Data-based reinforcement learning for lane keeping with input saturation

Rui Luo* and Dianwei Qian

School of Control and Computer Engineering,
North China Electric Power University,
Beijing, China
Email: hhu_neu@163.com
Email: dianwei.qian@ncepu.edu.cn
*Corresponding author

Qichao Zhang

State Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, Chinese Academy of Sciences,
University of Chinese Academy of Sciences,
Beijing, China
Email: zhangqichao2014@ia.ac.cn

Abstract: With the development of artificial intelligence, autonomous driving has received extensive attention. As a very complex integrated system, the autonomous vehicle has several modules. This paper is related to the control module, which is used to design an optimal or near-optimal controller to control the desired trajectory of the vehicle. In this paper, lateral control strategy for lane keeping task is proposed based on the model-free reinforcement learning. Different from the model-based methods such as linear quadratic regulator and model predictive control, our method only requires the generated data rather than the perfect knowledge of the system model to guarantee the optimal performance. At the same time, in order to meet two needs of passengers' comfort and fuel economy, input saturation should be considered in the design of the control module. A low-gain state feedback control method is adopted. It mainly solves some algebraic Riccati equations for data-based lateral control. Finally, the corresponding simulation is given and the validity of the algorithm is verified.

Keywords: lateral control; lane keeping; input saturation; model-free reinforcement learning.

Reference to this paper should be made as follows: Luo, R., Qian, D. and Zhang, Q. (2020) 'Data-based reinforcement learning for lane keeping with input saturation', *Int. J. Advanced Mechatronic Systems*, Vol. 8, No. 1, pp.9–15.

Biographical notes: Rui Luo is a Master student in the School of Control and Computer Engineering, The North China Electric Power University, Beijing, China. She received the BE degree in automation from Northeastern Electric Power University, Jilin, China, in 2017. Her research interest is autonomous driving lateral control.

Dianwei Qian is currently an Associate Professor in the School of Control and Computer Engineering, The North China Electric Power University, Beijing, China. He received his BE degree from the Hohai University (HHU), Nanjing, China, in 2003, and PhD degree from Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, in 2008. He also received his Master degree from the Northeastern University (NEU), Shenyang, China, in 2005. His research draws on the diverse control methods to aid in the analysis and design of complex dynamical systems.

Qichao Zhang is currently an Assistant Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. He received his BE in Automation from Northeastern Electric Power University, Jilin, China, in 2012, the Master degree in control theory and control engineering from Northeast University, Shenyang, China, in 2014, and PhD in Control Theory and Control Engineering from the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2017. His current research interests include reinforcement learning, game theory, and multi-agent systems.

1 Introduction

As the main driving force for the next generation of innovation, artificial intelligence (AI) has been booming in many fields. Autonomous driving, as one of the hot topics, has attracted wide attention on academia and industry. Many kinds of technologies are widely used in this field, especially reinforcement learning (RL). As a very complex integrated system, autonomous driving consists of perception module, positioning module, planning module, control module and so on, according to Pendleton et al. (2017). The main content of this paper is related to the control module, which is used to design the optimal or near-optimal controller to control the vehicle to follow the desired trajectory of the planning module. Mammar and Netto (2004) proposed that the dynamics of vehicle lateral and longitudinal are coupled. Therefore, they need to be designed together. Only under certain driving conditions (low vehicle speed or small steering angle), the lateral and longitudinal control can be designed separately. For advanced auxiliary driving systems, such as adaptive cruise control and lane-keeping system, some work focusing on lateral or longitudinal control has been proposed, according to Zhao et al. (2017) and Tamaddoni et al. (2011).

For the lateral control problem, the commonly used methods are linear quadratic regulator (LQR) and model predictive control (MPC) in the past decades. However, the above control methods need accurate system models. In fact, the parameters of various vehicles are not only different, but also difficult to determine. Therefore, a general system model is difficult to obtain. After introducing reinforcement learning theory, many researchers try to design vehicle stability controllers using data-based method. Model-free RL is a kind of data-driven method, such as Zhang et al. (2017) and Deng et al. (2006). It usually applies to a given environment. The agent that perceives and chooses the best action by maximizing the cumulative reward, which is pointed out by Kaelbling et al. (1996) and Deng and Inoue (2008). In order to design the longitudinal controller of connecting vehicles, the data-driven reinforcement learning method is adopted, it is pointed by Gao et al. (2017a) and Zhu et al. (2018). To solve the two-player nonzero-sum differential games with completely unknown linear discrete-time dynamics, Li et al. (2019) proposed a data-driven algorithm to learn the Nash equilibrium based on off-policy reinforcement learning. This algorithm is a fully model-free method, which solves the couple algebraic Riccati equations forward in time using measured data onto the system trajectories. The model-free RL scheme for the lateral control lane keeping is proposed by Zhang et al. (2019) the controller can be learned by measurable data without the perfect knowledge of the system dynamics.

