COSC-6590/GSCS-6390

Games: Theory and Applications

Lecture 16 - One-Player Differential Games

Luis Rodolfo Garcia Carrillo

School of Engineering and Computing Sciences
Texas A&M University - Corpus Christi, USA

Table of contents

One-Player Continuous-Time Differential Games

One-Player Continuous-Time Differential Games

Solution methods for the optimal control of a continuous-time dynamical systems

One-player continuous-time differential game with dynamics

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

with state $x(t) \in \mathbb{R}^n$ 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 $\mathcal{U}$.

One-Player Discrete-Time Games

Assume finite horizon ($T < \infty$) integral costs of the form

$$J := \underbrace{\int_0^T g(t, x(t), u(t))dt}_{\text{cost along trajectory}} + \underbrace{q(x(T))}_{\text{final cost}}$$

that the (only) player wants to minimize using either:

Open-Loop (OL) policy

$$u(t) = \gamma_k^{\text{OL}}(t, x_0), \qquad \forall t \in [0, T]$$

State-Feedback (FB) policy

$$u(t) = \gamma_k^{\text{FB}}(t, x(t)), \qquad \forall t \in [0, T]$$

Continuous-Time Cost-To-Go

Continuous-Time Cost-To-Go

Player is at some **state** x at **time** τ

- player wants to **estimate the cost**, if playing optimally from this point on, so as **to minimize costs** incurred from this point forwards until the end of the game.

Cost-to-Go from state x at time τ

$$V(\tau, x_\tau) := \inf_{u(t)\in\mathcal{U}, \forall t\in[\tau,T]} \int_\tau^T g(t, x(t), u(t))dt + q(x(T))$$

with the state $x(t)$, $t \in [\tau, T]$ initialized at $x(\tau) = x_\tau$

and satisfying the dynamics

$$\dot{x}(t) = f(t, x(t), u(t)) \qquad \forall t \in [\tau, T]$$

Continuous-Time Cost-To-Go

Note: The **cost-to-go** is a function of x and τ.
Often called the **value function** of the game/optimization.

Computing the cost-to-go $V(0, x_0)$ from the initial state x_0 at time $\tau = 0$ essentially amounts to minimizing the cost

$$J := \underbrace{\int_0^T g(t, x(t), u(t))dt}_{\text{cost along trajectory}} + \underbrace{q(x(T))}_{\text{final cost}}$$

for the dynamics

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

This leads to two important conclusions.

Continuous-Time Cost-To-Go

Conclusion 1. Regardless of the information structure considered (OL, FB, other), it is not possible to obtain a cost

$$J := \underbrace{\int_0^T g(t, x(t), u(t))dt}_{\text{cost along trajectory}} + \underbrace{q(x(T))}_{\text{final cost}}$$

lower than $V(0, x_0)$.

This is because in the minimization

$$V(\tau, x_\tau) := \inf_{u(t)\in\mathcal{U},\forall t\in[\tau,T]} \int_\tau^T g(t, x(t), u(t))dt + q(x(T))$$

we place no constraints on what information may or may not be available to compute the optimal $u(t)$, $\forall t \in [\tau, T]$.

Continuous-Time Cost-To-Go

Conclusion 2. If the infimum in the minimization

$$V(\tau, x_\tau) := \inf_{u(t)\in\mathcal{U}, \forall t\in[\tau,T]} \int_\tau^T g(t, x(t), u(t))dt + q(x(T))$$

is achieved for a specific signal

$$u^*(t) \in \mathcal{U}, \qquad t \in [\tau, T]$$

computed before the game starts just with knowledge of x_0, then this action signal provides an **optimal OL policy**

$$\gamma_1^{\text{OL}}(t, x_0) := u^*(t), \qquad \forall t \in [\tau, T]$$

This would not be the case, e.g., if there were stochastic events

Continuous-Time Dynamic Programming

Continuous-Time Dynamic Programming

DP is a computationally efficient recursive technique

- **useful to compute the cost-to-go**

For the final time T, the cost-to-go $V(T, x_T)$ is simply

$$V(T, x_T) = q(x(T)) = q(x_T)$$

because for $t = \tau$ the integral term in

$$V(\tau, x_\tau) := \inf_{u(t)\in\mathcal{U}, \forall t\in[\tau,T]} \int_\tau^T g(t, x(t), u(t))dt + q(x(T))$$

