نظریه بازیها
Game Theory

ارائه کننده: امیرحسین نیکوفرد
مهندسی برق و کامپیوتر دانشگاه خواجه نصیر

InfiniteDynamic Games

Material

- Dynamic Non-cooperative Game Theory: Second Edition
  - Chapter5: Sections 5:5 and Chapter6: Sections 6:2
- An Introductory Course in Non-cooperative Game Theory
  - Chapter 18

InfiniteDynamic Games

- Zero sum games
- Non-zero sum games
- Infinite Games
- **Infinite Dynamic Games**
  - Dynamic games in discrete time
  - Information structures
  - Continuous-time differential games
  - Discrete-time dynamic programming
  - Continuous-time dynamic programming
  - Discrete-time dynamic programming for zero sum games
  - **Continuous time dynamic programming for zero sum games**

Zero-sum dynamic games in continuous time

We now discuss the solution for two-player zero-sum dynamic games in continuous time, which corresponds to dynamics of the form

$$\underbrace{\dot{x}(t)}_{\substack{state\\derivative}} = \underbrace{f}_{\substack{game\\dynamics}}(\underbrace{t}_{time}, \underbrace{x(t)}_{\substack{current\\state}}, \underbrace{u(t)}_{\substack{P_1's\ action\\at\ time\ t}}, \underbrace{d(t)}_{\substack{P_2's\ action\\at\ time\ t}}), \qquad \forall\, t \in [0, T] \qquad (1)$$

with state $x(t) \in R^n$ initialized at a given $x(0) = x_0$. For every time $t \in [0, T]$, the action *u(t)* is required to belong to a given action space *U* and P_2's action *d(t)* is required to belong to an action space *D*. We assume a finite horizon $(T < \infty)$ integral cost of the form

$$J = \int_0^T \underbrace{g(t, x(t), u(t), d(t))dt}_{cost\ along\ trajctory} + \underbrace{q(x(T))}_{final\ cost} \qquad (2)$$

Zero-sum dynamic games in continuous time

that P_1 wants to minimize and P_2 wants to maximize. In this part we consider a **state-feedback information structure**, which correspond to policies of the form

$$u(t) = \gamma(\mathrm{t}, x(t)), \quad , d(t) = \sigma(\mathrm{t}, x(t)), \quad \forall\, \mathrm{t} \in [0, \mathrm{T}],$$

For continuous-time we can also use dynamic programming to construct saddle-point equilibria in state-feedback policies. The following result is the equivalent of Theorem about zero-sum dynamic games in discrete time for continuous time.

Zero-sum dynamic games in continuous time

Theorem 17.1. Assume that there exists a continuously differentiable function $V(t, x)$ that satisfies the following Hamilton-Jacobi-Bellman-Isaac equation

$$-\frac{\partial V(t,x)}{\partial t} = \min_{u \in U} \sup_{d \in D} \left(g(t,x,u,\mathrm{d}) + \frac{\partial V(t,x)}{\partial x} f(t,x,u,d) \right) \qquad (3)$$

$$= \max_{d \in D} \inf_{u \in U} \left(g(t,x,u,\mathrm{d}) + \frac{\partial V(t,x)}{\partial x} f(t,x,u,d) \right), \forall t \in [0,\mathrm{T}], \mathrm{x} \in \mathrm{R}^n$$

with

$$V(\mathrm{T},x) = q(x), \qquad \forall \mathrm{x} \in \mathrm{R}^n \qquad (4)$$

Zero-sum dynamic games in continuous time

Then the pair of policies (γ^*, σ^*)defined as follows is a saddle-point equilibrium in state-feedback policies:

$$\gamma^*(t,x) = \arg\min_{u\in U} \sup_{d\in D}\left(g(t,x,u,\mathrm{d}) + \frac{\partial V(t,x)}{\partial x} f(t,x,u,d)\right)$$

$$\sigma^*(t,x) = \arg\max_{d\in D} \inf_{u\in U}\left(g(t,x,u,\mathrm{d}) + \frac{\partial V(t,x)}{\partial x} f(t,x,u,d)\right)$$

$\forall t \in [0,\mathrm{T}], \mathrm{x} \in \mathrm{R}^n$ Moreover, the value of the game is equal to $V(0, x_0)$.