Nowadays, most scholars believe that, in order to meet two needs of passengers' comfort and fuel economy, serious acceleration and deceleration should be avoided. In other words, input saturation should be considered in the design of the control module. Some scholars have made achievements on this issue. A trajectory tracking control scheme for USV in the presence of unknown disturbances and input saturation has been proposed by Qian et al. (2016). By virtue of that there is input saturation in the control system, an auxiliary system has been designed to address input constraint to ensure the USV system control performance. In order to control a class of uncertain nonlinear systems in the presence of input saturation and external disturbances, two new schemes are proposed by Gao et al. (2017b) to design adaptive controllers to compensate for the effects of the nonlinear saturation and disturbances. Inspired by the papers mentioned above, the research results of this paper are considering input saturation and developing data-based lateral control strategy. We adopt a low-gain state feedback control method, according to Wen et al. (2011). It mainly solves some algebraic Riccati equations for data-based lateral control.

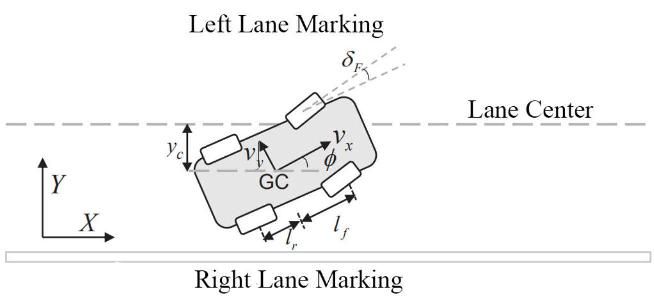
The rest of this paper is organized as follows. In Section 1, we mainly introduce the problem of the lateral control lane keeping. The data-based RL for the optimal control problem is proposed in Section 2. Then we design a low-gain controller in Section 3. At last, we give the simulation results and the conclusion in Sections 4 and 5.

2 Mathematical modelling

As shown in Figure 1, our model is a simplified linear single-track bicycle model. In order to control the vehicle travelling along the lane centre line, the system state vector is $x = (y_c, v_y, \varphi, \dot{\varphi})^T$ and four parameters represent the lateral offset of the vehicle's gravity centre, the velocity of the vehicle's gravity centre, the yaw angle and the yaw rate, respectively. From papers published by Rajamani et al. (2011) and Qian et al. (2008) we get the lateral vehicle dynamics:

$$\begin{aligned} \dot{x} &= Ax + B\delta(t) + C\rho \\ y &= Dx \end{aligned} \tag{1}$$

Figure 1 Vehicle lateral control model illustration



The control input of this system is steering wheel angle δ. ρ is the road curvature, which can be considered as a disturbance. The observation is $y = y_c$, which indicates the lateral offset at the centre of gravity. In this paper, matrix A, B, C and D in the lateral vehicle dynamics equation are expressed as follows.

$$A=\begin{bmatrix}0 & 1 & 0 & 0\\ 0 & -\frac{2C_{af}+2C_{ar}}{mV_x} & \frac{2C_{af}+2C_{ar}}{m} & \frac{-2C_{af}l_f+2C_{ar}l_r}{mV_x}\\ 0 & 0 & 0 & 1\\ 0 & \frac{-2C_{af}l_f+2C_{ar}l_r}{I_zV_x} & \frac{2C_{af}l_f-2C_{ar}l_r}{I_z} & \frac{-2C_{af}l_f^2+2C_{ar}l_r^2}{I_zV_x}\end{bmatrix}$$

$$B=\begin{bmatrix}0\\ \frac{2C_{af}}{r_{st}m}\\ 0\\ \frac{2C_{af}l_f}{r_{st}l_z}\end{bmatrix}\quad C=\begin{bmatrix}0\\ -\frac{2C_{af}l_f-2C_{ar}l_r}{m}-V_x^2\\ 0\\ -\frac{2C_{ar}l_f^2-2C_{ar}l_r^2}{I_z}\end{bmatrix}\quad D=\begin{bmatrix}1\\0\\0\\0\end{bmatrix} \tag{2}$$

l_f and l_r are the longitudinal distances from the centre of gravity to the front and rear wheels, respectively. m is mass. I_s is yaw moment of inertia. V_x means the longitudinal velocity, and is a common value for the sake of simplicity. The formula for calculating the steering gear ratio r_{st} is $r_{st} = \delta/\delta_F$ where δ and δ_F represent the steering wheel angle and the steering angle of the front wheels, respectively.

