Краткий конспект лекций по курсу «Игры среднего поля»

Лекция 4

Задача оптимального управления

Пусть $f\colon \mathbb{R}^d \times A \to \mathbb{R}^d$, где A — подмножество $\mathbb{R}^m$, является непрерывной по совокупности переменных и удовлетворяет условию

$$|f(y,a) - f(z,a)| \le L_f |y - z|.$$

Назовём измеримое отображение $\alpha\colon [0,T] \to A$ допустимым контролем, если

$$|f(0,\alpha(t))| \in L^1[0,T].$$

В случае ограниченной функции f всякое измеримое отображение α является допустимым контролем. Заметим, что имеет место неравенство

$$|f(y,a)| \le L_f|y| + |f(0,a)|.$$

Пусть $0 \le t < T$ и α — допустимый контроль. Рассмотрим задачу Коши

$$\dot{y}_x(s) = f(y_x(s), \alpha(s)), \quad y_x(t) = x, \quad s \in [t,T], x \in \mathbb{R}^d. \tag{1}$$

Решением y_x называется абсолютно непрерывное отображение $[t,T] \to \mathbb{R}^d$, для которого верно равенство

$$y_x(s) = x + \int_t^s f(y_x(\tau), \alpha(\tau))\, d\tau.$$

Существование и единственность решения можно обосновать с помощью теоремы о сжимающем отображении. Рассмотрим в $L^1[t, t+\Delta t]$ отображение

$$F(z)(s) = x + \int_t^s f(z(\tau), \alpha(\tau))\, d\tau.$$

Так как

$$\int_t^{t+\Delta t} \Big| F(z_1)(s) - F(z_2(s)) \Big|\, ds \le L_f |\Delta t| \int_t^{t+\Delta t} \Big| z_1(s) - z_2(s) \Big|\, ds,$$

то при $|\Delta t| L_f < 1$ отображение F сжимающее и у него существует единственная неподвижная точка. Применяя это утверждение к отрезкам $[t, t+\Delta t/2]$, $[t+\Delta t/2, t+\Delta t], \ldots$, получаем существование на отрезке $[0,T]$. Обоснуем теперь единственность. Пусть y_x^1 и y_x^2 — два решение. Множество $E = \{t \in [0,T]\colon\ y_x^1(t) = y_x^2(t)\}$ замкнуто из-за непрерывности y_x^1 и y_x^2. С другой стороны для всякой точки t_0, в которой $y_x^1(t_0) = y_x^2(t_0)$, по доказанному выше существует окрестность, в которой $y_x^1 = y_x^2$. Множество E одновременно открыто и замкнуто. Следовательно, $E = [0,T]$.

Заметим, что для двух решений y_x и y_z уравнения с одним и тем же управлением α верна оценка

$$|y_x(s) - y_z(s)| \le |x - z| + L_f \int_t^s |y_x(\tau) - y_z(\tau)|\, d\tau,$$

из которой следует неравенство

$$\sup_{s \in [0,T]} |y_x(s) - y_z(s)| \le |x - z| e^{L_f T}.$$

Пусть $l\colon \mathbb{R}^d \times A \times [0,T] \to \mathbb{R}$ — непрерывная и ограниченная снизу функция, причем

$$|l(y,a,s) - l(z,a,s)| \le L_l |y - z|.$$

Пусть также $g\colon \mathbb{R}^d \to \mathbb{R}$ — ограниченная липшицева функция. Рассмотрим задачу оптимального контроля, которая состоит в минимизации функционала

$$\alpha \to \int_t^T l(y_x(s), \alpha(s), s)\, ds + g(y_x(T)).$$

Принципиальную роль в решение этой задачи играет функция значения

$$u(x,t) = \inf_\alpha \Big\{ \int_t^T l(y_x(s), \alpha(s), s)\, ds + g(y_x(T)) \Big\}.$$

Предложение 1. (i) *Существует такое число $C>0$, что*

$$|u(x,t)-u(z,t)| \le C|x-z|$$

для всех $x,z\in\mathbb{R}^d$, $t\in[0,T]$.