disappears and the game starts t(and ends) precisely at $x(T) = x_T$

Continuous-Time Dynamic Programming

Consider now some time $\tau < T$. Pick some small positive constant h so that $\tau + h$ is still smaller than T. Then

$$
\begin{aligned}
V(\tau, x_\tau) &= \inf_{u(t)\in\mathcal{U},\forall t\in[\tau,T]} \int_\tau^T g(t,x(t),u(t))dt + q(x(T)) \\
&= \inf_{u(t)\in\mathcal{U},\forall t\in[\tau,T]} \underbrace{\int_\tau^{\tau+h} g(t,x(t),u(t))dt}_{\text{independent of } u(t),t\in[\tau+h,T]} \\
&\qquad + \underbrace{\int_{\tau+h}^T g(t,x(t),u(t))dt + q(x(T))}_{\text{depends on all } u(t),t\in[\tau,T]} \\
&= \inf_{u(t)\in\mathcal{U},\forall t\in[\tau,\tau+h)} \Bigg(\int_\tau^{\tau+h} g(t,x(t),u(t))dt \\
&\qquad + \inf_{u(t)\in\mathcal{U},\forall t\in[\tau+h,T]} \int_{\tau+h}^T g(t,x(t),u(t))dt + q(x(T)) \Bigg)
\end{aligned}
$$

Continuous-Time Dynamic Programming

Observation: **inner** infimum is the cost-to-go from the state $x(\tau+h)$ at time $\tau+h$. Then, we can re-write these equations compactly as

$$V(\tau, x_\tau) = \inf_{u(t)\in\mathcal{U}, \forall t\in[\tau,\tau+h]} \Big(\int_\tau^{\tau+h} g(t, x(t), u(t))dt + V(\tau+h, x(\tau+h)) \Big)$$

Subtracting $V(\tau, x_\tau) = V(\tau, x(\tau))$ from both sides and dividing both sides by $h > 0$, we can re-write the equation as

$$0 = \inf_{u(t)\in\mathcal{U}, \forall t\in[\tau,\tau+h)} \Big(\frac{1}{h} \int_\tau^{\tau+h} g(t, x(t), u(t))dt + \frac{V(\tau+h, x(\tau+h)) - V(\tau, x(\tau))}{h} \Big)$$

Since LHS must be equal to zero for every $h \in (0, T-\tau)$, the limit of the RHS as $h \to 0$ must also be equal to zero.

Continuous-Time Dynamic Programming

Optimistic assumption: the limit of the infimum is the same as the infimum of the limit and also that all limits exist. Then, we can use the equalities

$$\lim_{h\to 0}\frac{1}{h}\int_{\tau}^{\tau+h} g(t,x(t),u(t))dt = g(\tau,x(\tau),u(\tau))$$

$$\lim_{h\to 0}\frac{V(\tau+h,x(\tau+h)) - V(\tau,x(\tau))}{h}$$
$$= \frac{dV(\tau,x(\tau))}{d\tau} = \frac{\partial V(\tau,x(\tau))}{\partial \tau} + \frac{\partial V(\tau,x(\tau))}{\partial x} f(\tau,x(\tau),u(\tau))$$

to transform the previous equation into the so-called **Hamilton-Jacobi-Bellman** (HJB) equation

$$0 = \inf_{u\in\mathcal{U}}\left(g(\tau,x(\tau),u(\tau)) + \frac{\partial V(\tau,x)}{\partial \tau} + \frac{\partial V(\tau,x)}{\partial x} f(\tau,x,u)\right), \quad \forall \tau\in[0,T], \quad x\in\mathbb{R}^n$$

HJB equation is useful to compute the cost-to-go.

Continuous-Time Dynamic Programming

Note. The infimum in

$$0 = \inf_{u(t)\in\mathcal{U},\forall t\in[\tau,\tau+h)} \Big(\frac{1}{h}\int_{\tau}^{\tau+h} g(t,x(t),u(t))dt + \frac{V(\tau+h,x(\tau+h)) - V(\tau,x(\tau))}{h}\Big)$$

is taken over all the values of the signal $u(t)$ in the interval $t \in [\tau + \tau + h)$, which is the subject of calculus of variations.