3 Adaptive optimal control with input saturation

In the model proposed in Section 2, if there is no disturbance ρ, only a feedback controller $\delta(t) = -Kx$ is needed to solve the control problem. We can evaluate a controller by the following cost function to obtain the optimal control.

$$J=\int_t^\infty \left(x^TQx+r\delta^2\right)dt \tag{3}$$

The two parameters Q and r in equation (3) represent a semi definite symmetric matrix and a normal number, respectively. After solving the following algebraic Riccati equation (ARE), we obtain the optimal control of minimizing the cost function (3).

$$A^TP+PA+Q-\frac{1}{r}PBB^TP=0 \tag{4}$$

According to the method of Huang et al. (2018), the controller is designed to fit the disturbance term of (1).

$$\delta(t)=-Kx+\left(U^*+KX^*\right)\rho \tag{5}$$

where $U^* \in R$ and $X^* \in R^n$ with $n = 4$ satisfy

$$\begin{aligned}&AX^*+BU^*+C=0\\ &DX^*=0\end{aligned} \tag{6}$$

If a control gain K in equation (5) satisfies then the asymptotic convergence of the system error is achieved, i.e., In other words, the task of lane keeping is achieved.

Let and the error dynamics can be described as

$$\begin{aligned}&\dot{\bar{x}}=A\bar{x}+B\bar{u}\\ &y=D\bar{x}\end{aligned} \tag{7}$$

Considering the optimal control problem of the error dynamics

$$\min_{\bar{u}}\int_t^\infty \left(\bar{x}^TQ\bar{x}+r\bar{u}^2\right)dt \tag{8}$$

The optimal control $\bar{u}=-B^TP\bar{x}$. Above, we express $\bar{u}$ as $\bar{u}=\delta(t)-U^*\rho$, so the optimal controller is

$$\delta^*(t)=-K^*x+\left(U^*+K^*X^*\right)\rho \tag{9}$$

Therefore, the optimal feedback gain is

$$K^*=r^{-1}B^TP \tag{10}$$

Consider the input saturation problem. Here we constrain the input $\bar{u}$

$$\bar{u}(t)=\begin{cases}\bar{u}_M & \text{if } \bar{u}>\bar{u}_M\\ \bar{u} & \text{if } |\bar{u}|\le\bar{u}_M\\ -\bar{u}_M & \text{if } \bar{u}<-\bar{u}_M\end{cases} \tag{11}$$

For input saturation as mentioned in equation (11), we first obtain some constants $\varepsilon = \varepsilon^* > 0$. These constants $\|K(\varepsilon^*)\bar{x}\|\le\bar{u}_M$ make with the optimal feedback gain

$$K\left(\varepsilon^*\right)=r^{-1}B^TP\left(\varepsilon^*\right) \tag{12}$$

After solving the following ARE, we get the unique solution $P(\varepsilon^*)$

$$A^TP\left(\varepsilon^*\right)+P\left(\varepsilon^*\right)A+\varepsilon^*Q-\frac{1}{r}P\left(\varepsilon^*\right)BB^TP\left(\varepsilon^*\right)=0 \tag{13}$$

Furthermore, $P(\varepsilon) \to 0$ as $\varepsilon \to 0$ refer to Gao et al. (2017c). On the basis of the inverse optimal control theory proposed in Krstic and Tsiotras (1999) it can be concluded that equation (12) is the optimal controller of the problem.

$$\min_{\bar{u}}\int_0^\infty \left(\varepsilon^*\bar{x}^TQ\bar{x}+r\bar{u}^2\right)dt \tag{14}$$

In order to obtain feasible control input under input saturation, we give the following lemmas.

Lemma 1 (2017): Considering any given bounded neighbourhood N^0 of the origin, there exists the constant $\varepsilon = \varepsilon^* > 0$ such that the system

$$\dot{\bar{x}}=\left(A-BK\left(\varepsilon^*\right)\right)\bar{x} \tag{15}$$

is asymptotically stable and $\left|K(\varepsilon^*)\bar{x}\right|\le\bar{u}_M$, the gain matrix $K(\varepsilon^*)$is given by $K(\varepsilon^*) = r^{-1}B^TP(\varepsilon^*)$.

Hence, the optimal control under the input constrained is given by

$$\delta^*(t)=-K\left(\varepsilon^*\right)x+\left(U^*+K\left(\varepsilon^*\right)X^*\right)\rho \tag{16}$$

Based on the formulas (7), (11) and (12), we can obtain the optimal control $P(\varepsilon^*)$, $K(\varepsilon^*)$, X^* and U^* on the premise that the system matrices A and B is known. In fact, most model parameters are usually different if the vehicle type is different, such as C_{af}, C_{ar}, l_f, l_r and the general model cannot be obtained. In order to avoid the time-consuming parameter adjustment process and design the optimal lateral control strategy for different types of vehicles, we need to design the optimal control without the knowledge of system dynamics. According to the adaptive optimal control methods of Gao et al. (2017b) and Zhu et al. (2018) we use the generated data to solve $P(\varepsilon^*)$, $K(\varepsilon^*)$, X^* and U^*.

