Multi-agents Reinforcement Learning for a Type of Platoon Control Problem

July 18, 2020

Abstract

MSC Class:

Key Words: Autonomous vehicle, Platoon control, Reinforcement learning, Stackelberg game, Q-learning, Nash equilibrium

1 Background

With the continuous development and expansion of the world's automobile industry, its position in the world's economic construction is becoming more and more prominent, and automobile industry has gradually became a pillar industry in major automobile producing countries. Safety, energy saving and efficient passage are the eternal themes of automobile industry. In recent years, the continuous increase in vehicle ownership has exacerbated traffic congestion and environmental pollution, and led to an increase in the frequency of traffic accidents. The platoon control of vehicles can achieve a stable queue driving, thereby increasing the vehicle density of the road and improving the traffic efficiency. What's more, it can also enhance the safety of the transportation system and save chemical energy.

The queue control of vehicles is to form adjacent vehicles in a single lane, and automatically adjust the vehicle's motion state according to the information of adjacent vehicles, so that the adjacent vehicles maintain a stable distance and a consistent driving speed. Research on vehicle platoon control began with the PATH project in California in the 1980s. This project talked about many fundamental topics in platoon control, such as the allocation of control tasks, the layout of control architecture, technologies for perception and actuation, and longitudinal/lateral control laws. After that, plenty of attractive topics in the platoon control have been discussed, such as the choice of distance between vehicles, the impact of homogeneity and heterogeneity.

On the other hand, the information flow topology has a very important influence on platoon control. In the former studies, the information flow topology used was relatively single, which focused only on a few common structures, such as predecessor-following topology, predecessor-leader following topology, bidirectional topology and so on. Along with the development of communication technology, the communication between vehicles (Vehicle-to-Vehicle,V2V) is becoming more and more popular. A large number of different kinds of information flow topology structures can be produced in the vehicle platoon, including two-predecessor following topology and multiple-predecessor following topology. But there are some new challenges arise naturally when we consider the variety of topologies, such as time delay, packet loss and quantization error in the communications.

From the view of control, the platoon can be regarded as a dynamic system composed of multiple single vehicle nodes, where the controls can be formed for the individual vehicle through the information interaction between the nodes. Therefore, the platoon of vehicles can be regarded as a special multi-agent system which is a dynamic system formed by multiple agents with independent autonomy through the interaction of certain information topological structures. Under this perspective, a platoon can be decomposed into four interrelated sub-components: Node dynamic(ND), Information flow topology(IFT) , Formation geometry(FG) and Distributed controller(DC). The Node dynamic mainly describes the dynamic behavior of a single vehicle, including the position, velocity and acceleration of the vehicle. Information flow topology is a graph which can model the topological relation of information transfer between individual vehicle nodes. Formation geometry gives us the desired distance between adjacent vehicle. And distributed controller mainly reflects how can each vehicle adjusts its own behavior through the obtained information.

In this paper, we utilize the above framework and modify the models in [1]. As there are many uncertainties in the driving process and signal delay when receiving the information, we add the randomness to the Node dynamic. [1] mainly talks about the control of each vehicle to keep a stable distance and same speed between adjacent vehicles, which is a stability problem. We propose an optimization problem that gives a cost function for each vehicle in the platoon and try to find the Nash equilibrium. To achieve the goal, we adopt reinforcement learning approach and utilize the Stacklberg game theory to simplify the problem.

2 Problem Setup

2.1 General $(N+1)$-player game

Consider a $(N+1)$-player game, for player $i = 0, 1, \ldots, N$, the state is

$$dX_i(t) = b(X_i(t), u_i(t))\, dt + \sigma\, dW(t), \tag{1}$$

where $u_i(t)$ is the control of player i and $W(t)$ is a standard Brownian motion. Denote $X(t) = (X_0(t), X_1(t), \ldots, X_N(t))$, $u(t) = (u_0(t), u_1(t), \ldots, u_N(t))$, and for each player $i = 0, 1, \ldots, N$, we set the cost function as following:

$$J_i(u) = \mathbb{E}\Big[\int_0^T l_i(X(t), u(t))\, dt + g_i(X(T))\Big]. \tag{2}$$

The value function of each player $i = 0, 1, \ldots, N$ is

$$v_i(t, x) = \inf_{u_i \in \Pi} \mathbb{E}\Big[\int_t^T l_i(X(s), u(s))\, ds + g_i(X(T))\Big].$$