However, in the HJB equation

$$0 = \inf_{u\in\mathcal{U}} \Big(g(\tau,x(\tau),u(\tau)) + \frac{\partial V(\tau,x)}{\partial \tau} + \frac{\partial V(\tau,x)}{\partial x} f(\tau,x,u)\Big), \quad \forall \tau \in [0,T], \quad x \in \mathbb{R}^n$$

the infimum is simply taken over the set $\mathcal{U}$ and can generally be solved using standard calculus.

Continuous-Time Dynamic Programming

Theorem 16.1 (Hamilton-Jacobi-Bellman).

Any continuously differentiable function $V(\tau, x)$ that satisfies the HJB equation

$$0 = \inf_{u \in \mathcal{U}} \left(g(\tau, x(\tau), u(\tau)) + \frac{\partial V(\tau, x)}{\partial \tau} + \frac{\partial V(\tau, x)}{\partial x} f(\tau, x, u) \right), \quad \forall \tau \in [0, T], \quad x \in \mathbb{R}^n$$

with

$$V(T, x) = q(x), \qquad \forall x \in \mathbb{R}^N$$

is equal to the cost-to-go $V(\tau, x)$.

In addition, if the infimum in the HJB equation is always achieved at some point in $\mathcal{U}$, we have the following.

Continuous-Time Dynamic Programming

1. For any given x_0, an optimal OL policy γ^{OL} is given by

$$\gamma^{\text{OL}}(t, x_0) := u^*(t), \qquad \forall t \in [0, T]$$

with $u^*(t)$ obtained from solving

$$u^*(t) = \arg\min_{u \in \mathcal{U}} g(t, x^*(t), u) + \frac{\partial V(t, x^*(t))}{\partial x} f(t, x^*(t), u)$$

$$\dot{x}^*(t) = f(t, x^*(t), u^*(t)), \quad \forall t \in [0, T], \quad x^*(0) = x_0$$

Note. In an OL setting both $x^*(t)$ and $u^*(t)$, $t \in [0, T]$ are precomputed before the game starts.

Continuous-Time Dynamic Programming

2. An optimal (time-consistent) state-FB policy γ^{FB} is given by

$$\gamma^{\mathrm{FB}}(t, x_t) := \arg\min_{u \in \mathcal{U}} g(t, x(t), u) + \frac{\partial V(t, x(t))}{\partial x} f(t, x(t), u), \quad \forall t \in [0, T]$$

Either of the above optimal policies leads to an optimal cost equal to $V(0, x_0)$

Note 15. OL and state-FB information structures are **optimal**, in the sense that it is not possible to achieve a cost lower than $V(0, x_0)$, regardless of the information structure.

Continuous-Time Dynamic Programming

Note 14 (Hamilton-Jacobi-Bellman equation).

Since $\frac{\partial V(\tau,x)}{\partial x}$ in

$$0 = \inf_{u \in \mathcal{U}} \Big(g(\tau, x(\tau), u(\tau)) + \frac{\partial V(\tau,x)}{\partial \tau} + \frac{\partial V(\tau,x)}{\partial x} f(\tau,x,u) \Big), \quad \forall \tau \in [0,T], \quad x \in \mathbb{R}^n$$

does not depend of u, we remove this term from inside the infimum. This leads to the common form for the HJB:

$$-\frac{\partial V(\tau,x)}{\partial x} = \inf_{u \in \mathcal{U}} \Big(g(\tau,x,u) + \frac{\partial V(\tau,x)}{\partial x} f(\tau,x,u) \Big)$$

This form highlights the fact that the HJB is a partial differential equation (PDE) and we can view

$$V(T,x) = q(x), \qquad \forall x \in \mathbb{R}^N$$

as a boundary condition for this PDE.

Continuous-Time Dynamic Programming

When we find a continuously differentiable solution to this PDE that satisfies the boundary condition, we automatically obtain the cost-to-go. Unfortunately, solving a PDE is often difficult, and many times

$$0 = \inf_{u \in \mathcal{U}} \left(g(\tau, x(\tau), u(\tau)) + \frac{\partial V(\tau, x)}{\partial \tau} + \frac{\partial V(\tau, x)}{\partial x} f(\tau, x, u) \right), \quad \forall \tau \in [0, T], \quad x \in \mathbb{R}^n$$

does not have continuously differentiable solutions.

