lateral displacement error while following the vehicle ahead. Figure 10 shows the desired trajectory of the vehicle for a single lane change or a step input in terms of lateral displacement at a constant speed. It shows that in order to achieve a smooth transition from one lateral position to another, the required steering signal trajectory is a specific delayed single sinusoidal cycle. Therefore, lateral vehicle movement can be considered a nonlinear control problem.

Our approach is to transform the control problem such the operating envelope can be subdivided into regions within which the behaviour of the plant approximates linearity. A common linear control law is formed where the gains would differ depending on the operating conditions. The control law minimizes a predicted (preview) angular error which gives this control system an anticipatory element. The common linear controller along with its collection of gains is considered a form of adaptive control referred to as gain scheduling (Astrom and Wittenmark 1994). The difference in our implementation of gain scheduling, is that the tedious task of determining the each gain is achieved using a form of machine learning called Monte Carlo ES reinforcement learning. In addition, the preview length used to determine the predicted angular error is also determined using Monte Carlo ES.

3.1 Reinforcement Learning

Reinforcement learning (RL) is a machine learning approach where a software agent senses the environment through its states and responds to it through its actions under the control of a policy, . This policy is improved upon iteratively based on its experiences with the environment through a reinforcement learning algorithm. The environment provides the agent with rewards, , which is numerical feedback for being in the current state. The challenge in reinforcement learning is to determine the actions which result in the maximum reward for every possible state, this state to action mapping is called the optimal policy .

The environment supplies the next state based on the current state and the actions taken using the transition function or model, . Since it may be difficult to obtain a transition model of a plant, one can use average sampled experiences as an equivalent transition model. In this case, the world becomes the transition model and leads to real-time learning. One could also create a simulated world that provides the simulated experiences to the agent for the process of learning.
In reinforcement learning, the control problem is formulated into a mathematical framework known as a finite Markov Decision Process (MDP) (Bellman 1957). To formulate an MDP, the states \( s \), actions \( a \), policy \( \pi(s,a) \), reward \( R(s) \) and transition model \( \sigma \) must be defined. The key feature of an MDP is that to be considered Markov, its current state must be independent of previous states so that for each visit to a state results in a path independent reward. Subsequent actions will result in new states giving rise to different rewards. For the current state, actions that result in more favourable future states lead to higher rewards. The favourability of a certain action given the current state is known as the \( Q-Value \). As an agent experiences its environment, it updates the \( Q-Value \) for each state-action pair it visits according to its reinforcement learning algorithm. As it repeatedly visits every state-action pair, it updates the policy so that the highest valued state-actions will dominate. The optimal policy is reached when the \( Q-Value \) function has been maximized, that is when every state-action pair results in the highest reward possible. The convergence of this maximization process requires that all states and actions be visited infinitely in order for the estimates of \( Q-Value \) to reach their actual values. To ensure this convergence criterion, policies leading to \( \pi^* \) are \( \varepsilon \)-soft, meaning that there is an \( \varepsilon \) probability that a random action is selected. Therefore, all actions and states will be reached as \( t \to \infty \). This process of policy improvement is referred to as a reinforcement learning algorithm. Specifically, Monte Carlo reinforcement learning algorithms improve the policy using the averaged sample returns experienced by the agent at the end of each episode (Sutton and Barto 1998).

The key to the process of improvement is the reward function which expresses the desirability of being in a current state. It is the method of communicating to the agent the task to be performed. The challenge of the designer is to be able to come up with a reward function that captures the essence of the task so that learning can be achieved.

### 3.2 Lateral Control

The problem of lateral control can be simply stated as a vehicle’s ability to follow another vehicle directly in front of it, in the lateral axis directly in front of it. It is assumed that the vehicle ahead is within sensing range and travelling at a constant speed so as to decouple the problem from longitudinal control (Ng et al 2008). Therefore, the lateral controller must maintain a relative lateral error of zero with the vehicle ahead while maintaining the fixed distance behind the forward vehicle using a separate longitudinal controller. Figure 13 shows the overall structure of the vehicle control system and how multiple vehicle’s are linked in terms of data to provide automated control for multiple vehicles.