Lemma 2 (1968): Let $K^0(\varepsilon^*)$ be any stabilizing feedback gain. For $j = 0, 1, \ldots$, the Lyapunov equation

$$\begin{aligned}&\left(A-BK^{(j)}(\varepsilon^*)\right)^T P^{(j)}+P^{(j)}\left(A-BK^{(j)}(\varepsilon^*)\right)\\&+\varepsilon Q+r\left(K^{(j)}\varepsilon^*\right)^T K^{(j)}(\varepsilon^*)=0\end{aligned} \quad 17)$$

is iteratively solved with the symmetric positive definite solution $P^{(j)}(\varepsilon^*)$. The feedback is updated by

$$K^{(j+1)}(\varepsilon^*)=r^{-1}B^T P^{(j)}(\varepsilon^*) \quad (18)$$

Then the following properties hold for all j

1 $(A - BK^{(j)}(\varepsilon^*))$ is Hurwitz

2 $P^* \le P^{(j+1)}(\varepsilon^*) \le P^{(j)}(\varepsilon^*)$

3 $\lim_{j\to\infty} K^{(j)}(\varepsilon^*) = K^*$, $\lim_{j\to\infty} P^{(j)} = P^*(\varepsilon^*)$

Based on Lemma 2, under the condition of the initial stability controller, we propose an adaptive optimal control, which can solve $P(\varepsilon^*)$, $K(\varepsilon^*)$, X^* and U^* according to generated data.

Motivated by Gao and Jiang (2016) can be described by

$$X^* = X^1 + \sum_{l=2}^{4} \alpha_l X^l \quad (19)$$

where $X^1 = 0_{4\times 1}$, $\alpha_l \in R$ and $X^l \in R^4$ with $DX^l = 0$ for $l = 2,3,4$.

Suppose an initial stabilizing control $\delta^{(0)}(x) = -K^0 x$ with given input constrained. We define $\hat{x}^l = x - X^l\rho$. Based on the vehicle lateral dynamics (1) with $\delta^{(0)}(x)$, we have

$$\dot{\hat{x}}_l = A\hat{x}_l - BK^0\hat{x}_l + (C + AX^l - BK^0X^l)\rho \quad (20)$$

Along the solution to (20), we differentiate $\hat{x}_l^T P^{(j)}(\varepsilon)\hat{x}_l$ for $\forall\varepsilon$. Then integrating it between the time interval $[t, t + T]$, we have

$$\begin{aligned}&\hat{x}_l^T(t+T)P^{(j)}\hat{x}_l(t+T)-\hat{x}_l^T(t)P^{(j)}\hat{x}_l(t)\\&=\int_t^{t+T}\Big(\hat{x}_l^T\left(A^TP^{(j)}+PA\right)\hat{x}_l-2\hat{x}_l^TP^{(j)}BK^{(0)}\hat{x}_l\\&\qquad+2\hat{x}_l^TP^{(j)}\left(C+AX^l-BK^0X^l\rho\right)\Big)d\tau\\&=\int_t^{t+T}\Big(2r\hat{x}_l^T\left(K^{(j+1)}\right)^T\left(\left(K^{(j)}-K^0\right)\hat{x}_l-K^0X^l\rho\right)\\&\qquad+2\hat{x}_l^TP^{(j)}\left(C+AX^l\right)\rho\\&\qquad-\hat{x}_l^T\left(\varepsilon Q+r\left(K^{(j)}\right)\right)^TK^{(j)}\Big)\hat{x}_l\Big)d\tau\end{aligned} \quad (21)$$

Following the Kronecker product representation, we have

$$\begin{aligned}&\hat{x}_l^T\left(\varepsilon Q+\left(K^{(j)}\right)^TK^{(j)}\right)\hat{x}_l=\left[\hat{x}_l\otimes\hat{x}_l\right]vec\left(\varepsilon Q+\left(K^{(j)}\right)^TK^{(j)}\right)\\&\hat{x}_l^T\left(K^{(j+1)}\right)^TX^l\rho=\left[\hat{x}_l\otimes\rho\right]\left[K^0X^l\otimes I_n\right]vec\left(K^{(j+1)}\right)\\&\hat{x}_l^TP^{(j)}\left(C+AX^l\right)\rho=\left[\hat{x}_l\otimes\rho\right]vec\left(\left(C+AX^l\right)P^{(j)}\right)\hat{x}_l^T\left(K^{(j+1)}\right)^T\\&\qquad\times\left(\left(K^{(j)}-K^0\right)\hat{x}_l\right)\\&\qquad=\left[\hat{x}_l\otimes\hat{x}_l\right]\left[\left(K^{(j)}-K^0\right)\otimes I_n\right]vec\left(K^{(j+1)}\right)^T\end{aligned}$$