Note The lack of differentiability of the solution to this equation is not an insurmountable difficulty.

There are methods to overcome this technical difficulty by making sense of what it means for a non-differentiable function to be a solution o this equation.

Continuous-Time Dynamic Programming

Proof of Theorem 16.1.

Let $u^*(t)$ and $x^*(t)$, $\forall t \in [0, T]$ be a trajectory arising from either the OL or the state-FB policies.

- note that both policies lead to the same trajectory.

Let $\bar{u}(t)$ and $\bar{x}(t)$, $\forall t \in [0, T]$ be another (arbitrary) trajectory.

To prove optimality, show that the latter trajectory cannot lead to a cost lower than the former.

Since $V(\tau, x)$ satisfies the HJB equation, and $u^*(t)$ achieves the infimum in the HJB equation, for every $t \in [0, T]$, we have that

$$\begin{aligned} 0 &= \inf_{u \in \mathcal{U}} g(t, x^*(t), u) + \frac{\partial V(t, x^*(t))}{\partial \tau} + \frac{\partial V(t, x^*(t))}{\partial x} f(t, x^*(t), u) \\ &= g(t, x^*(t), u^*(t)) + \frac{\partial V(t, x^*(t))}{\partial \tau} + \frac{\partial V(t, x^*(t))}{\partial x} f(t, x^*(t), u^*(t)) \end{aligned}$$

Continuous-Time Dynamic Programming

However, since $\bar{u}(t)$ does not necessarily achieve the infimum, we have that

$$
\begin{aligned}
0 &= \inf_{u \in \mathcal{U}} g(t, \bar{x}(t), u) + \frac{\partial V(t, \bar{x}(t))}{\partial \tau} + \frac{\partial V(t, \bar{x}(t))}{\partial x} f(t, \bar{x}(t), u) \\
&\leq g(t, \bar{x}(t), \bar{u}(t)) + \frac{\partial V(t, \bar{x}(t))}{\partial \tau} + \frac{\partial V(t, \bar{x}(t))}{\partial x} f(t, \bar{x}(t), \bar{u}(t))
\end{aligned}
$$

Integrating both sides of this and the previous eq. over the interval $[0, T]$ we conclude

$$
\begin{aligned}
0 &= \int_0^T \Big(g(t, x^*(t), u^*(t)) + \underbrace{\frac{\partial V(t, x^*(t))}{\partial \tau} + \frac{\partial V(t, x^*(t))}{\partial x} f(t, x^*(t), u^*(t))}_{\frac{dV(t, x^*(t))}{dt}} \Big) dt \\
&\leq \int_0^T \Big(g(t, \bar{x}(t), \bar{u}(t)) + \underbrace{\frac{\partial V(t, \bar{x}(t))}{\partial \tau} + \frac{\partial V(t, \bar{x}(t))}{\partial x} f(t, \bar{x}(t), \bar{u}(t))}_{\frac{dV(t, \bar{x}(t))}{dt}} \Big) dt
\end{aligned}
$$

Continuous-Time Dynamic Programming

From which we obtain

$$\begin{aligned} 0 &= \int_0^T g(t, x^*(t), u^*(t))dt + V(T, x^*(T)) - V(0, x_0) \\ &\leq \int_0^T g(t, \bar{x}(t), \bar{u}(t))dt + V(T, \bar{x}(T)) - V(0, x_0) \end{aligned}$$

Using

$$V(T, x) = q(x), \quad \forall x \in \mathbb{R}^n$$

and adding $V(0, x_0)$ to all terms, one concludes that

$$V(0, x_0) = \int_0^T g(t, x^*(t), u^*(t))dt + q(x^*(T)) \leq \int_0^T g(t, \bar{x}(t), \bar{u}(t))dt + q(\bar{x}(T))$$

Continuous-Time Dynamic Programming

Two conclusions can be drawn:

1. signal $\bar{u}(t)$ does not lead to a cost smaller than that of $u^*(t)$
2. $V(0, x_0)$ is equal to the optimal cost obtained with $u^*(t)$

If we had carried out the above proof on an interval $[\tau, T]$ with initial state $x(\tau) = x$, we would have concluded that $V(\tau, x)$ is the (optimal) value of the cost-to-go from state x at time τ.

