Управление обратной связи в некооперативных играх

Явный недостаток постановки задач в программных управлениях $u_i(t)$ в том, что при заданных функционалах, динамике и моменте окончания выбор управления базируется только на знании начального положения $x(t_0)=x_0$. Если по какой-то причине в некоторый момент $t_1>t_0$ положение системы будет отличаться от $x^*(t_1)$ (которое должно быть в силу выбранных в момент t_0 программных управлений), то уже ничего нельзя сказать о равновесном результате в момент окончания T. Например, пусть в случае двух игроков первый придерживался своего управления $u_1^*(t)$, а второй отклонился от $u_2^*(t)$. По определению равновесия по Нэшу, если первый будет продолжать придерживаться до момента T управления $u_1^*(t)$, то значение функционала J_2 второго игрока может быть хуже равновесного значения J_2^*. Но ничего нельзя сказать о функционале для первого игрока.

(1) В связи с этим вводят такие понятия. Пусть в момент t_0 выбраны управления $u_i^*(0,T)$, $i=\overline{1,N}$.

а) Набор $u_i^*(0,T)$ называется слабо согласованным по времени (weakly time consistent), если продолжение $u_i^*(t_1,T)$ с момента t_1 выбранных управлений даёт ситуацию равновесия для подигры $\Gamma(t_1, x^*(t_1))$, где $x^*(t_1) = x(0, t_1, u_1^*(0,t_1), \dots, u_N^*(0,t_1))$ для любого $t_1 \in (0, T)$.

в) Набор $u_i^*(0,T)$ называется сильно согласованным (strongly time consistent), если продолжение $u_i^*(t_1,T)$ с момента t_1 выбранных управлений даёт ситуацию равновесия для подигры $\Gamma(t_1, x_{t_1})$, где $x_{t_1} \in R^n$ – произвольное состояние, которое достижимо из некоторого начального состояния в момент t_0 (т. е. даже отличного от x_0).

В этих определениях не уточняется, единственно или нет ситуация равновесия (особенно во втором определении). Для наглядности можно рассматривать случай единственности равновесия для любого t_1 и любого $x(t_1)$.

Ясно, что open-loop равновесие по Нэшу является weakly time consistent.

Но выполнение strongly time consistent в программных управлениях – большая редкость.

На стр. 361 приведён пример, в котором strongly time consistent отсутствует.

② Теперь вводятся управления обратной связи $u_i(t) = f_i(t, x(t))$. Вначале общего вида (потом f_i будет линейная функция от $x(t)$). Определение 8.3 определяет набор $u_i^*(t) = f_i^*(t, x(t))$ feedback Nash equilibrium, если эти стратегии обеспечивают ситуацию равновесия по Нэшу для подигры $\Gamma(t_1, x_{t_1})$ для всех $t_1 \in [0, T)$ и для всех $x_{t_1} \in R^n$. Подчеркивается обратная связь Марковского типа, когда выбираемое в момент t управление зависит только от состояния $x(t)$ в этот момент, но не зависит от состояний, предшествующих моменту t. Уязвимость в том, что надо говорить о решении уравнения системы, когда игроки используют функции $f_i(t, x(t))$. Это сильно сужает класс функций f_i, либо надо говорить о решении как о пределе движений в силу дискретных схем.

Теорема 8.2. при предположении о гладкости функции $V_i(t, x) = J_i^*(t, x)$ формулирует необходимое и достаточное условие существования стратегии feedback Nash equilibrium. В сторону необходимость: пусть функции $u_i^*(\cdot)$, $i=1,2$ обеспечивают feedback Nash equilibrium. Предположим, что функции $V_i(t,x)$ дифференцируемы по t и по x, причем по x непрерывно дифференцируемы, и существует полная производная $\frac{d}{dt}V_i(t,x(t))$. Тогда функции V_i удовлетворяют уравнениям в частных производных (8.2.2), (8.2.4) с краевыми условиями (8.2.3) и (8.2.5). [В формулировке не очень понятно, как задаются $V_i(t,x)$ для произвольных x, если начальное условие $x(0)$ четко оговорено.]

?

В сторону достаточности: пусть функции V_i удовлетворяют уравнениям (8.2.2) и (8.2.4) с краевыми условиями (8.2.3) и (8.2.5), где V_i – гладкие функции. Пусть управления u_1^*, u_2^* задаются через условие минимума (стр. 363). Тогда $u_i^*(\cdot)$ обеспечивают feedback Nash equilibrium. Главное уязвимость – предположение о гладкости.