Figure 14 shows the design of the lateral adaptive control. Central to the control system is the common linear control law in the form of a digital Proportional-Derivative (PD) controller which controls \( ey' \), the angular error between the current vehicle and the vehicle ahead. The difference equation of this controller which produces the final steering command \( m_n \) is shown below

\[
m_n = m_{n-1} + k_p (e_n - e_{n-1}) + k_d \frac{\Delta \theta}{\Delta T} (e_n - e_{n-1}) \quad (5)
\]

where \( n \) is the current iteration of the control cycle, \( e \) is \( ey' \), and \( \Delta T \) is the period of the control cycle; \( k_p \) and \( k_d \) are provided by the learned optimal policy and are functions of MDP state variables \( s_f = Vx \), the longitudinal vehicle velocity as described in Table 1.
The angular error $e\psi_1$ is determined within the Polar Transform block using the following equation

$$e\psi_1 = \arctan\left( \frac{y_1}{x_1} \right) - \psi$$  \hspace{1cm} (6)

where $\psi$ is the current yaw angle or heading angle of the vehicle, $e_x = x_{1\text{-}} - \psi$ is the longitudinal following between the vehicle and the vehicle ahead and $e_y = y_{1\text{-}} - \psi$ is the lateral previewed error. The lateral previewed displacement is $y' = \psi + k_{\text{preview}} V_x \sin(\psi)$ where the previewed distance $\Delta y_1$ is calculated in the Preview block using the following equation

$$\Delta y_1 = k_{\text{preview}} V_x \sin(\psi)$$  \hspace{1cm} (7)

where $k_{\text{preview}}$ is a learned gain provided by the optimal policy and $V_x$ is the vehicle speed.

For a given operating point, there are three parameters or gains which must be provided in a lookup table or schedule. One could obtain each value in the gain schedule for $k_p$ and $k_d$ using classical control techniques or state-space techniques. In this thesis, it is proposed that all gains can be learned using machine learning if the optimization problem can be formulated into a Markov Decision Process (MDP).

The nature of this optimization problem is episodic, that is, feedback as to the desirability of a given solution can only be provided at the end of an episode. The episode is defined as starting at the onset of a change in lateral position such as a commanded lane change. This follows the logic that when a new lateral command is required, a set of gains should be selected from the gain schedule and applied until the lateral error is eliminated. The goodness of a set of gains can therefore only be assessed once the command is complete. Since the MDP is episodic in nature, the Monte Carlo ES reinforcement learning algorithm described in Figure 12 is used to learn the gain schedule.

The MDP that describes this optimization problem requires: states, actions, a reward function and a transition model. The transitions from actions to states are provided by the vehicle. Since a detailed vehicle model was developed, the gain schedule can be obtained by offline Monte Carlo reinforcement learning using simulated experiences provided by the vehicle model.

The choice in the selection of states lies in the nonlinear nature of the steering plant. At different speeds the lateral dynamics respond differently, therefore, the controller gains will differ from a given speed. The actions are the preview gain and the two gains used in the digital control system. The states and actions along with their associated digitization sets are shown in Table 1 and Table 2.

Table 1 States of the lateral MDP

<table>
<thead>
<tr>
<th>State</th>
<th>Description</th>
<th>Digitization Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_j$</td>
<td>$V_x$: vehicle speed</td>
<td>{5, 10, 15, 20, 25, 30, 35, 40} m/s</td>
</tr>
</tbody>
</table>

Table 2 Actions of the lateral MDP

<table>
<thead>
<tr>
<th>Action</th>
<th>Description</th>
<th>Digitization Sets</th>
</tr>
</thead>
</table>
Reinforcement Learning of Dynamic Collaborative Driving II: Lateral Adaptive Control

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$k_{\text{preview}}$: Preview Length Gain</td>
<td>{0.1, 0.2, 0.3, … 9.9} $n_s = 100$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$k_p$: Proportional Gain</td>
<td>{0.01, 0.02, 0.03, … 0.99} $n_s = 100$</td>
</tr>
<tr>
<td>$a_3$</td>
<td>$k_d$: Derivative Gain</td>
<td>{0.01, 0.02, 0.03, … 0.99} $n_s = 100$</td>
</tr>
</tbody>
</table>

The reward function reflects the specification of the control problem. The reward is a piecewise continuous function of the normalized angular error variable $e\psi'$, the current normalized relative heading angle and is expressed below:

$$R(e\psi) = \begin{cases} \int_{0}^{10} & |e\psi| \leq 0.01 \\ \int_{0.01}^{1} & 0.01 > |e\psi| \geq 1 \end{cases}$$

For a given episode, the solution which maximizes the reward, or minimizes the $e\psi'$, will be favoured. These favoured solutions will be explored to determine the optimal solution.