NOTE: Theorem 17.1 provides a sufficient condition for the existence of Nash equilibria, but this condition is not necessary. In particular, two security levels may not commute for some state x at some stage t, but there may still be a saddle-point for the game.

Zero-sum dynamic games in continuous time

Proof of Theorem 17.1. From the fact that the inf and sup commute in (3) and the definitions of $\gamma^*(t,x)$ *and* $\sigma^*(t,x)$, we conclude that the pair $(\gamma^*(t,x),\sigma^*(t,x))$ is a saddle-point equilibrium for a zero-sum game with criterion

$$g(t,x,u,\mathrm{d})+\frac{\partial V(t,x)}{\partial x}f(t,x,u,d)$$

which means that

$$g(t,x,\gamma^*(t,x),\mathrm{d})+\frac{\partial V(t,x)}{\partial x}f(t,x,\gamma^*(t,x),d) \quad \le$$

$$g(t,x,\gamma^*(t,x),\sigma^*(t,x))+\frac{\partial V(t,x)}{\partial x}f(t,x,\gamma^*(t,x),\sigma^*(t,x)) \quad \le$$

$$g(t,x,u,\sigma^*(t,x))+\frac{\partial V(t,x)}{\partial x}f(t,x,u,\sigma^*(t,x))$$

Zero-sum dynamic games in continuous time

Moreover, since the middle term in these inequalities is also equal to the right-hand-side of (3), we have that

$$-\frac{\partial V(t,x)}{\partial t} = g(t,x,\gamma^*,\sigma^*) + \frac{\partial V(t,x)}{\partial x} f(t,x,\gamma^*,\sigma^*), \mathrm{x} \in \mathrm{R}^n$$

$$= \sup_{d \in D}(g(t,x,\gamma^*(t,x),\mathrm{d}) + \frac{\partial V(t,x)}{\partial x} f(t,x,\gamma^*(t,x),d)), \forall t \in [0,\mathrm{T}]$$

which, because of Theorem Continuous-time dynamic programming in lecture 15, shows that $\sigma^*(t,x)$ is an optimal (maximizing) state-feedback policy against $\gamma^*(t,x)$ and the maximum is equal to $V(0, x_0)$.

Zero-sum dynamic games in continuous time

Moreover, since we also have that

$$-\frac{\partial V(t,x)}{\partial t} = g(t,x,\gamma^*,\sigma^*) + \frac{\partial V(t,x)}{\partial x} f(t,x,\gamma^*,\sigma^*), \mathrm{x} \in \mathrm{R}^n$$

$$= \inf_{u \in U} (g(t,x,u,\sigma^*(t,x)) + \frac{\partial V(t,x)}{\partial x} f(t,x,u,\sigma^*(t,x))), \forall t \in [0,\mathrm{T}]$$

we can also conclude that $\gamma^*(t,x)$ is an optimal (minimizing) state-feedback policy against $\sigma^*(t,x)$ and the minimum is also equal to $\mathrm{V}(0,\mathrm{x}_0)$. This proves that (γ^*,σ^*) is indeed a saddle-point equilibrium in state-feedback policies with value $\mathrm{V}(0, \mathrm{x}_0)$.

Zero-sum dynamic games in continuous time

Note: we actually conclude that

1.P_2 cannot get a reward larger than $V(0, x_0)$ against $\gamma^*(t,x)$, regardless of the information structure available to P_2, and

2.P_1 cannot get a cost smaller than $V(0, x_0)$ against $\sigma^*(t,x)$, regardless of the information structure available to P_1.

In practice, this means that $\gamma^*(t,x)$ and $\sigma^*(t,x)$ are "extremely safe" policies for P_1 and P_2, respectively, since they guarantee a level of reward regardless of the information structure for the other player.