!

[При чтении текста о теореме 8.2 надо смотреть стр. 362 и 363 книги.]

3. Теоремой 8.2 об уравнениях в частных производных исчерпывается материал об управлениях обратной связи общего вида. Далее рассматриваются управления $u_i^*(t) = F_i^*(t)x(t)$ линейные по фазовой переменной.

Со ссылкой на теорему 5.1 (глава 5 о задачах управления) доказывается на стр. 364–365 следующая теорема 8.3.

Теорема 8.3. Линейно-квадратичная дифференциальная игра двух лиц имеет для каждого начального состояния a linear feedback Nash equilibrium тогда и только тогда, когда следующая пара связанных матричных уравнений Риккати имеет симметричные решения K_1, K_2 на $[0,T]$:

$$\dot{K}_1(t) = -(A - S_2K_2(t))^T K_1(t) - K_1(t)(A - S_2K_2(t)) + K_1(t)S_1K_1(t) - Q_1 - K_2(t)S_{21}K_2(t), \tag{8.2.6}$$

$$K_1(t) = Q_{1T}$$

$$\dot{K}_2(t) = -(A - S_1K_1(t))^T K_2(t) - K_2(t)(A - S_1K_1(t)) + K_2(t)SK_2(t) - Q_2 - K_1(t)S_{12}K_1(t), \tag{8.2.7}$$

$$K_2(t) = Q_{2T}.$$

Более того, точка равновесия является единственной и соответствует управлениям

$$u_i^*(t) = -R_i^{-1}B_i^T K_i(t)x(t), \quad i = 1.2$$

Значение платы имеет вид $x_0^T K_i(0)x_0$, $i = 1,2$.

Как уже сказано, доказательство теоремы опирается на теорему 5.1. Следует разобрать это доказательство

Д.з. самим.

④ Очень интересным является вопрос о соотношении решений в классе управлений обратной связи (линейных по x) и в классе программных управлений. На стр. 366–367 рассмотрен пример, в котором терминальный член в функционале платы

$$J_1(u_1,u_2)=\int_0^T\{u_1^2(t)-u_2^2(t)\}dt+\alpha x^2(T)$$

$$J_2(u_1,u_2)=-J_1(u_1,u_2)$$

зависит от параметра α. Динамика имеет вид

$$\dot x(t)=\sqrt2\,u_1(t)-u_2(t),\quad x(0)=x_0.$$

Показано, что при $\alpha>0$ игра в рамках линейных обратных связей имеет ситуацию равновесия для любого T. При $\alpha<0$, наоборот, при любом T не имеет.

В рамках программных управлений игра не имеет ситуации равновесия при любом T и при $\alpha>0$ и при $\alpha<0$.

Д.з. Разобрать пример 8.2.

(5) Теорема 8.4. формулирует необходимые и достаточные условия для антагонистических линейно-квадратичных игр в случае линейных управлений (стр. 369–371).

Теорема 8.4. Динамика

$$\dot{x}(t) = Ax(t) + B_1 u_1(t) + B_2 u_2(t), \quad x(0) = x_0. \quad (8.2.14)$$

Функционал платы для первого игрока

$$J_1(u_1, u_2) = \int_0^T \{x^T(t) Q x(t) + u_1^T(t) R_1 u_1(t) - u_2^T(t) R_2 u_2(t)\} dt + x^T(T) Q_T x(T)$$

$$J_2(u_1, u_2) = -J_1(u_1, u_2).$$

Матрицы Q и Q_T, R_i симметричны, R_i - положительно-определённые.

Эта игра имеет linear feedback Nash equilibrium для каждого начального состояния тогда и только тогда, когда уравнение Риккати

$$\dot{K}(t) = -A^T K(t) - K(t)A + K(t)(S_1 - S_2)K(t) - Q, \quad (8.2.15)$$
$$K(T) = Q_T$$

имеет симметричное решение на $[0, T]$.

Более того, если (8.2.15) имеет решение, то игра имеет единственное равновесие. Оно доставляется управлениями

$$u_1^*(t) = -R_1^{-1} B_1^T K(t) x(t), \quad u_2^*(t) = -R_2^{-1} B_2^T K(t) x(t).$$

Равновесный результат есть $x_0^T K(0) x_0$ и $-x_0^T K(0) x_0$.

Д.з. Разобрать доказательство теоремы на стр. 370.