Note 15. In the proofs of **Theorems 16.1** (and later in the proof of **Theorem 16.2**) we show that it is not possible to achieve a cost lower than $V(0, x_0)$, regardless of the information structure.

This is because the signal $\bar{u}(t)$ considered could have been generated by a policy using any information structure and we showed $\bar{u}(t)$ cannot lead to a cost smaller than $V(0, x_0)$.

Linear Quadratic Dynamic Games

Linear Quadratic Dynamic Games

Continuous-time **linear quadratic one-player games** are characterized by linear dynamics like

$$\dot{x} = \underbrace{Ax(t) + Bu(t)}_{f(t,x(t),u(t))}, \quad x \in \mathbb{R}^n, u \in \mathbb{R}^{n)u}, \quad t \in [0, T]$$

and an integral quadratic cost of the form

$$J := \int_0^T \underbrace{\left(||y(t)||^2 + u(t)' R u(t)\right)}_{g(t,x(t),u(t))} dt + \underbrace{x'(T) P_T x(T)}_{q(x(T))}$$

where

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

Linear Quadratic Dynamic Games

Cost function J captures scenarios in which the (only) player wants to make $y(t)$ small over the interval $[0, T]$ without **spending** much effort in the action $u(t)$.

R : symmetric positive definite matrix

- conversion factor that maps units of $u(t)$ into units of $y(t)$.

The HJB equation for this game is then

$$-\frac{\partial V(t,x)}{\partial t} = \min_{u \in \mathbb{R}^{n_u}} \left(x'C'Cx + u'Ru + \frac{\partial V(t,x)}{\partial x}(Ax + Bu) \right)$$

$x \in \mathbb{R}^n$, $t \in [0, T]$, with boundary condition

$$V(T, x) = x'P_T x, \qquad \forall x \in \mathbb{R}^n$$

Linear Quadratic Dynamic Games

Inspired by the boundary condition $V(T,x)$, we will try to find a solution to the HJB of the form

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

for some selected symmetric $n \times n$ matrix $P(t)$.

For $V(T,x)$ to hold, we need $P(T) = P_T$.

For the HJB to hold, we need

$$-x'\dot{P}(t)x = \min_{u \in \mathbb{R}^{n_u}} \left(x'C'Cx + u'Ru + 2x'P(t)(Ax + Bu) \right)$$

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

Linear Quadratic Dynamic Games

Function to optimize is quadratic:

- to compute the minimum in previous eq., make the appropriate gradient equal to zero

$$\frac{\partial}{\partial u}\big(x'C'Cx + u'Ru + 2x'P(Ax+Bu)\big) = 0$$
$$\Leftrightarrow 2u'R + 2x'PB = 0 \Leftrightarrow u = -R^{-1}B'Px$$

The critical point obtained by setting $\frac{\partial(\cdot)}{\partial u}$ is a minimum because $R > 0$

- the value for the minimum will provide the optimal policy.

Therefore

$$-x'\dot{P}(t)x = \min_{u\in\mathbb{R}^{n_u}} \big(\underbrace{x'C'Cx + u'Ru + 2x'P(t)(Ax+Bu)}_{u=-R^{-1}B'Px}\big)$$
$$= x'(PA + A'P + C'C - PBR^{-1}B'P)x$$

Linear Quadratic Dynamic Games

Note. Since P is symmetric, we wrote

$$2x'PAx \qquad \text{as} \qquad x'(PA + A'P)x$$

From the previous equation we have that

$$-x'\dot{P}(t)x = x'(PA + A'P + C'C - PBR^{-1}B'P)x$$

$\forall x \in \mathbb{R}^n$, $t \in [0, T]$, which holds provided that

$$-\dot{P}(t) = PA + A'P + C'C - PBR^{-1}B'P, \quad \forall t \in [0, T]$$

Theorem 16.1 can then be used to compute the optimal policies for this game.

Linear Quadratic Dynamic Games

Corollary 16.1. Suppose there exists a symmetric solution to the matrix-valued ODE

$$-\dot{P}(t) = PA + A'P + C'C - PBR^{-1}B'P, \quad \forall t \in [0, T]$$

with final condition $P(T) = P_T$.