Our goal is to minimize these cost functions and obtain the Nash equilibrium $(u_0^*, u_1^*, \ldots, u_N^*)$, which can be achieved when

$$J_i(u_i, u_{-i}^*) \geq J_i(u_i^*, u_{-i}^*) \tag{3}$$

holds for all $i = 0, 1, \ldots, N$. The corresponding Hamilton–Jacobi–Bellman(HJB) equation for $v_i(x, t)$ is

$$\begin{cases} \partial_t v_i + \sum_{j=0}^N b(x_j, u_j^*)\nabla_j v_i + \frac{1}{2}(\sigma\sigma^\mathsf{T} D^2 v_i) + l_i(x, u_i^*) = 0 \\ u_i^* = \arg\min\limits_{u \in \Pi}\{b(x_i, u)\nabla_i v_i + l_i(x, a)\} \\ v_i(x, T) = g_i(x) \end{cases} \tag{4}$$

2.2 Platoon control

We consider a platoon running on a flat road with $N+1$ vehicles, including a leading vehicle (LV, indexed by 0) and N following vehicles (FVs, indexed from 1 to N). Motivated by the Node dynamic model in [1], we neglects the inertial delay in powertrain dynamics and assume that the vehicle dynamics is ideal double integrators. But considering the uncertainties in the driving process and signal delay, we add the randomness to the Node dynamic model, for $i = 0, 1, \ldots, N$,

$$\begin{cases} dp_i(t) = v_i(t)\, dt + \sigma_1\, dW_i(t) \\ dv_i(t) = u_i(t)\, dt + \sigma_2\, dB_i(t) \end{cases} \tag{5}$$

where $p_i(t)$ and $v_i(t)$ denote the position and velocity of vehicle i, σ_1 and σ_2 are two constants, $W_i(t)$ and $B_i(t)$ are standard Brownian motions and the control input $u_i(t)$ for this model is the acceleration of each vehicle. As we can not change the acceleration very sharply, we add a condition for the control input $-k \leq u_i(t) \leq k$, where k is a constant. By denoting

$$X_i(t) = \begin{pmatrix} p_i(t) \\ v_i(t) \end{pmatrix}, \quad b(X_i(t), u_i(t)) = \begin{pmatrix} v_i(t) \\ u_i(t) \end{pmatrix}, \quad \sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}, \quad \widetilde{W}_i(t) = \begin{pmatrix} W_i(t) \\ B_i(t) \end{pmatrix},$$

then the formula (5) can be transformed as

$$dX_i(t) = b(x_i(t), u_i(t))\, dt + \sigma\, d\widetilde{W}_i(t),$$

which is consistent with the dynamic system (1) of general $(N+1)$-player game.

The goal of the platoon control is to maintain the best distance between adjacent vehicles, keep the speed of each following vehicle consistent with that of the leader vehicle, and make the traffic efficiency as high as possible, so for the vehicle $i = 1, 2, \ldots, N$ in the platoon we can set the cost function as follows:

$$J_i(u_i) = \mathbb{E}\Big[\int_0^T \Big(u_i^2(t) + |p_i(t) - p_{i-1}(t) - d|^2 + |v_i(t) - v_{i-1}(t)|^2\Big)\, dt - p_i^2(T)\Big], \tag{6}$$

where d is the desired space between node $i-1$ and node i. And for the leading vehicle, we consider the cost function as follows

$$J_0(u_0) = \mathbb{E}\Big[\int_0^T u_0^2(t)\, dt - p_0^2(T)\Big]. \tag{7}$$

Our goal is to minimize these cost functions and obtain the Nash equilibrium $(u_0^*, u_1^*, \ldots, u_N^*)$, which can be achieved when

$$J_i(u_i, u_{-i}^*) \geq J_i(u_i^*, u_{-i}^*) \tag{8}$$

holds for all $i = 0, 1, \ldots, N$. So, to arrive the Nash equilibrium, the optimal strategy sequence $(u_0^*, u_1^*, \ldots, u_N^*)$ should satisfy $N+1$ inequalities. It is difficult for us to verify the condition of Nash equilibrium, thus we want to use the Stackelberg game theory to simplify this problem.