4. Reinforcement Learning Experiments

The objective of the reinforcement learning (RL) experiments is to obtain an optimal policy $n^*$ for the lateral control of the vehicle. In these experiments the agent must follow another vehicle placed ahead of it travelling at a constant speed but with a different lateral position. For example, assuming the agent is at $y = 0$ m and the vehicle ahead is at the lateral position of $y = 3.6$ m, the vehicle ahead is considered to be in the adjacent lane to the left of the agent. Therefore, the agent must automatically steer itself at a constant speed to eliminate the lateral error between the two vehicles. It is assumed that the agent is using a longitudinal cruise control system (Ng et al 2008) and travelling at a constant speed with a fixed range to the vehicle ahead. Once the vehicle ahead has reached the end of the test track, the episode is complete. An experiment consists of 1000 episodes where $\varepsilon = 0.25$ of the $\varepsilon$-soft greedy policy for a particular combination of the three states. The distance of the test track is dependent on the speed of the lead vehicle using the following equation.

$$x_{\text{max}} = 1 + 0.2v_{\text{lead}} \times 300 \text{ m}$$

Each step of an episode generates a reward as per equation (7), this reward is accumulated during the course of an episode to measure the controller's tracking performance using a particular set of actions. Since it is possible to collide with the vehicle ahead during an episode, it would be beneficial if the reward were averaged to reflect how well the vehicle was able to follow the vehicle ahead during the course of the episode. Therefore, the average reward for the course of the entire episode is provided by the following equation.

$$R_{\text{avg}} = \frac{\sum_{t=1}^{T} r_t}{x_{\text{max}} - x_{\text{final}}}$$

Figure 15 shows the average reward as the agent progresses through the learning experiment for a particular state combination. The learning performance is similar for all combinations. One can observe a series of plateaus which increase as learning progress.

The resulting learned optimal policy is a collection of three functions of vehicle speed which represents each of the gains used in the adaptive controller.
Controller Performance Experiments

These experiments demonstrate the lateral tracking performance of the optimal policy for different speeds and for varying lateral positions. In these experiments, the vehicle is driven using the longitudinal controller (Ng et al 2008) at a constant initial speed. At time $t = 50$ s, a step input is fed to the lateral control system and the vehicle begins lateral tracking to reach its new lateral position.

Figure 16 shows the vehicle paths for executing a single lane change ($y_f = 3.6$ m) at various speeds (10, 20, and 30 m/s). Due to the different vehicle speeds the position where the lane change occurs varies. The vehicle path using the optimal policy gains for each velocity is fairly smooth. At 10 m/s, the path shows slight under-damping in the lateral axis with a very minimal overshoot. At 20 and 30 m/s, the path is critically damped with no overshoot. Despite increasing speed the paths remain very similar and the lane change is accomplished within 500 m indicating that lateral acceleration is increasing with velocity.

In Figures 17, 18 and 19 the vehicle paths are shown with respect to changing lateral position under a constant speed. The vehicle paths shown correspond to 1 m, 1.8 m (½ lane), 3.6 m (single lane) and 7.2 m (double lane). At 10 m/s (Figure 17) all paths reach the steady-state lateral position within 300 m. Figure 18 and 19 show the paths for 20 m/s and 30 m/s respectively, despite the different vehicle speeds the vehicle paths remain similar for a given lateral displacement. Therefore with higher speeds, the maneuvers are being accomplished under higher lateral accelerations. All vehicle paths are shown to be critically damped and smooth.

These experiments are important because they demonstrate that the lateral adaptive control system is capable of relatively quick and accurate lateral tracking. This tracking capability forms the basis for platoon maneuvers which allow vehicles to enter or exit its current platoon for decentralized dynamic collaborative driving control.

6. Multi-Vehicle Performance Experiments

These experiments demonstrate the performance of the lateral control system within a five car formation or platoon. Each car is equipped with a longitudinal control system (Ng et al 2008) which controls the throttle and brake so as to maintain a constant speed and a constant inter-vehicle spacing. In these experiments, the initial speed is set to 20 m/s (72 km/h) and inter-vehicle spacing is set to 20 m. At $t = 0$ s or at $x = 2.1$ km, the lead vehicle is placed at a specific lateral distance, each subsequent vehicle responds by following the vehicle ahead. The vehicle path for each of the five vehicles is recorded for the next 2 km. Four experiments are shown to illustrate the lateral controller performance, they are conducted at $y = 1$ m, 1.8 m (½ lane), 3.6 m (single lane) and 7.2 m (double lane).
Reinforcement Learning of Dynamic Collaborative Driving II: Lateral Adaptive Control

Figure 20 shows the 5 car formation for a lateral change in position of $y = 1 \text{ m}$. The response is slightly under-damped and the overshoot is propagated and increases with each subsequent vehicle. The resulting oscillations in the vehicle path decrease as it progresses through the experiment.

The response remains under-damped at a lateral position change of $y = 1.8 \text{ m}$ ($(\frac{1}{2} \text{ lane})$ as seen in Figure 21, the overshoot is smaller than at $y = 1 \text{ m}$. This trend continues as the lateral position change increases to $y = 3.6 \text{ m}$ (single lane) in Figure 22. At $y = 7.2 \text{ m}$ (double lane), the responses is almost considered critically damped (Figure 23). These experiments demonstrate that the lateral controller can be deployed effectively in a multi-vehicle platoon with minimal overshoot and oscillation for varying degrees of lateral tracking.