(ii) *Если f и l ограничены, то существует такое число $C>0$, что*

$$|u(x,t)-u(z,s)| \le C\big(|x-z|+|t-s|\big)$$

для всех $x,z\in\mathbb{R}^d$, $t,s\in[0,T]$. Более того, $u(x,T-0)=g(x)$.

Доказательство. Обоснуем первую оценку. Для всякого $\varepsilon>0$ найдем α, при котором

$$-u(z,t)\le \varepsilon-\int_t^T l(y_z(s),\alpha(s),s)\,ds-g(y_z(T)).$$

Имеем

$$u(x,t)-u(z,t)\le\varepsilon+\int_t^T\Big|l(y_x(s),\alpha(s),s)-l(y_z(s),\alpha(s),s)\Big|\,ds+|g(y_x(T)-g(y_z(T))|.$$

Правая часть оценивается выражением

$$(TL_l+L_g)e^{L_fT}|x-z|,$$

где L_g — константа Липшица функции g.

Пусть теперь f и l ограничены. Получим вторую оценку. Для всякого $\varepsilon>0$ найдем $\alpha\colon[s,T]\to A$, при котором

$$-u(x,s)\le\varepsilon-\int_s^T l(y_x^1(\tau),\alpha(\tau),\tau)\,d\tau-g(y_x^1(T)).$$

Если $t<s$, то продолжаем α на $[t,s)$ произвольным значением из A. Пусть y_x^2 — решение уравнения $\dot y_x^2(\tau)=f(y_x^2(\tau),\alpha(\tau))$ с начальным условием $y_x^2(t)=x$. Тогда

$$u(x,t)-u(x,s)\le\varepsilon+\int_t^T l(y_x^2(\tau),\alpha(\tau),\tau)\,d\tau+g(y_x^2(T))-\int_s^T l(y_x^1(\tau),\alpha(\tau),\tau)\,d\tau-g(y_x^1(T)).$$

Пусть $|l|\le M_l$ и $|f|\le M_f$. Предположим, что $s<t$. Тогда правая часть оценивается выражением

$$\varepsilon+M_l(t-s)+L_l\int_t^T|y_x^2(\tau)-y_x^1(\tau)|\,d\tau+L_g|y_x^2(T)-y_x^1(T)|.$$

Заметим, что при $\tau\in[t,T]$ верно неравенство

$$|y_x^2(\tau)-y_x^1(\tau)|\le M_f(t-s)+L_f\int_t^\tau|y_x^2(\tau)-y_x^1(\tau)|\,d\tau,$$

из которого следует оценка

$$|y_x^2(\tau)-y_x^1(\tau)|\le M_fe^{L_fT}(t-s).$$

Следовательно, приходим к неравенству

$$u(x,t)-u(x,s)\le\varepsilon+M_l(t-s)+TM_fe^{L_fT}(t-s)+L_gM_fe^{L_fT}(t-s).$$

Устремляя ε к нулю получаем

$$u(x,t)-u(x,s)\le C|t-s|.$$

Случай $t<s$ исследуется полностью аналогично. □

В общем случае, функция значения u не является дифференцируемой.

Принцип динамического программирования

Теорема 1. *Для всех $0\le t\le\tau\le T$ имеет место равенство*

$$u(x,t)=\inf_\alpha\Big\{\int_t^\tau l(y_x(s),\alpha(s),s)\,ds+u(y_x(\tau),\tau)\Big\}.$$

Доказательство. Обозначим правую часть доказываемого равенства через $U(t,x)$. Для всякого $\varepsilon > 0$ найдется управление α, при котором

$$u(x,t) + \varepsilon \geq \int_t^T l(y_x(s), \alpha(s), s)\, ds + g(y_x(T)).$$

Правую часть можно записать в виде

$$\int_t^\tau l(y_x(s), \alpha(s), s)\, ds + \int_\tau^T l(y_x(s), \alpha(s), s)\, ds + g(y_x(T))$$