Note: The function $P(t)$ could be found by numerically solving the (**differential Ricatti Equation**) matrix-valued ODE backwards in time.

Then the state-FB policy

$$\gamma^*(t, x) = -R^{-1}B'Px, \quad \forall x \in \mathbb{R}^n, t \in [0, T]$$

is an optimal state-FB policy, leading to an optimal cost $x_0'P(0)x_0$.

Differential Games with Variable Termination Time

Differential Games with Variable Termination Time

Consider a one-player continuous-time differential game with dynamics

$$\dot{x}(t) = f(t, x(t), u(t)), \quad x(t) \in \mathbb{R}^n, u(t) \in \mathcal{U}, \geq 0$$

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

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

where T_{end} is the first time at which state $x(t)$ enters a closed set $\mathcal{X}_{\text{end}} \subset \mathbb{R}^n$ or $T_{\text{end}} = +\infty$ in case $x(t)$ never enters $\mathcal{X}_{\text{end}}$.

$\mathcal{X}_{\text{end}}$: set of states at which the game terminates. Evolution of $x(t)$ is irrelevant after this time. (Game Over States)

Differential Games with Variable Termination Time

Cost-to-go (value function) from state x at time τ is

$$V(\tau, x) := \inf_{u(t) \in \mathcal{U}, \forall t \geq \tau} \int_{\tau}^{T_{\text{end}}} g(t, x(t), u(t)) dt + q(T_{\text{end}}, x(T_{\text{end}}))$$

where the state $x(t)$, $t \geq \tau$ satisfies the dynamics

$$x(\tau) = x, \qquad \dot{x}(t) = f(t, x(t), u(t)), \quad \forall t \geq \tau$$

T_{end} : first time at which $x(t)$ enters the closed set $\mathcal{X}_{\text{end}}$.

When we compute $V(\tau, x)$ for some $x \in \mathcal{X}_{\text{end}}$, we have $T_{\text{end}} = \tau$ and therefore

$$V(\tau, x) = q(\tau, x), \qquad \forall \tau \geq 0, x \in \mathcal{X}_{\text{end}}$$

instead of the boundary condition $V(T, x)$. It turns out that the HJB equation is still the same.

Differential Games with Variable Termination Time

Theorem 16.2. A continuously differentiable function $V(\tau, x)$ that satisfies the HJB equation with (a boundary condition)

$$V(\tau, x) = q(\tau, x), \qquad \forall \tau \geq 0, \quad x \in \mathcal{X}_{\text{end}}$$

is equal to the cost-to-go $V(\tau, x)$. If the infimum in the HJB is always achieved at some point in $\mathcal{U}$, we have that:

1. For any given x_0, an optimal OL policy γ^{OL} is given by

$$\gamma^{\text{OL}}(t, x_0) := u^*(t), \qquad \forall t \in [0, T_{\text{end}}]$$

with $u^*(t)$ obtained from solving

$$u^*(t) = \arg\min_{u \in \mathcal{U}} g(t, x^*(t), u) + \frac{\partial V(t, x^*(t))}{\partial x} f(t, x^*(t), u)$$

$$x^*(t) = f(t, x^*(t), u^*(t)), \qquad \forall t \in [0, T_{\text{end}}], \quad x^*(0) = 0$$

Differential Games with Variable Termination Time

2. An optimal (time-consistent) state-FB policy γ^{FB} is given by

$$\gamma^{\text{FB}}(t, x_t) := \arg\min_{u \in \mathcal{U}} g(t, x(t), u) + \frac{\partial V(t, x(t))}{\partial x} f(t, x(t), u), \quad \forall t \in [0, T_{\text{en}}$$

Either of the optimal policies γ^{OL} or γ^{FB} leads to an optimal cost equal to $V(0, x_0)$.

Proof of Theorem 16.2.
Let $u^*(t)$ and $x^*(t)$, $\forall t \geq 0$ be a trajectory arising from the OL or the state-FB policies and let $\bar{u}^*(t)$ and $\bar{x}^*(t)$, $\forall t \geq 0$ be another (arbitrary) trajectory.

To prove optimality, show that the latter trajectory cannot lead to a cost lower than the former.