where I_n represents a $n \times n$ unit matrix, and $vec(\ )$ is a vectorization operator, according to Qian et al. (2015) and Deng et al. (2011), which transforms a matrix into a vector by stacking its elements along the column direction. Defining the collected data between intervals $\{t_k, t_{k+1}\}$ as

$$\delta_{\hat{x}_l}=\begin{bmatrix}\hat{x}_l(t_1+T)-\hat{x}_l(t_1)\\\vdots\\\hat{x}_l(t_N+T)\hat{x}_l(t_N)\end{bmatrix}\quad I_{\hat{x}_l\hat{x}_l}=\begin{bmatrix}\int_{t_1}^{t_1+T}(\hat{x}_l\otimes\hat{x}_l)d\tau\\\vdots\\\int_{t_N}^{t_N+T}(\hat{x}_l\otimes\hat{x}_l)d\tau\end{bmatrix}$$

$$I_{\hat{x}_l\rho}=\begin{bmatrix}\int_{t_1}^{t_1+T}(\hat{x}_l\otimes\rho)d\tau\\\vdots\\\int_{t_N}^{t_N+T}(\hat{x}_l\otimes\rho)d\tau\end{bmatrix}$$

Then (21) becomes

$$\begin{aligned}&\delta_{\hat{x}_l}vec\left(P^{(j)}\right)-2I_{\hat{x}_l\rho}vec\left(\left(C+AX^l\right)P^{(j)}\right)\\&-2r\left(I_{\hat{x}_l\hat{x}_l}\left[\left(K^{(j)}-K^0\right)\otimes I_n\right]-I_{\hat{x}_l\rho}\left[K^0X^l\otimes I_n\right]\right)\\&\times vec\left(K^{(j+1)}\right)+I_{\hat{x}_l\hat{x}_l}vec\left(\varepsilon Q+r\left(K^{(j)}\right)^TK^{(j)}\right)=0\end{aligned} \quad (22)$$

Let

$$\Theta_l=\begin{bmatrix}\delta_{\hat{x}_l}^T\\-2r\left[\left(K^{(j)}-K^0\right)\otimes I_n\right]-I_{\hat{x}_l\rho}\left[K^0X^l\otimes I_n\right]^T\\-2I_{\hat{x}_l\rho}\end{bmatrix}^T$$

$$\Xi_l=I_{\hat{x}_l\hat{x}_l}vec\left(\varepsilon Q+r\left(K^{(j)}\right)^TK^{(j)}\right)$$

Assumption 1: If the following l = 1, ..., 4, there is a positive integer N^*, when all $N > N^*$, for all $j \ge 0$, Θ_l has full column rank.

With Assumption 1, the least-squares method can be used to solve (21)

$$\begin{bmatrix}vec\left(P^{(j)}\right)\\vec\left(K^{(j+1)}\right)\\vec\left(\left(C+AX^l\right)P^{(j)}\right)\end{bmatrix}=-\left(\Theta_l^T\Theta_l\right)^{-1}\Theta_l^T\Xi_l \quad (23)$$

Due to $X^1 = 0$ and B can be estimated by $B = rP^{(j)}(K^{(j+1)})^T$, C can be computed. AX^l is determined for $l = 2, 3, 4$. Then the parameters α_l and U^* can be computed by

$$\sum_{l=2}^{4}\alpha_l AX^l + BU^* + C = 0 \quad (24)$$

Algorithm 1 Data-based RL for lane keeping

Start with an initial admissible feedback gain K^0. Choose the parameters Q, e, g, ε, the time interval T and convergence bound η. $i \leftarrow 0, j \leftarrow 0$.
repeat
$i \leftarrow i + 1$
Let $\varepsilon = \varepsilon \times g$
Let $X^1 = 0$. Compute X^l for $l = 2, 3, 4$ with $CX^l = 0$. Collect the generated data Θ_l and Ξ_l under the initial controller for each X^l
repeat
$j \leftarrow j + 1$
Solve $P^{(j)}$, $K^{(j+1)}$ and $(C + AX^l)P^{(j)}$ from (23)
Until $\|K^{(j+1)} - K^{(j+1)}\| \le \eta$ with $j^* = j$
Determine AX^l for $l = 1, 2, 3, 4$, and compute the parameter α_l and U^* from (24) and X^* (from (19)
Until $|\bar{u}(t)| \le \bar{u}_M$ for the testing phase
Obtain the constrained optimal controller $\delta^* = -K^f x + (U^* + K^f X^*)\rho$

4 Simulation

We use numerical simulation method to prove that Algorithm 1 is effective against the lateral control lane keeping. The system parameters are chosen as follows.

$$m = 1150kg, I_z = 2000kgm^2, r_{st} = 1.78,$$
$$m_x = 80km/h, l_f = 1.27m, l_r = 1.37m,$$
$$C_{af} = C_{ar} \approx 84000N/rad$$

For the lane keeping task with input saturation, the range of steering wheel angle is constrained by [–0.6π, 0.6 π] under the road curve shown in the Figure. 2. Let the initial state $x(0) = [1, 0.5, 0.2, 0]$, $\varepsilon = 1$, g =0.9, $Q = diag(2, 0, 0, 0.2)$ and r =1.