7. Conclusions

In this paper, the nonlinear nature of a vehicle’s lateral dynamics is shown. This is due to the nonlinearities present in the tire, suspension and the rigid body dynamics of the modeled vehicle. From this, we conclude that linearization of the lateral model may not be suitable for the entire operating range of the vehicle. The linear lateral controllers resulting from using a simplified linear model of the vehicle dynamics in the design process may only be adequate for a particular operating point.

The use of a more accurate nonlinear vehicle dynamics model in the design process should result in better nonlinear control systems for lateral control. In this paper, an adaptive control system using gain scheduling is introduced whereby the gains are learned using reinforcement learning. An element of anticipatory control is incorporated into a common PD control system. This element is a learned preview or look-ahead distance used is to calculate a predicted lateral error feedback which is controlled using the adaptive PD controller. Even with a simple reward function, it is possible for Monte Carlo ES reinforcement learning to converge upon an optimal policy within 1000 episodes for a particular operating regime; therefore, the MDP described in this paper properly models the task of lateral control.

When the learned optimal policies are combined to provide an adaptive control surface or a gain schedule, nonlinear control is achieved throughout the operating range. The performance of the controller at specific operating points shows accurate, smooth and quick tracking. In a multi-vehicle convoy or platoon, the lateral tracking performance is also quick, accurate and smooth with minimal overshoot and oscillation. The oscillations, although small do decrease as the tracking progresses implying stability. The performance of the adaptive controller in the multi-vehicle convoy or platoon is promising and forms the basis of higher level platoon maneuvers.

Acknowledgement

This research is funded by the Auto21 Network of Centres of Excellence, an automotive research and development program focusing on issues relating to the automobile in the 21st century. AUTO21 is a member of the Networks of Centres of Excellence of Canada program. Web site: http://www.auto21.ca
L. Ng, C.M. Clark, J. P. Huissoon

References


Reinforcement Learning of Dynamic Collaborative Driving II: Lateral Adaptive Control


Figures:
Figure 1 Schematic of the vehicle model

Figure 2 Longitudinal (left) and Lateral (right) Force-slip characteristics of the Yokohama P205/60R14 87H tire for $\mu = 1.0$ (ideal)

Figure 3 1997 Jeep Cherokee steering step input responses (Salaani and Heydinger 2000)

Figure 4 1997 Jeep Cherokee steering ramp input responses (Salaani and Heydinger 2000)

Figure 5 Adams Car steering step input responses
Reinforcement Learning of Dynamic Collaborative Driving II: Lateral Adaptive Control

Figure 6 Adams Car steering ramp input responses

Figure 7 Adams Car (left) and Simulation (right) steering impulse input response

Figure 8 Adams Car (left) and Simulation (right) steering step input response

Figure 9 Adams Car (left) and Simulation (right) steering ramp input response
Figure 10 Adams Car (left) and Simulation (right) single lane change

Figure 11 Schematic of lateral control
Reinforcement Learning of Dynamic Collaborative Driving II: Lateral Adaptive Control

Initialize, for all $s \in S, a \in A(s)$:

- $Q(s, a) \leftarrow \text{arbitrary}$
- $\pi(s) \leftarrow \text{arbitrary}$
- $\text{Returns}(s, a) \leftarrow \text{empty list}$

Repeat forever:
(a) Generate an episode using exploring starts $\pi$
(b) For each pair $(s, a)$ appearing in the episode
   - $R \leftarrow \text{return following the first occurrence of } (s, a)$
   - Append $R$ to $\text{Returns}(s, a)$
   - $Q(s, a) \leftarrow \text{average} (\text{Returns}(s, a))$
(c) For each $s$ in the episode:
   - $\pi(s) \leftarrow \text{argmax}_a Q(s, a)$

Figure 12: Monte Carlo ES-algorithm

Figure 13: Overview of lateral control system

Figure 14: Block diagram of lateral control system
Figure 15 Performance of Reinforcement Learning experiments

Figure 16 Single lane change performance
Reinforcement Learning of Dynamic Collaborative Driving II: Lateral Adaptive Control

Figure 17 Lateral performance at 10 m/s

Figure 18 Lateral performance at 20 m/s
Figure 19 Lateral performance at 30 m/s

Figure 20 Lateral performance for 5 car platoon at 1 m
Reinforcement Learning of Dynamic Collaborative Driving II: Lateral Adaptive Control

Figure 21 Lateral performance for 5 car platoon at 1.8 m

Figure 22 Lateral performance for 5 car platoon at 3.6 m
Lateral Performance of 5 Car Platoon at 7.2 m
(Double Lane Change)

Figure 23 Lateral performance for 5 car platoon at 7.2 m