и оценить снизу следующим образом:

$$\int_t^\tau l(y_x(s), \alpha(s), s)\, ds + u(y_x(\tau), \tau) \geq U(t,x).$$

В силу произвольности ε получаем неравенство $u(x,t) \geq U(x,t)$. Установим противоположное неравенство. Для всякого $\varepsilon > 0$ найдутся такие α_1 на $[t,\tau]$ и α_2 на $[\tau, T]$, что

$$U(x,t) + \varepsilon \geq \int_t^\tau l(y_x^1(s), \alpha_1(s), s)\, ds + u(y_x^1(\tau), \tau)$$

и

$$u(y_x^1(\tau), \tau) + \varepsilon \geq \int_\tau^T l(y_z^2(s), \alpha_2(s), s)\, ds + g(y_z^2(T)), \quad z = y_x^1(\tau).$$

Положим $\alpha(s) = \alpha_1(s)$ на $[t,\tau)$ и $\alpha(s) = \alpha_2(s)$ на $[\tau, T]$. Пусть y_x — решение задачи Коши с таким α. Тогда $y_x(s) = y_x^1(s)$ на $[t,\tau]$ и $y_x(s) = y_x^2(s)$ на $[\tau, T]$. Имеем

$$U(x,t) + 2\varepsilon \geq \int_t^T l(y_x(s), \alpha(s), s)\, ds + g(y_x(T)) \geq u(x,t).$$

Устремляя ε к нулю получаем $U(x,t) \geq u(x,t)$. □

Заметим, что для всякого контроля α функция

$$\tau \to \int_t^\tau l(y_x(s), \alpha(s), s)\, ds + u(y_x(\tau), \tau)$$

монотонно не убывает.

Предложение 2. *Контроль α является оптимальным тогда и только тогда, когда функция*

$$\tau \to \int_t^\tau l(y_x(s), \alpha(s), s)\, ds + u(y_x(\tau), \tau)$$

является константой на $[t, T]$.

Доказательство. Если рассматриваемая функция является константой, то ее значения при $\tau = t$ и $\tau = T$ равны и

$$u(x,t) = \int_t^T l(y_x(s), \alpha(s), s)\, ds + g(y_x(T)).$$

Пусть теперь α — оптимальное управление на $[t, T]$. Тогда из монотонности по τ следует неравенство

$$\int_t^\tau l(y_x(s), \alpha(s), s)\, ds + u(y_x(\tau), \tau) \leq u(x,t),$$

а из принципа динамического программирования следует неравенство

$$u(x,t) \leq \int_t^\tau l(y_x(s), \alpha(s), s)\, ds + u(y_x(\tau), \tau).$$

Следовательно, при каждом τ

$$u(x,t) = \int_t^\tau l(y_x(s), \alpha(s), s)\, ds + u(y_x(\tau), \tau).$$

□

Уравнение Гамильтона–Якоби–Беллмана

Предположим, что функция u непрерывно дифференцируема по x и t. Пусть $\Delta t > 0$. Принцип динамического программирования можно записать в виде

$$\inf_{\alpha}\Big\{\int_t^{t+\Delta t} l(y_x(s), \alpha(s), s)\, ds + u(y_x(t+\Delta t), t+\Delta t) - u(x,t)\Big\} = 0.$$

Поделим это равенство на Δt и устремим Δt к нулю. Получаем уравнение Гамильтона–Якоби–Беллмана

$$-u_t(x,t) + H(x,t,\nabla u(x,t)) = 0,$$

где

$$H(x,t,p) = \sup_{a\in A}\Big\{-l(x,a,t) - \langle p, f(x,a)\rangle\Big\}.$$

Таким образом, функция значения u является решением задачи Коши

$$-u_t(x,t) + H(x,t,\nabla u(x,t)) = 0, \quad u(x,T) = g(x). \tag{2}$$

Теорема 2. *Если непрерывно дифференцируемая функция v является решением задачи Коши (2). Тогда*

(i) $v(x,t) \le u(x,t)$*;*

(ii) *если для контроля α выполняется*

$$-l(y_x(s), \alpha(s), s) - \langle \nabla v(y_x(s)), f(y_x(s), \alpha(s))\rangle = H(y_x(s), s, \nabla v(y_x(s))),$$

то α является оптимальным контролем и $v = u$.