Choose X^l for $l = 2, 3, 4$ as follows

$$X^2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad X^3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad X^4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

and the initial feedback gain is K^0 = [1.4 0.7 4.5 0.4] for X^l with $X^{1,2,4}$, and K^0 = [1 0.8 3.5 0.1] for X^3. Note that the initial feedback gain is a very important parameter for Algorithm 1. Here we choose them based on the experiments. The convergence bound η is chosen as 1e-7. For some ε, it is hard to find the converged solution, thus we let the max iteration number be 800 for j. After the fifth iteration for i, i.e., $\varepsilon = 0.656$ the system states are shown in Figure 3. At the 85s, the probing noise is removed, we can see that the system states are converged to zero under the learned controller. The optimal feedback gain is K = [1.14; 0.067; 3.544; 0.269] for i = 4. The convergence of $K^{(j)}$ for is shown in Figure 4. To compare the control input, we also record the optimal feedback gain K = [1.423; 0.084; 4.416; 0.335] for i = 0. We can get the sequence $\alpha_3 \approx 0.0165, \alpha_2 = \alpha_4 = 0$ and $U^* \approx 2.7491$.

Figure 2 Road curve

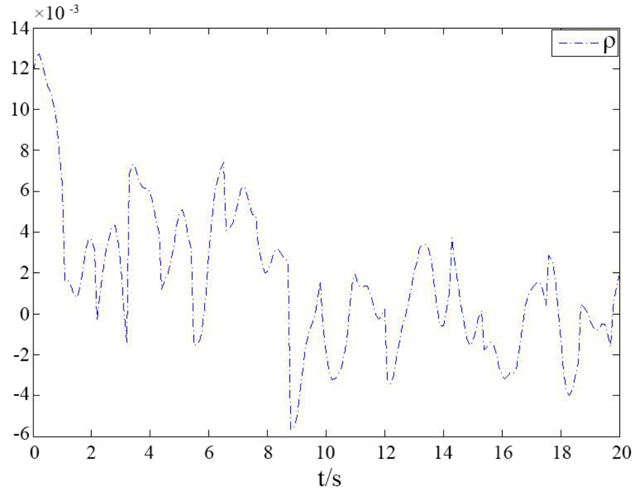


Figure 3 The trajectories of system states for $i = 4$ (see online version for colours)

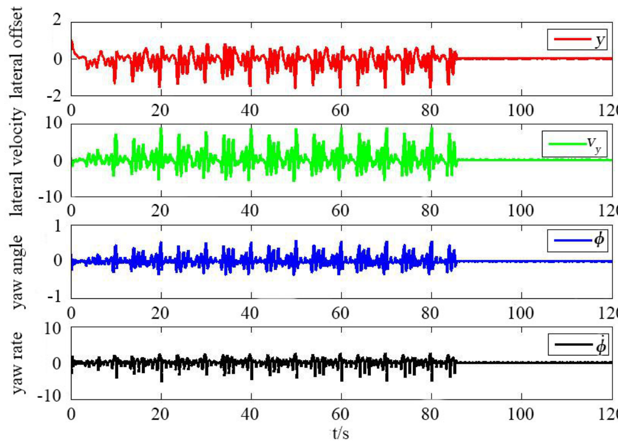
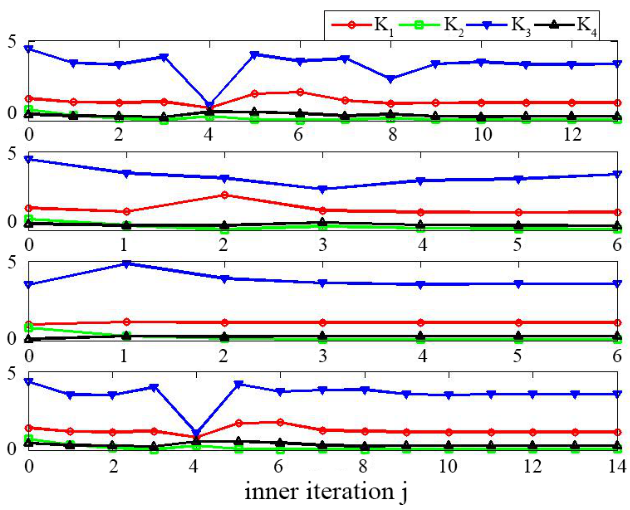


Figure 4 The iteration of feedback gain for i = 4 (see online version for colours)



Then we test the controller for $i = 4$ and $i = 0$ under the road curve in Figure 2. In Figure 5, the optimal control input for $i = 0$ is not constrained by [–0.6π, 0.6 π], and the minimum control input is –2.335, i.e., –0.74 π. Note that the optimal control input for $i = 4$ is constrained by [–0.6π, 0.6 π], and the minimum control input is –1.882, i.e., –0.599π.