2.2.1 Case: N = 0

First we consider the case when there is only one vehicle in the platoon system. For $i = 0$, the state of the system governed by

$$dX_0(t) = \begin{pmatrix} dp_0(t) \\ dv_0(t) \end{pmatrix} = \begin{pmatrix} v_0(t) \\ u_0(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} dW_0(t) \\ dB_0(t) \end{pmatrix}, \tag{9}$$

where $u_0(t)$ is a progressively measurable process valued in Π, which is the admissible control set. The cost function is

$$J_0(u_0) = \mathbb{E}\Big[\int_0^T u_0^2(t)\, dt - p_0^2(T)\Big], \tag{10}$$

and the object is to minimize the cost function $J_0(t, x, u_0)$ over the control process. We introduce the associated value function

$$v(t, x) = \inf_{u_0 \in \Pi} \mathbb{E}\Big[\int_t^T u_0^2(s)\, ds - p_0^2(T)\Big].$$

The corresponding HJB equation for the above stochastic control problem is

$$\begin{cases} -\frac{\partial w}{\partial t} - \inf_{u_0 \in \Pi} \Big[x_{0,2} \frac{\partial w}{\partial x_{0,1}} + u_0 \frac{\partial w}{\partial x_{0,2}} + \frac{1}{2}\sigma_1^2 \frac{\partial^2 w}{\partial x_{0,1}^2} + \frac{1}{2}\sigma_2^2 \frac{\partial^2 w}{\partial x_{0,2}^2} + u_0^2\Big] = 0 \\ w(T, x) = -x_{0,1}^2. \end{cases} \tag{11}$$

Suppose the equation (11) admits the solution $w \in C^{1,2}([0, \infty), \mathbb{R}^2)$. Under this assumption, $u_0^*(t, x) = -\frac{1}{2}\frac{\partial w}{\partial x_{0,2}}$, and then the HJB equation (11) can be simplified as

$$\begin{cases} \frac{\partial w}{\partial t} - \frac{1}{4}\Big(\frac{\partial w}{\partial x_{0,2}}\Big)^2 + x_{0,2} \frac{\partial w}{\partial x_{0,1}} + \frac{1}{2}\sigma_1^2 \frac{\partial^2 w}{\partial x_{0,1}^2} + \frac{1}{2}\sigma_2^2 \frac{\partial^2 w}{\partial x_{0,2}^2} = 0 \\ w(T, x) = -x_{0,1}^2. \end{cases}$$

Denote $\tau = T - t$ and $w(t,x) = \widetilde{w}(\tau, x)$, then the following PDE can be obtained

$$\begin{cases} -\frac{\partial \widetilde{w}}{\partial \tau} - \frac{1}{4}\Big(\frac{\partial \widetilde{w}}{\partial x_{0,2}}\Big)^2 + x_{0,2}\frac{\partial \widetilde{w}}{\partial x_{0,1}} + \frac{1}{2}\sigma_1^2\frac{\partial^2 \widetilde{w}}{\partial x_{0,1}^2} + \frac{1}{2}\sigma_2^2\frac{\partial^2 \widetilde{w}}{\partial x_{0,2}^2} = 0 \\ \widetilde{w}(0,x) = -x_{0,1}^2. \end{cases}$$

Suppose

$$\widetilde{w}(\tau, x) = \phi_1(\tau)x_{0,1}^2 + \phi_2(\tau)x_{0,2}^2 + \phi_3(\tau)x_{0,1}x_{0,2} + \phi_4(\tau)x_{0,1} + \phi_5(\tau)x_{0,2} + \phi_6(\tau), \tag{12}$$

we have

$$\begin{cases} \frac{\partial \widetilde{w}}{\partial \tau} = x_{0,1}^2\phi_1^{'}(\tau) + x_{0,2}^2\phi_2^{'}(\tau) + x_{0,1}x_{0,2}\phi_3^{'}(\tau) + x_{0,1}\phi_4^{'}(\tau) + x_{0,2}\phi_5^{'}(\tau) + \phi_6^{'}(\tau) \\ \frac{\partial \widetilde{w}}{\partial x_{0,1}} = 2x_{0,1}\phi_1(\tau) + x_{0,2}\phi_3(\tau) + \phi_4(\tau) \\ \frac{\partial^2 \widetilde{w}}{\partial x_{0,1}^2} = 2\phi_1(\tau) \\ \frac{\partial \widetilde{w}}{\partial x_{0,2}} = 2x_{0,2}\phi_2(\tau) + x_{0,1}\phi_3(\tau) + \phi_5(\tau) \\ \frac{\partial^2 \widetilde{w}}{\partial x_{0,2}^2} = 2\phi_2(\tau) \end{cases}$$