Доказательство. Пусть α — какой-либо допустимый контроль. Имеем

$$\frac{d}{d\tau} v(y_x(\tau), \tau) = \langle \nabla v(y_x(\tau), \tau), f(y_x(\tau), \alpha(\tau))\rangle + v_t(y_x(\tau), \tau).$$

Заметим, что

$$\langle \nabla v(y_x(s), s), f(y_x(s), \alpha(s))\rangle \le -l(y_x(s), \alpha(s), s) - H(y_x(s), s, \nabla v(y_x(s), s)).$$

Так как v является решением уравнения Гамильтона–Якоби–Беллмана, то

$$\frac{d}{d\tau} v(y_x(tau), \tau) \le l(y_x(s), \alpha(s), s).$$

Интегрируя это неравенство по s от t до T получаем

$$v(x,t) \le \int_t^T l(y_x(s), \alpha(s), s)\, ds + g(y_x(T)).$$

В силу произвольности α приходим к оценке $v(x,t) \le u(x,t)$. Если α удовлеиворяет условию (ii), то получим равенства

$$v(x,t) = \int_t^T l(y_x(s), \alpha(s), s)\, ds + g(y_x(T)), \quad v(x,t) = u(x,t).$$

$\square$

Предположим, что A – компактное множество в $\mathbb{R}^d$ и для всех x, t, p существует единственное $a(x,t,p)$, при котором

$$H(x,t,p) = -l(x,a,t) - \langle p, f(x,a)\rangle.$$

Тогда H дифференцируемо по p и

$$H_p(x,t,p) = -f(x, a(x,t,p)).$$

Это наблюдение является частным случаем следующего более общего утверждения.

Предложение 3. *Пусть A — компактное метрическое пространство, Q — открытое множество в $\mathbb{R}^d$ и $F(x,a)$ — непрерывная функция на $A \times Q$, причем существует и непрерывна производная F_x. Если a_z является единственной точкой максимума функции $a \to F(z,a)$, то функция*

$$G(x) = \sup_{a \in A} F(x,a)$$

дифференцируема в точке z и $G_x(z) = F_x(z, a_z)$.

Доказательство. Пусть a_x — какая-либо точка максимума функции $a \to F(x,a)$. Из единственности a_z и компактности A следует, что $a_x \to a_z$, если $x \to z$. Имеем

$$G(z) - G(x) = F(z,a_z) - F(x,a_x) \le F(z,a_z) - F(x,a_z) = F_x(\xi, a_z)(z-x),$$
$$G(x) - G(z) \le F_x(\eta, a_x)(x-z),$$

где $\xi, \eta \in [x,z]$. Следовательно,

$$|G(x) - G(z) - F_x(z,a_z)(x-z)| \le$$
$$\le |F_x(\xi,a_z) - F_x(z,a_z)||x-z| + |F_x(\eta,a_x) - F_x(z,a_z)||x-z| = o(|x-z|).$$

□

Итак, если α — оптимальный контроль, то

$$f(y_x(s), \alpha(s)) = -H_p(y_x(s), s, \nabla v(y_x(s), s)).$$

Следовательно, оптимальное решение y_x является решением задачи Коши

$$\dot{y}_x(s) = -H_p(y_x(s), s, \nabla v(y_x(s), s)), \quad y_x(t) = x,$$

и может быть найдено без предварительного вычисления α.