Then we compare the constrained controller with the LQR controller. The system states under the LQR controller are shown in Figure 6, where the system states are fluctuating with the road curve, especially the lateral offset y. From Figure 7, the system states are converged to zero quickly, which is robust for the disturbance term ρ. The results demonstrate the effectiveness of the designed controller.

Figure 5 The control inputs for $i = 4$ and $i = 0$ (see online version for colours)

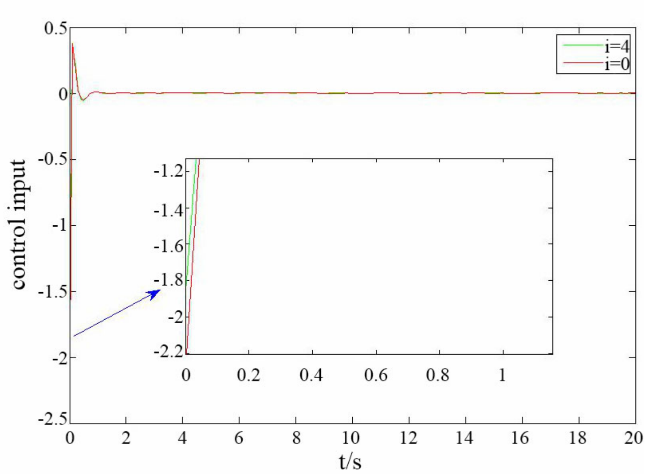


Figure 6 The trajectories of system states under LQR controller (see online version for colours)

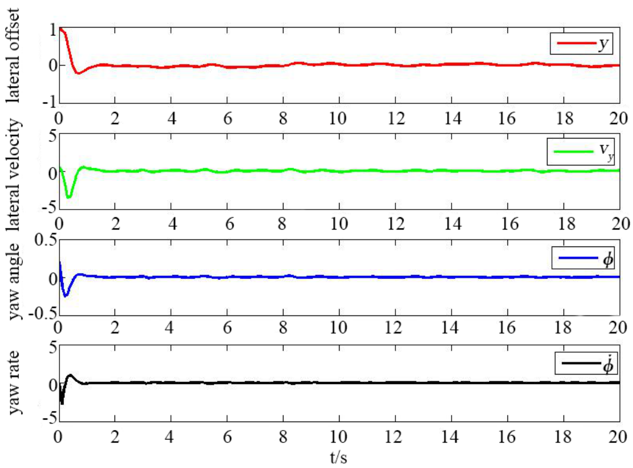
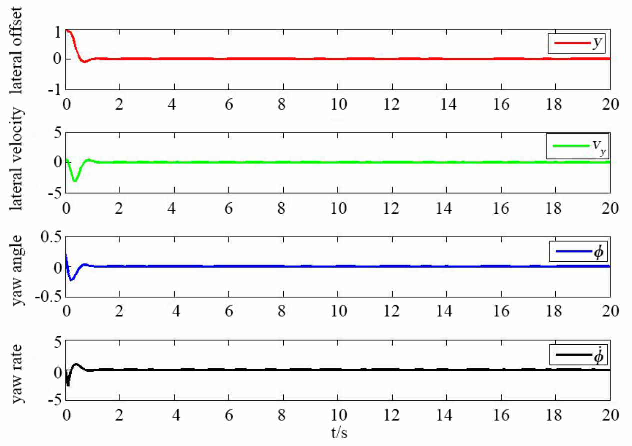


Figure 7 The trajectories of system states under constrained controller (see online version for colours)



5 Conclusions

In this paper, we investigate a data-based RL scheme for lane-keeping lateral control on the premise of input saturation. By solving the algebraic Riccati equation, a low gain control method is finally obtained. By learning from measurable data, we design a controller without perfect knowledge of system dynamics. Finally, the corresponding simulation is given and the validity of the algorithm is verified.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (NSFC) under Grants No. 61803371 and Beijing Science and Technology Plan under Grants Z181100004618003, partly by No. 61573353, No. 61533017 and No. 61873268.

References

Deng, M.C. and Inoue, A. (2008) 'Networked non-linear control for an aluminum plate thermal process with time-delays', *International Journal of Systems Science*, Vol. 39, No. 11, pp.1075–1080.

Deng, M.C., Inoue, A. and Ishikawa, K. (2006) 'Operator-based nonlinear feedback control design using robust right coprime factorization', *IEEE Transactions on Automatic Control*, Vol. 51, No. 4, pp.645–648.