InfiniteDynamic Games

- Zero sum games
- Non-zero sum games
- Infinite Games
- **Infinite Dynamic Games**
  - Dynamic games in discrete time
  - Information structures
  - Continuous-time differential games
  - Discrete-time dynamic programming
  - Continuous-time dynamic programming
  - Discrete-time dynamic programming for zero sum games
  - **Continuous time dynamic programming for zero sum games**
    - **Linear quadratic dynamic games**
    - **Differential games with variable termination time**

Linear quadratic dynamic games

Continuous-time linear quadratic games are characterized by linear dynamics of the form

$$\dot{x}(t) = \underbrace{Ax(t) + Bu(t) + Ed(\mathrm{t})}_{f(\mathrm{t},x(t),\mathrm{u(t)},\mathrm{d(t)})}, \qquad x \in R^n, u \in R^{n_u}, d \in R^{n_d}, \mathrm{t} \in [0, \mathrm{T}]$$

and an integral quadratic cost of the form

$$J := \int_0^T \underbrace{(\|y(t)\|^2 + \|u(t)\|^2 - \mu^2 \|d(t)\|^2)}_{g(t,x(t),u(t),d(t))} dt \quad + \underbrace{x'(T)\mathrm{P}_T\, x(T)}_{q(x(T))}$$

where

$$y(t) = Cx(t), \qquad \forall t \in [0, \mathrm{T}]$$

Linear quadratic dynamic games

This cost function captures scenarios in which

1) player P_1 wants to make *y(t)* small over the interval $[0, T]$ without "spending" much effort in its action *u(t)*,

2) whereas player P_2 wants to make *y(t)* large without "spending" much effort in its action *d(t)*.

The constant μ can be seen as a conversion factor that maps units of *d(t)* into units of *u(t)* and *y(t)*

NOTE: If needed, a "conversion factor" between units of *u* and y could be incorporated into the matrix C that defines y.

Linear quadratic dynamic games

The Hamilton-Jacobi-Bellman-Isaac equation for this game is

$$-\frac{\partial V(t,x)}{\partial t} = \min_{u\in U}\sup_{d\in D}(\mathrm{x'C'Cx} + u'u - \mu^2 d'd + \frac{\partial V(t,x)}{\partial x}(Ax + Bu + Ed))$$

$$= \max_{d\in D}\inf_{u\in U}(\mathrm{x'C'Cx} + u'u - \mu^2 d'd + \frac{\partial V(t,x)}{\partial x}(Ax + Bu + Ed))$$

$\forall t \in [0, \mathrm{T}], \mathrm{x} \in \mathrm{R}^n$,with

$$V(T,x) = x'(T)\mathrm{P}_T\, x(T), \qquad \forall \mathrm{x} \in \mathrm{R}^n$$

Inspired by the boundary condition , we will try to find a solution to the Hamilton-Jacobi-Bellman-Isaac equation of the form

$$V(t,x) = x'\mathrm{P(t)}x, \qquad \forall \mathrm{x} \in \mathrm{R}^n, \forall t \in [0, \mathrm{T}]$$

Linear quadratic dynamic games

for some appropriately selected symmetric *n* x *n* matrix *P(t)*. For boundary condition to hold, we need to have $P(T) = P_T$. For the Hamilton-Jacobi-Bellman-Isaac equation to hold, we need

$$-x'\dot{P}(t)x = \min_{u\in U}\sup_{d\in D}(x'C'Cx + u'u - \mu^2 d'd + 2x'P(t)(Ax + Bu + Ed))$$

$$= \max_{d\in D}\inf_{u\in U}(x'C'Cx + u'u - \mu^2 d'd + 2x'P(t)(Ax + Bu + Ed)) \quad (5)$$

$\forall t \in [0, T], x \in R^n$, Since the functions to optimize are quadratic, to compute the inner supremum and infimum in (5), we simply need to make the appropriate gradients equal to zero:

$$\frac{\partial}{\partial d}(x'C'Cx + u'u - \mu^2 d'd + 2x'P(t)(Ax + Bu + Ed)) = 0$$

$$\Leftrightarrow -2\mu^2 d' + 2x'PE = 0 \Leftrightarrow d = \mu^{-2} E'Px$$

Linear quadratic dynamic games

$$\frac{\partial}{\partial u}(\mathrm{x}'\mathrm{C}'\mathrm{Cx}+u'u-\mu^2 d'd+2x'\mathrm{P(t)}(Ax+Bu+Ed))=0$$

$$\Leftrightarrow 2u'+2x'\mathrm{P}B=0 \Leftrightarrow u=-B'\mathrm{Px}$$

Therefore

$$\sup_{d\in D}\underbrace{(\mathrm{x}'\mathrm{C}'\mathrm{Cx}+u'u-\mu^2 d'd+2x'\mathrm{P(t)}(Ax+Bu+Ed))}_{\mathrm{d}=\mu^{-2}\mathrm{E}'\mathrm{Px}}$$