Plugging it back to the PDE, we can get the Ricatti system of ODEs as follows

$$\begin{cases} \phi_1^{'}(\tau) + \frac{1}{4}\phi_3^2(\tau) = 0 \\ \phi_2^{'}(\tau) + \phi_2^2(\tau) - \phi_3(\tau) = 0 \\ \phi_3^{'}(\tau) + \phi_2(\tau)\phi_3(\tau) - 2\phi_1(\tau) = 0 \\ \phi_4^{'}(\tau) + \frac{1}{2}\phi_3(\tau)\phi_5(\tau) = 0 \\ \phi_5^{'}(\tau) + \phi_2(\tau)\phi_5(\tau) - \phi_4(\tau) = 0 \\ \phi_6^{'}(\tau) + \frac{1}{4}\phi_5^2(\tau) - \sigma_1^2\phi_1(\tau) - \sigma_2^2\phi_2(\tau) = 0 \end{cases}$$

with the initial condition

$$\phi_1(0) = -1, \phi_2(0) = \phi_3(0) = \phi_4(0) = \phi_5(0) = \phi_6(0) = 0.$$

2.2.2 Case: $N = 1$

In this section we consider the case when $N = 1$. There are two vehicles in the system. Suppose for the leading vehicle, we got the optional control $u_0^*(t)$ and the state function $X_0^*(t)$, then the dynamic of this system is governed by

$$\begin{cases} dX_0^*(t) = b(X_0^*(t), u_0^*(t))\,dt + \sigma\,d\widetilde{W}_0(t) \\ dX_1(t) = b(X_1(t), u_1(t))\,dt + \sigma\,d\widetilde{W}_1(t) \end{cases} \tag{13}$$

We want to minimize the cost function

$$J_1(u_1) = \mathbb{E}\Big[\int_0^T \Big(u_1^2(t) + |p_1(t) - p_0^*(t) - d|^2 + |v_1(t) - v_0^*(t)|^2\Big)\,dt - p_1^2(T)\Big],$$

and the value function is denoted by

$$v(t, x_1, x_0^*) = \inf_{u_1 \in \Pi} \mathbb{E}\Big[\int_t^T \Big(u_1^2(s) + |p_1(s) - p_0^*(s) - d|^2 + |v_1(s) - v_0^*(s)|^2\Big)\,ds - p_1^2(T)\Big].$$

The corresponding HJB equation is as following

$$\begin{cases} -\frac{\partial w}{\partial t} - \inf_{u_1 \in \Pi} \Big[x^*_{0,2} \frac{\partial w}{\partial x^*_{0,1}} + u^*_0 \frac{\partial w}{\partial x^*_{0,2}} + x_{1,2} \frac{\partial w}{\partial x_{1,1}} + u_1 \frac{\partial w}{\partial x_{1,2}} + \frac{1}{2}\sigma_1^2 \frac{\partial^2 w}{\partial x^{*2}_{0,1}} + \frac{1}{2}\sigma_2^2 \frac{\partial^2 w}{\partial x^{*2}_{0,2}} \\ \qquad + \frac{1}{2}\sigma_1^2 \frac{\partial^2 w}{\partial x^2_{1,1}} + \frac{1}{2}\sigma_2^2 \frac{\partial^2 w}{\partial x^2_{1,2}} + u_1^2 + (x_{1,2} - x^*_{0,2})^2 + (x_{1,1} - x^*_{0,1} - d)^2 \Big] = 0 \\ w(T,x) = -x^2_{1,1} \end{cases} \tag{14}$$

where $w(t,x) = w(t, x^*_{0,1}, x^*_{0,2}, x_{1,1}, x_{1,2})$. Suppose the equation (14) has the unique solution $w(t,x) \in C^{1,2}([0,\infty), \mathbb{R}^4)$. Under this assumption, we have $u_1^*(t,x) = -\frac{1}{2}\frac{\partial w}{\partial x_{1,2}}$, and then the HJB equation (14) can be simplified as

$$\begin{cases} \frac{\partial w}{\partial t} + x^*_{0,2} \frac{\partial w}{\partial x^*_{0,1}} + u^*_0 \frac{\partial w}{\partial x^*_{0,2}} + x_{1,2} \frac{\partial w}{\partial x_{1,1}} - \frac{1}{4}\Big(\frac{\partial w}{\partial x_{1,2}}\Big)^2 + \frac{1}{2}\sigma_1^2 \frac{\partial^2 w}{\partial x^{*2}_{0,1}} + \frac{1}{2}\sigma_2^2 \frac{\partial^2 w}{\partial x^{*2}_{0,2}} \\ \qquad + \frac{1}{2}\sigma_1^2 \frac{\partial^2 w}{\partial x^2_{1,1}} + \frac{1}{2}\sigma_2^2 \frac{\partial^2 w}{\partial x^2_{1,2}} + (x_{1,2} - x^*_{0,2})^2 + (x_{1,1} - x^*_{0,1} - d)^2 = 0 \\ w(T,x) = -x^2_{1,1} \end{cases}$$