Deng, M.C., Jiang, C.A., Inoue, A. and Su, C.Y. (2011) 'Operator-based robust control for nonlinear systems with Prandtl-Ishlinskii hysteresis', *International Journal of Systems Science*, Vol. 42, No. 4, pp.643–652.

Gao, W.N. and Jiang, Z-P. (2016) 'Adaptive dynamic programming and adaptive optimal output regulation of linear systems', *IEEE Transactions on Automatic Control*, Vol. 61, No. 12, pp.4164–4169.

Gao, W.N., Gutierrez, F.R. and Tong, W. (2017a) 'Cooperative adaptive cruise control of connected and autonomous vehicles subject to input saturation', *Electronics and Mobile Communication Conference*, pp.418–423.

Gao, W.N., Jiang, Z-P. and Ozbay, K. (2017b) 'Data-driven adaptive optimal control of connected vehicles', *IEEE Transactions on Intelligent Transportation Systems*, Vol. 18, No. 5, pp.1122–1133.

Gao, Y., Wang, N. and Zheng, Z.J. (2017c) 'Disturbance observer-based trajectory tracking control of unmanned surface vehicles with unknown disturbances and input saturation', *Chinese Control Conference*, pp.4859–4863.

Huang, M., Gao, W.N., Wang, Y. and Jiang, Z-P. (2018) 'Data driven shared steering control design for lane keeping', *IFAC-PapersOnLine*, Vol. 51, No. 13, pp.155–160.

Kaelbling, L.P., Littman, M.L. and Moore, A.W. (1996) 'Reinforcement learning: a survey', *Journal of Artificial Intelligence Research*, Vol. 4, pp.237–285.

Krstic, M. and Tsiotras, P. (1999) 'Inverse optimal stabilization of a rigid spacecraft', *IEEE Transactions on Automatic Control*, Vol. 44, No. 5, pp.1042–1049.

Li, L., Wang, L.M. and Yang, Y.L. (2019) 'Model-free solution to the discrete-time coupled riccti equation using off-policy reinforcement learning', the *38th Chinese Control Conference*, 27–30 July, Guangzhou, China, pp.6813–6818.

Mammar, S. and Netto, M. (2004) 'Integrated longitudinal and lateral control for vehicle low speed automation', *International Conference on Control Applications*, Vol. 1, No. 1, pp.350–355.

Pendleton, S.D., Andersen, H., Du, X.X., Shen, X.T. and Meghjani, M. (2017) 'Perception, Planning, control, and coordination for autonomous vehicles', *Machines*, Vol. 5, No. 6, pp.1–54.

Qian, D.W., Tong, S. and Lee, S.G. (2015) 'Fuzzy-logic-based control of payloads subjected to double-pendulum motion in overhead cranes', *Automation in Construction*, Vol. 65, pp.133–143.

Qian, D.W., Tong, S.W., Liu,H. and Liu, X.J. (2016) 'Load frequency control by neural-network-based integral sliding mode for nonlinear power systems with wind turbines', *Neurocomputing*, Vol. 173, pp.875–885.

Qian, D.W., Yi, J.Q. and Zhao, D.B. (2008) 'Hierarchical sliding mode control for a class of SIMO under-actuated systems', *Control and Cybernetics*, Vol. 37, No. 1, pp.159–175.

Rajamani, R. (2011) *Vehicle Dynamics and Control*, Springer Science & Business Media, Springer, Boston, MA.

Tamaddoni, S.H., Taheri, S. and Ahmadian, M (2011) 'Optimal preview game theory approach to vehicle stability controller design', *Vehicle System Dynamics*, Vol. 49, No. 12, pp.1967–1979.

Wen, C.Y., Zhou, J., Liu, Z.T. and Su, H.Y. (2011) 'Robust adaptive control of uncertain nonlinear systems in the presence of input saturation and external disturbance', *IEEE Transactions on Automatic Control*, Vol. 56, No. 7, pp.1672–1678.

Zhang, Q.C., Zhao, D.B. and Luo, C.M. (2019) 'Model-free reinforcement learning based lateral control for lane keeping', *International Joint Conference on Neural Networks*, Budapest, Hungary, 14–19 July, pp.1–7.

Zhang, Q.C., Zhao, D.B. and Zhu, Y.H. (2017) 'Data-driven adaptive dynamic programming for continuous-time fully cooperative games with partially constrained inputs', *Neurocomputing*, Vol. 238, pp.377–386.

Zhao, D.B., Xia, Z.P. and Zhang, Q.C. (2017) 'Model-free optimal control based intelligent cruise control with hardware-in-the-loop demonstration [research frontier]', *IEEE Computational Intelligence Magazine*, Vol. 12, No. 2, pp.56–69.

Zhu, Y.H., Zhao, D.B. and Zhong, Z. (2018) 'adaptive optimal control of heterogeneous CACC system with uncertain dynamics', *IEEE Transactions on Control Systems Technology*, Vol. 27, No. 4, pp.1772–1779.