Differential Games with Variable Termination Time

As in the proof of **Theorem 16.1**, $V(\tau, x)$ satisfies the HJB equation, and $u^*(t)$ achieves the infimum in it.

Let $T^*_{\text{end}} \in [0, \infty]$, and $\bar{T}_{\text{end}} \in [0, \infty]$ denote the times at which $x^*(t)$ and $\bar{x}(t)$, respectively, enter the set $\mathcal{X}_{\text{end}}$.

Integrating both sides of respective equations over the intervals $[0, T_{\text{end}^*}]$ and $[0, \bar{T}_{\text{end}}]$, respectively, we conclude that

$$0 = \int_0^{T^*_{\text{end}}} \Big(g(t, x^*(t), u^*(t)) + \underbrace{\frac{\partial V(t, x^*(t))}{\partial \tau} + \frac{\partial V(t, x^*(t))}{\partial x} f(t, x^*(t), u^*(t))}_{\frac{dV(t, x^*(t))}{dt}} \Big) dt$$

$$\leq \int_0^{\bar{T}_{\text{end}}} \Big(g(t, \bar{x}(t), \bar{u}(t)) + \underbrace{\frac{\partial V(t, \bar{x}(t))}{\partial \tau} + \frac{\partial V(t, \bar{x}(t))}{\partial x} f(t, \bar{x}(t), \bar{u}(t))}_{\frac{dV(t, \bar{x}(t))}{dt}} \Big) dt$$

Differential Games with Variable Termination Time

From which we obtain

$$
\begin{aligned}
0 &= \int_0^{T^*_{\text{end}}} g(t, x^*(t), u^*(t))dt + V(T^*_{\text{end}}, x^*(T^*_{\text{end}})) - V(0, x_0) \\
&\leq \int_0^{\bar{T}^*_{\text{end}}} g(t, \bar{x}(t), \bar{u}(t))dt + V(\bar{T}_{\text{end}}, \bar{x}(\bar{T}_{\text{end}})) - V(0, x_0)
\end{aligned}
$$

Using

$$V(\tau, x) = q(\tau, x), \quad \forall \tau \geq 0, \quad x \in \mathcal{X}_{\text{end}}$$

two conclusions can be drown from here.

Differential Games with Variable Termination Time

First, the signal $\bar{u}(t)$ does not lead to a cost smaller than that of $u^*(t)$, because

$$\int_0^{T^*_{\text{end}}} g(t, x^*(t), u^*(t))dt + q(T^*_{\text{end}}, x^*(T^*_{\text{end}})) \leq \int_0^{\bar{T}^*_{\text{end}}} g(t, \bar{x}(t), \bar{u}(t))dt + q(\bar{T}_{\text{end}}, \bar{x}(\bar{T}_{\text{end}}))$$

Second, $V(0, x_0)$ is equal to the optimal cost obtained with $u^*(t)$, because

$$V(0, x_0) = \int_0^{T^*_{\text{end}}} g(t, x^*(t), u^*(t))dt + q(T^*_{\text{end}}, x^*(T^*_{\text{end}}))$$

If we had carried out the proof starting at time τ with initial state $x(\tau) = x$, we would have concluded that $V(\tau, x)$ is the (optimal) value of the cost-to-go from state x at time τ.

Practice Exercise

Practice Exercise

16.1. Prove the following result, which permits the construction of a state-FB policy based on a function that only satisfies the HJB equation approximately.

Theorem 16.3. Suppose that there exist constants $\epsilon, \delta \geq 0$, a continuously differentiable function $V(\tau, x)$ that satisfies

$$\Big| \inf_{u \in \mathcal{U}} \Big(g(\tau, x, u) + \frac{\partial V(\tau, x)}{\partial \tau} + \frac{\partial V(\tau, x)}{\partial x} f(\tau, x, u) \Big) \Big| \leq \epsilon, \quad \forall \tau \in [0, T], x \in \mathbb{R}^n$$

with

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

and a state-feedback policy $\gamma(\cdot)$ for which

$$g(\tau, x, u) + \frac{\partial V(\tau, x)}{\partial x} f(\tau, x, u) \Big|_{u=\gamma(x)} \leq \delta + \inf_{u \in \mathcal{U}} \Big(g(\tau, x, u) + \frac{\partial V(\tau, x)}{\partial x} f(\tau, x, u) \Big)$$

$\forall \tau \in [0, T], x \in \mathbb{R}^n$

Practice Exercise

Then the policy $\gamma(\cdot)$ leads to a cost $J(\gamma)$ that satisfies

$$J(\gamma) \leq J(\bar{\gamma}) + (2\epsilon + \delta)T$$

for any other state-FB policy $\bar{\gamma}(\cdot)$ with cost $J(\bar{\gamma})$.

