# PLATOON

THE AUTHOR

## 1. Introduction

## 2. State constraint differential game

The original problem on the platoon control problem of $N$ vehicles can be formulated as a state-constraint stackleburg differential game.

Consider a $N$-players game, for player $i = 1, \dots, N$, the state is governed by a SDE

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

where $u_i(t) \in \mathbb{R}$ is the control or acceleration of player $i$ and $W_i(t)$ is a standard Brownian motion. $X_i(t) = (X_{i0}(t), X_{i1}(t))$ is 2 dimensional process, where $X_{i0}$ is displacement, $X_{i1}$ is the velocity.

Denote $X(t) = (X_1(t), \dots, X_N(t))$, $u(t) = (u_1(t), \dots, u_N(t))$, and for each player $i = 1, \dots, N$, we set the cost function as following:

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

Our goal is to minimize these cost functions and obtain the Nash equilibrium $(u_1^*, \dots, u_N^*)$, which satisfies

$$J_i(u_i, u_{-i}^*) \geq J_i(u_i^*, u_{-i}^*) := v_i, \forall (u_i, u_{-i}^*), u^* \in \Pi_c \tag{3}$$

where $\Pi_c$ is the admissible control space defined by

$$\Pi_c = \{u \in L^2_{\mathbb{F}} : X_{i-1}(t) - X_i(t) \geq d_1, \forall t, \text{ w.p.1 } \}.$$

In the platoon control problem, $\ell_1(x, a)$ is a function depending only on $(x_1, a)$, and $\ell_i$ is a function depending only on $(x_i, x_{i-1}, a)$ for $i \geq 2$. This leads to the above Nash equilibrium becomes Stackelberg differential game.

Compared to traditional differential game, this problem is difficult due to the state constraint $X_{i-1}(t) - X_i(t) \geq d$ for all $t$ w.p.1. One has to verify its well posedness for such a constraint first. If $\sigma > 0$, then one can prove that the admissible control space is empty set due to the unboundedness of the Brownian motion. [1] If $\sigma = 0$, it is well posed. However, Bellman principle fails again due to the state constraint and the associated HJB equation does not exist.

[1] proof is needed

## 3. Stackelberg differential game with penalization

To remove the state constraint, we revise the running cost by, for large $\lambda > 0$,

$$\ell_i^\lambda(x,a) = \ell_i(x,a) + \lambda(x_{i-1} - x_i - d_1)^-, \forall i \geq 2.$$

Correspondingly, we revise the total cost by

$$J_i^\lambda(u) = \mathbb{E}\Big[\int_0^T \ell_i^\lambda(X(t), u(t))\, dt + g_i(X(T))\Big]. \tag{4}$$

and the Nash Equilibrium by

$$J_i^\lambda(u_i, u_{-i}^*) \geq J_i^\lambda(u_i^*, u_{-i}^*) := v_i^\lambda, \forall (u_i, u_{-i}^*), u^* \in \Pi := L^2_{\mathbb{F}}. \tag{5}$$

One can show that [2]

**Lemma 1.** *$v_i^\lambda \downarrow v_i$ as $\lambda \downarrow 0+$.*

Next verification theorem provides the connection of penalized problem to PDE. [3]

**Theorem 2.** *If $w_1, \ldots, w_N$ solves*

$$\begin{cases} \partial_t w_i + \sum_{j=0}^N b(x_j, u_j^*)\nabla_j w_i + \frac{1}{2}\sigma^2 \Delta w_i + \ell_i^\lambda(x, u_i^*) = 0 \\ u_i^* = \arg\min_a \{b(x_i, a)\nabla_i w_i + \ell_i^\lambda(x, a)\} \\ w_i(T, x) = g_i(x), \end{cases} \tag{6}$$

*then, $v_i^\lambda = w_i(0, x)$ and $u^*$ is the feedback form of Nash equilibrium.*

## 4. MDP

Next we provide approximating MDP for the penalized problem with penalization parameter $\lambda > 0$ and discretization parameter $h$, and it will be referred to [MDP($\lambda, h$)].

- State space: $\mathbb{R}^+ \times (\mathbb{R}^2)^N$
- Action space: $\mathbb{R}^N$.
- 1-step transition probability with step size $h$ and action $a$:
  - current state: $(t, x) \in \mathbb{R}^+ \times (\mathbb{R}^2)^N$
  - next state: If $t \geq T$, then stop. Otherwise,
    $$\hat{t} = t + h, \ \hat{x}_i = b(x_i, a_i)h + \sigma\sqrt{h}Z$$
    where $Z$ is standard normal random variable.
- Total cost: If $t < T$, then
  $$\Delta J_i = \ell_i^\lambda(x_i, a_i)\Delta t.$$
  If $t = T$, then
  $$\Delta J_i = g_i(x_i).$$
- objective: Find equilibrium $u_i^* : \mathbb{R}^+ \times (\mathbb{R}^2)^N \mapsto \mathbb{R}^N$.

---

[2] proof is needed

[3] proof is needed

The next theorem justifies the convergence of the approximating MDP. [4]

**Theorem 3.** *Let* $v_i^{\lambda,h}$ *be the value associated to the* $[\mathrm{MDP}(\lambda, h)]$*. Then,* $\lim_{h\to 0} v_i^{\lambda,h} = v_i^{\lambda}$.

## 5. Computation

We want compute the Nash Equilibrium of MDP$(\lambda, h)]$ by reinforcement learning with the following data: [5]

- $T = 1, \lambda = 1, h = .01$
- $g_i = 0$
- $\sigma = .1$
- $$b_i(x, a) = \begin{bmatrix} x_{i2} \\ a \end{bmatrix}$$
- $\ell_1(x, a) = (x_{12} - 60)^2 + a^2$
- $\ell_i(x, a) = (x_{i2} - x_{i-1,2})^2 + a^2$ for $i \geq 2$
- $$\ell_i^{\lambda}(x, a) = \ell_i(x, a) + \lambda(x_{i-1} - x_i - d_1)^-, \forall i \geq 2.$$

[4] proof is needed

[5] coding is needed