$$=x'(\mathrm{PA}+A'P+\mathrm{C}'\mathrm{C}+\mu^{-2}\mathrm{P}E\mathrm{E}'\mathrm{P})\mathrm{x}+u'u+2x'\mathrm{P}Bu$$

$$\inf_{u\in U}\underbrace{(\mathrm{x}'\mathrm{C}'\mathrm{Cx}+u'u-\mu^2 d'd+2x'\mathrm{P(t)}(Ax+Bu+Ed))}_{u=-B'\mathrm{Px}}$$

$$=x'(\mathrm{PA}+A'P+\mathrm{C}'\mathrm{C}-\mathrm{P}BB'\mathrm{P})\mathrm{x}-\mu^2 d'd+2x'\mathrm{P}Ed.$$

Linear quadratic dynamic games

This means that (5) is of the form

$$-x'\dot{\mathrm{P}}(\mathrm{t})x = \min_{u\in U}\left(x'(\mathrm{PA}+A'P+\mathrm{C'C}+\mu^{-2}\,\mathrm{P}\,E\,\mathrm{E'P})\mathrm{x}+u'u+2x'\,\mathrm{P}\,Bu\right)$$

$$= \max_{d\in D}\left(x'(\mathrm{PA}+A'P+\mathrm{C'C}-\mathrm{P}\,BB'\,\mathrm{P})\mathrm{x}-\mu^{2}d'd+2x'\,\mathrm{P}\,Ed\right)$$

Once again we have quadratic functions to optimize so all we need to do is to make their gradients equal to zero:

$$\frac{\partial}{\partial u}\left(x'(\mathrm{PA}+A'P+\mathrm{C'C}+\mu^{-2}\,\mathrm{P}\,E\,\mathrm{E'P})\mathrm{x}+u'u+2x'\,\mathrm{P}\,Bu\right)=0 \Leftrightarrow u=-B'\,\mathrm{Px}$$

$$\frac{\partial}{\partial d}\left(x'(\mathrm{PA}+A'P+\mathrm{C'C}-\mathrm{P}\,BB'\,\mathrm{P})\mathrm{x}-\mu^{2}d'd+2x'\,\mathrm{P}\,Ed\right)=0 \Leftrightarrow \mathrm{d}=\mu^{-2}\,\mathrm{E'Px}$$

Linear quadratic dynamic games

Therefore

$$\min_{u\in U}(\underbrace{x'(PA+A'P+C'C+\mu^{-2}PEE'P)x+u'u+2x'PBu}_{u=-B'Px})=0$$

$$=x'(PA+A'P+C'C+\mu^{-2}PEE'P-PBB'P)x$$

$$\max_{d\in D}(\underbrace{x'(PA+A'P+C'C-PBB'P)x-\mu^{2}d'd+2x'PEd}_{d=\mu^{-2}E'Px})=0$$

$$=x'(PA+A'P+C'C+\mu^{-2}PEE'P-PBB'P)x$$

Therefore the inf and sup commute and (5) simply becomes

$$-x'\dot{P}(t)x=x'(PA+A'P+C'C+\mu^{-2}PEE'P-PBB'P)x$$

Linear quadratic dynamic games

which holds provided that

$$-\dot{\mathrm{P}}(\mathrm{t}) = \mathrm{PA} + A'P + \mathrm{C'C} + \mu^{-2}\,\mathrm{P}\,E\,\mathrm{E'P} - \mathrm{P}\,BB'\,\mathrm{P}, \quad \forall t \in [0, \mathrm{T}]$$

The following then follows from Theorem 17.1:

Corollary 17.1. Suppose that there is a symmetric solution to the following matrix-valued ordinary differential equation

$$-\dot{\mathrm{P}}(\mathrm{t}) = \mathrm{PA} + A'P + \mathrm{C'C} + \mu^{-2}\,\mathrm{P}\,E\,\mathrm{E'P} - \mathrm{P}\,BB'\,\mathrm{P}, \quad \forall t \in [0, \mathrm{T}]$$

with final condition $\mathrm{P(T)} = \mathrm{P_T}$. Then the state-feedback policies

$$\gamma^*(t,x) = -B'\,\mathrm{Px}, \qquad \sigma^*(t,x) = \mu^{-2}\,\mathrm{E'Px}, \qquad x \in R^n, \forall t \in [0, \mathrm{T}]$$

is a saddle-point equilibrium in state-feedback policies with value $x'(0)\,\mathrm{P}(0)\,x(0)$