Solution to Exercise 16.1. Proof of Theorem 16.3.

Let $u^*(t)$ and $x^*(t)$, $\forall t \in [0, T]$ be a trajectory arising from the state-FB policy $\gamma(\cdot)$.

Let $\bar{u}(t)$ and $\bar{x}(t)$, $\forall t \in [0, T]$ be another trajectory, e.g., resulting from the state-FB policy $\bar{\gamma}(\cdot)$ that appears in $J(\gamma) \leq J(\bar{\gamma}) + (2\epsilon + \delta)T$.

Since $V(\tau, x)$ satisfies the ϵ condition for $x = x^*(t)$, and $\gamma(\cdot)$ satisfies the δ condition for $x = x^*(t)$, and $u = \gamma(x^*(t)) = u^*(t)$ satisfies the δ condition, we have the following.

Practice Exercise

$$\epsilon \geq \inf_{u\in\mathcal{U}} \left(g(t, x^*(t), u) + \frac{\partial V(t, x^*(t))}{\partial t} + \frac{\partial V(t, x^*(t))}{\partial x} f(t, x^*(t), u)\right)$$
$$\geq -\delta + g(t, x^*(t), u^*(t)) + \frac{\partial V(t, x^*(t))}{\partial t} + \frac{\partial V(t, x^*(t))}{\partial x} f(t, x^*(t), u^*(t))\Big)$$

On the other hand, using the ϵ condition for $x = \bar{x}(t)$ and the fact that $\bar{u}(t)$ does not necessarily achieve the infimum, we have

$$-\epsilon \leq \inf_{u\in\mathcal{U}} \left(g(t, \bar{x}(t), u) + \frac{\partial V(t, \bar{x}(t))}{\partial t} + \frac{\partial V(t, \bar{x}(t))}{\partial x} f(t, \bar{x}(t), u)\right)$$
$$\geq g(t, \bar{x}(t), \bar{u}(t)) + \frac{\partial V(t, \bar{x}(t))}{\partial t} + \frac{\partial V(t, \bar{x}(t))}{\partial x} f(t, \bar{x}(t), \bar{u}(t))\Big)$$

Practice Exercise

Integrating both sides of the previous equations over the interval $[0, T]$, we conclude that

$$(\epsilon + \delta)T \geq \Big(g(t, x^*(t), u^*(t) + \underbrace{\frac{\partial V(t, x^*(t))}{\partial t} + \frac{\partial V(t, x^*(t))}{\partial x} f(t, x^*(t), u^*(t)}_{\frac{dV(t,x^*(t))}{dt}}\Big)dt$$

$$= \int_0^T (g(t, x^*(t), u^*(t))dt + V(T, x^*(T)) - V(0, x_0)$$

and

$$-\epsilon T \leq \Big(g(t, \bar{x}(t), \bar{u}(t) + \underbrace{\frac{\partial V(t, \bar{x}(t))}{\partial t} + \frac{\partial V(t, \bar{x}(t))}{\partial x} f(t, \bar{x}(t), \bar{u}(t)}_{\frac{dV(t,\bar{x}(t))}{dt}}\Big)dt$$

$$= \int_0^T (g(t, \bar{x}(t), \bar{u}(t))dt + V(T, \bar{x}(T)) - V(0, x_0)$$

Practice Exercise

From which we obtain

$$\int_0^T g(t, x^*(t), u^*(t))dt + V(T, x^*(T)) \leq V(0, x_0) + (\epsilon + \delta)T$$

$$\leq (2\epsilon + \delta)T + \int_0^T g(t, \bar{x}(t), \bar{u}(t))dt + V(T, \bar{x}(T))$$

Which proves

$$J(\gamma) \leq J(\bar{\gamma}) + (2\epsilon + \delta)T$$

End of Lecture

16 - One-Player Differential Games

Questions?