3 Numerical method

Inspired by the numerical framework for generalized HJB equation in [3], we propose the similar numerical method to solve the HJB equation in our project. Suppose the real-value function $w(\cdot,\cdot)$ is defined on $[0,T] \times \bar{O}$, where O is bounded open set in $\mathbb{R}^d$. Π is a compact subset of Euclidean space $\mathbb{R}^d$, and $\sigma(\cdot,\cdot) : \bar{O} \times \Pi \mapsto \mathbb{R}^{d\times d}$ is a matrix-valued function, and $a(x,u) = \sigma(x,u)\sigma^{\mathsf{T}}(x,u)$. Next we give the numerical framework for the following HJB equation: for $\forall (t,x) \in [0,T] \times O$

$$\frac{\partial w}{\partial t} + \inf_{u \in \Pi} \Big[b(x,u) D_x w + \frac{1}{2}\mathbf{tr}(a(x,u) D_x^2 w) + l(x,u) \Big] = 0 \tag{15}$$

and for $(t,x) \in ([0,T) \times \partial O) \cup (\{T\} \times \bar{O})$,

$$w(t,x) = g(x).$$

Let e_i be the i-th unit basis of $\mathbb{R}^d$ for $i = 1, 2, \ldots, d$. For a given positive discretized parameter δ, h, define discrete spaces in state and time by

$$O^\delta = \{x \in O : x = \sum_{i=1}^n k_i \delta e_i, k_i \in \mathbb{Z}\} \tag{16}$$

and

$$[t,T]^h = [t,T] \cap \{t = kh + T : k \in \mathbb{Z}\}. \tag{17}$$

Next we introduce the central finite difference(CFD) numerical scheme and the upwind finite difference(UFD) numerical scheme.

3.1 CFD scheme

For $\forall \phi(t,x)$, the operators of the CFD scheme given by

$$\begin{cases} \Delta^\delta_{x_i} \phi = (2\delta)^{-1}[\phi(t, x + \delta e_i) - \phi(t, x - \delta e_i)] \\ \Delta^{2,\delta}_{x_i} \phi = \delta^{-2}[\phi(t, x + \delta e_i) + \phi(t, x - \delta e_i) - 2\phi(t,x)] \\ \Delta^{h,-}_t \phi = h^{-1}[\phi(t,x) - \phi(t-h,x)] \\ \Delta^{h,+}_t \phi = h^{-1}[\phi(t+h,x) - \phi(t,x)] \end{cases}$$

For simplicity, we use $w^h(t,x)$ to substitute $w^{\delta,h}(t,x)$. Applying the CFD scheme to the HJB equation (15), one can write the explicit numerical scheme as

$$\Delta_t^{h,-}W^h + \inf_u \Big[\sum_{i=1}^d b_i(x,u)\delta_{x_i}^{\delta} w^h + \frac{1}{2}\sum_{i=1}^d a_{ii}(x,u)\Delta_{x_i}^{2,\delta} w^h + l(x,u)\Big] = 0. \tag{18}$$

Plugging the operators into the formula (18), one can give the equivalent Markov chain approximation interpretation of above CFD scheme as follows

$$w^h(t-h,x) = \inf_u \Big[\sum_{y\in O^\delta} p^h(x,y,u)w^h(t,y) + hl(x,u)\Big]. \tag{19}$$

where

$$p^h(x,x\pm\delta e_i,u) = \frac{h}{2\delta^2}\sum_{i=1}^d \Big(a_{ii}(x,u) \pm \delta b_i(x,u)\Big),$$

and

$$p^h(x,x,u) = 1 - \frac{h}{\delta^2}\sum_{i=1}^d a_{ii}(x,u).$$

For the boundary condition, we have

$$w^h(T,x) = g(x).$$

To proceed, we need the following assumptions:

(A1) $a(x,u)$ satisfies

$$h\sum_{i=1}^d a_{ii}(x,u) \le \delta^2. \tag{20}$$

(A2) Discrete parameter $\delta = \delta(h)$ is a function of h, such that

$$h\sum_{i=1}^d \Big[a_{ii}(x,u) \pm \delta b_i(x,u)\Big] \ge 0. \tag{21}$$

Note that under the assumptions (20) and (21), we have

$$\sum_{y\in O^\delta} p^h(x,y,u) = 1, \quad p^h(x,y,u) \ge 0.$$

3.2 UFD scheme

The finite difference operators under the upwind scheme is as following: for $\forall\phi(t,x)$,

$$\begin{cases} \Delta_{x_i}^{\delta,+}\phi = \delta^{-1}[\phi(t,x+\delta e_i) - \phi(t,x)] \\ \Delta_{x_i}^{\delta,-}\phi = \delta^{-1}[\phi(t,x) - \phi(t,x-\delta e_i)] \\ \Delta_{x_i}^{2,\delta}\phi = \delta^{-2}[\phi(t,x+\delta e_i) + \phi(t,x-\delta e_i) - 2\phi(t,x)] \\ \Delta_t^{h,-}\phi = h^{-1}[\phi(t,x) - \phi(t-h,x)] \\ \Delta_t^{h,+}\phi = h^{-1}[\phi(t+h,x) - \phi(t,x)] \end{cases}$$

Applying the upwind difference scheme to the HJB equation (15), onw can write the explicit numerical scheme as

$$\Delta_t^{h,-} w^h + \inf_u \Big[\sum_{i=1}^d \Big(b_i^+(x,u)\Delta_{x_i}^{\delta,-} w^h - b_i^-(x,u)\Delta_{x_i}^{\delta,+} w^h \Big) + \frac{1}{2}\sum_{i=1}^d a_{ii}(x,u)\Delta_{x_i}^{2,\delta} w^h + l(x,u) \Big] = 0. \quad (22)$$

Plugging the UFD operators into the formula (22), then one can give the equivalent Markov chain approximation interpretation of the above UFD scheme as follows

$$w^h(t-h,x) = \inf_u \Big[\sum_{y\in O^\delta} p^h(x,y,u) w^h(t,y) + hl(x,u) \Big], \quad (23)$$

where

$$p^h(x, x-\delta e_i, u) = \frac{h}{2\delta^2}\sum_{i=1}^d \Big(a_{ii}(x,u) - 2\delta b_i^+(x,u) \Big),$$

$$p^h(x, x+\delta e_i, u) = \frac{h}{2\delta^2}\sum_{i=1}^d \Big(a_{ii}(x,u) - 2\delta b_i^-(x,u) \Big),$$

and

$$p^h(x,x,u) = 1 + \frac{h}{\delta^2}\sum_{i=1}^d \Big(\delta|b_i(x,u)| - a_{ii}(x,u) \Big).$$

Similarly, we need the following assumptions:

(A3) $a(x,u)$ and $b(x,u)$ satisfies

$$\sum_{i=1}^d \Big(a_{ii}(x,u) - 2\delta b_i^+(x,u) \Big) \geq 0, \quad \sum_{i=1}^d \Big(a_{ii}(x,u) - 2\delta b_i^-(x,u) \Big) \geq 0 \quad (24)$$

(A4) Discrete parameter $\delta = \delta(h)$ is a function of h, such that

$$h\sum_{i=1}^d \Big[a_{ii}(x,u) - \delta|b_i(x,u)| \Big] \leq \delta^2, \quad \sum_{i=1}^d \Big(\delta|b_i(x,u)| - a_{ii}(x,u) \Big) \leq 0 \quad (25)$$

Thus under the assumptions (24) and (25), one can get that

$$\sum_{y\in O^\delta} p^h(x,y,u) = 1, \quad p^h(x,y,u) \geq 0.$$

References

[1] S. E. Li, Y. Zheng, K. Li, L. Wang and H. Zhang. Platoon Control of Connected Vehicles from a Networked Control Perspective: Literature Review, Component Modeling, and Controller Synthesis, IEEE Transactions on Vehicular Technology, doi: 10.1109/TVT.2017.2723881.

[2] Pham, H. Continuous-time Stochastic Control and Optimization with Financial Applications, Springer-Verlag, 2009.

[3] Q. Song. Convergence of markov chain approximation on generalized HJB equation and its applications, Automatica, 44(3):761–766, 2008.