Linear quadratic dynamic games

Note (Induced norm). Since (γ^*, σ^*) is a saddle-point equilibrium with value $x'(0)\,\mathrm{P}(0)x(0)$, when P_1 uses

$$u(t) = \gamma^*(t, x) = -B'\,\mathrm{P}\mathrm{x}$$

for every policy

$$d(t) = \sigma(t, x(t))$$

for P_2, we have that

$$J(\gamma^*, \sigma^*) = x_0'\,\mathrm{P}(0)x_0 \ge J(\gamma^*, \sigma) = \int_0^T (\|y(t)\|^2 + \|u(t)\|^2 - \mu^2 \|d(t)\|^2)dt$$

$$+ x'(T)\,\mathrm{P}_T\, x(T)$$

and therefore

$$\int_0^T \|y(t)\|^2\, dt \le x_0'\,\mathrm{P}(0)x_0 + \int_0^T \mu^2 \|d(t)\|^2\, dt - \int_0^T \|u(t)\|^2\, dt - x'(T)\,\mathrm{P}_T\, x(T)$$

Linear quadratic dynamic games

When P_T is positive semi-definite and $x_0 = 0$, this implies that

$$\int_0^T \|y(t)\|^2 \, dt \le \int_0^T \mu^2 \|d(t)\|^2 \, dt$$

Moreover, this holds for every possible *d(t)*, regardless of the information structure available to P_2, and therefore we conclude that

$$\sup_{d \in D} \frac{\sqrt{\int_0^T \|y(t)\|^2 \, dt}}{\sqrt{\int_0^T \|d(t)\|^2 \, dt}} \le \mu \qquad (6)$$

Linear quadratic dynamic games

In view of (6), the control law is said to achieve an L2-induced norm in the interval $[0, T]$ from the disturbance d to the output y lower than μ.

NOTE: When $T = \infty$, the left-hand side of (6) is called the H-infinity norm of the closed-loop and control low guarantees a H-infinity norm smaller than μ.

InfiniteDynamic Games

- Zero sum games
- Non-zero sum games
- Infinite Games
- **Infinite Dynamic Games**
  - Dynamic games in discrete time
  - Information structures
  - Continuous-time differential games
  - Discrete-time dynamic programming
  - Continuous-time dynamic programming
  - Discrete-time dynamic programming for zero sum games
  - **Continuous time dynamic programming for zero sum games**
    - Linear quadratic dynamic games
    - **Differential games with variable termination time**

Zero-sum Differential games with variable termination time

Consider now a two-player zero-sum differential game with the usual dynamics

$$\underbrace{\dot{x}(t)}_{\substack{state\\derivative}} = \underbrace{f}_{\substack{game\\dynamics}}(t, x(t), u(t), d(t)), \qquad x(t) \in R^n, u(t) \in U, d(t) \in D, \mathrm{t} \geq 0$$

and initialized at a given $x(0) = x_0$, but with an integral cost with variable horizon:

$$J = \int_0^{T_{end}} \underbrace{g(t, x(t), u(t), \mathrm{d(t)})dt}_{\text{cost along trajctory}} + \underbrace{q(T_{end}, x(T_{end}))}_{\text{final cost}}$$

- where T_{end} is the first time at which the state *x(t)* enters a closed set $\chi_{end} \subset R^n$ or $T_{end} = \infty$ in case *x(t)* never enters χ_{end} .

Zero-sum Differential games with variable termination time

Also for this game we can use dynamic programming to construct saddle-point equilibria in state-feedback policies. The following result is the equivalent of Theorem 17.1 for this game with variable termination time.

Theorem 17.2. Assume that there exists a continuously differentiable function $V(t, x)$ that satisfies the following Hamilton-Jacobi-Bellman-Isaac equation (3) with

$$V(t,x) = q(\mathrm{t},x), \qquad \forall t > 0,\, x \in \chi_{end} \qquad (7)$$

Zero-sum Differential games with variable termination time

Then the pair of policies (γ^*, σ^*) defined as follows is a saddle-point equilibrium in state-feedback policies:

$$\gamma^*(t,x) = \arg\min_{u \in U} \sup_{d \in D} (g(t,x,u,\mathrm{d}) + \frac{\partial V(t,x)}{\partial x} f(t,x,u,d))$$

$$\sigma^*(t,x) = \arg\max_{d \in D} \inf_{u \in U} (g(t,x,u,\mathrm{d}) + \frac{\partial V(t,x)}{\partial x} f(t,x,u,d))$$

$\forall t \in [0, \mathrm{T}], \mathrm{x} \in \mathrm{R}^n$ Moreover, the value of the game is equal to $V(0, x_0)$.

NOTE: We can view (7) as a boundary condition for the

Hamilton -Jacobi-Beilman-Isaac equation (3). From that perspective, Theorems 17.1 and 17.2 share the same Hamilton - Jacobi - Bellman-Isaac PDE and only differ by the boundary conditions.

Zero-sum Differential games with variable termination time

Proof of Theorem 17.2. From the fact that the inf and sup commute in (3) and the definitions of $\gamma^*(t,x)$ *and* $\sigma^*(t,x)$, we have that

$$g(t,x,\gamma^*(t,x),\mathrm{d})+\frac{\partial V(t,x)}{\partial x}f(t,x,\gamma^*(t,x),d) \quad \le$$

$$g(t,x,\gamma^*(t,x),\sigma^*(t,x))+\frac{\partial V(t,x)}{\partial x}f(t,x,\gamma^*(t,x),\sigma^*(t,x)) \quad \le$$

$$g(t,x,u,\sigma^*(t,x))+\frac{\partial V(t,x)}{\partial x}f(t,x,u,\sigma^*(t,x))$$

Zero-sum Differential games with variable termination time

Moreover, since the middle term in these inequalities is also equal to the right-hand-side of (3), we have that

$$-\frac{\partial V(t,x)}{\partial t} = g(t,x,\gamma^*,\sigma^*) + \frac{\partial V(t,x)}{\partial x} f(t,x,\gamma^*,\sigma^*), \mathrm{x} \in \mathrm{R}^n$$

$$= \sup_{d \in D}(g(t,x,\gamma^*(t,x),\mathrm{d}) + \frac{\partial V(t,x)}{\partial x} f(t,x,\gamma^*(t,x),d)), \forall t \in [0,\mathrm{T}]$$

which, because of Theorem Continuous-time dynamic programming in lecture 15, shows that $\sigma^*(t,x)$ is an optimal (maximizing) state-feedback policy against $\gamma^*(t,x)$ and the maximum is equal to $V(0, x_0)$.

Zero-sum Differential games with variable termination time

Moreover, since we also have that

$$-\frac{\partial V(t,x)}{\partial t} = g(t,x,\gamma^*,\sigma^*) + \frac{\partial V(t,x)}{\partial x} f(t,x,\gamma^*,\sigma^*), \mathrm{x} \in \mathrm{R}^n$$

$$= \inf_{u \in U}(g(t,x,u,\sigma^*(t,x)) + \frac{\partial V(t,x)}{\partial x} f(t,x,u,\sigma^*(t,x))), \forall t \in [0,\mathrm{T}]$$

we can also conclude that $\gamma^*(t,x)$ is an optimal (minimizing) state-feedback policy against $\sigma^*(t,x)$ and the minimum is also equal to $\mathrm{V}(0,\mathrm{x}_0)$. This proves that (γ^*,σ^*) is indeed a saddle-point equilibrium in state-feedback policies with value $\mathrm{V}(0,\,\mathrm{x}_0)$.

Zero-sum Differential games with variable termination time

Note: we actually conclude that

1. P_2 cannot get a reward larger than $V(0, x_0)$ against $\gamma^*(t,x)$, regardless of the information structure available to P_2, and

2. P_1 cannot get a cost smaller than $V(0, x_0)$ against $\sigma^*(t,x)$, regardless of the information structure available to P_1.

In practice, this means that $\gamma^*(t,x)$ and $\sigma^*(t,x)$ are "extremely safe" policies for P_1 and P_2, respectively, since they guarantee a level of reward regardless of the information structure for the other player.