Reinforcement Learning and Optimal Control

IFT6760C, Fall 2021

Pierre-Luc Bacon

December 4, 2021

Continuous-Time OCP

We consider problems of the form:

$$
\begin{aligned}
&\text{minimize} \;\; c(x(t_f)) + \int_{t_0}^{t_f} c(x(t), u(t))dt \\
&\text{subject to} \;\; \dot{x}(t) = f(x(t), u(t)) \\
&\text{given} \;\; x(t_0) = x_0 \quad .
\end{aligned}
$$

The Hamilton-Jacobi Equations

We can show that the cost-to-go for the above continuous-time problem function satisfies for all x, t:

$$J(x, T) = c(x) \quad \text{and} \quad 0 = \min_{u \in \mathcal{U}} \left\{ c(x, u) + \underbrace{D_2 J(x, t)}_{\text{time derivative}} + \overbrace{D_1 J(x, t)}^{\text{space derivative}} f(x, u) \right\}$$

This is a partial differential equation (PDE), with both time and space partial derivatives.

HJB as a PDE

Theorem (Sufficiency). Let $V(x,t)$ be a solution to the PDE:

$$0 = \min_{u \in \mathcal{U}} \{c(x,u) + D_2 V(x,t) + D_1 V(x,t) f(x,u)\} \text{ for all } x, t \quad ,$$

and boundary condition $V(x, t_f) = c(x)$ for all x. Then V is the cost-to-go function J and an optimal policy $\mu(x,t)$ can be obtained by minimizing the expression above given $V(x,t)$.

HJB in the Infinite-Horizon Case

$$\int_{t_0}^{\infty} c(x(t), u(t))dt$$

$$\text{subject to } \dot{x}(t) = f(x(t), u(t))$$

$$\text{given } x(t_0) = x_0 \ .$$

The time derivative now disappears from the cost-to-go function:

$$0 = \min_{u \in \mathcal{U}} \{c(x, u) + DJ(x) f(x, u)\} \quad \text{for all } x$$

A "v-improving" policy can then be obtained as:

$$\mu^\star(x) \in \arg\min_{u \in \mathcal{U}} \{c(x, u) + DJ(x) f(x, u)\}$$

Control Affine Dynamics

In the general case, the minimizer in the HJB equation cannot be computed in closed-form. However, this is possible for the class of control affine systems of the form:

$$f(x, u) \triangleq g(x) + h(x)u \quad .$$

for some given functions g and h. Furthermore, we can further restrict our attention to immediate cost functions of the form:

$$c(x, u) \triangleq l(x) + u^\top Ru \quad .$$

Note that the the above setting is more general the the plain LQR one.

Infinite-Horizon HJB under Control Affine Assumptions

$$0 = \min_{u \in \mathcal{U}} \left\{ l(x) + u^\top R u + DJ(x) \left(g(x) + h(x) u \right) \right\} \text{ for all } x, t \quad ,$$

If we set the derivative with respect to u of the quantity inside the min operator to zero, we get:

$$2u^\top R + DJ(x) h(x) = 0 \quad ,$$

which means that:

$$u^\star = -\frac{1}{2} R^{-1} h(x)^\top DJ(x)^\top$$

LQR in Continuous-Time

Consider the following finite-horizon continuous-time LQR problem:

$$
\begin{aligned}
&\text{minimize} \;\; x(t_f)^\top Q_{t_f} x(t_f) + \int_{t_0}^{\infty} \left(x(t)^\top Q x(t) + u(t)^\top R u(t) \right) \\
&\text{such that} \;\; \dot{x}(t) = Ax(t) + Bu(t) \\
&\text{given} \;\; x(t_0) = x_0 \quad .
\end{aligned}
$$

Solving the HJB for LQR

The HJB equation is then:

$$
\begin{aligned}
J(x, t_f) &= x^\top Q_{t_f} x \\
0 &= \min_{u \in \mathcal{U}} \left\{ x^\top Q x + u^\top R u + D_2 J(x,t) + D_1 J(x,t)\,(Ax + Bu) \right\}
\end{aligned}
$$

Remember that in the discrete-time case, we had that the cost-to-go function was a quadratic. We can attempt to solve the above in the same way and start with $J(x,t) = x^\top K(t) x$ where $K(t)$ is a symmetric matrix. Note that:

$$
D_1 J(x,t) = 2x^\top K(t) \qquad \text{and} \qquad D_2 J(x,t) = x^\top \dot{K}(t) x
$$

Substitution

If we substitute our guess into the HJB, we get:

$$0 = \min_{u \in \mathcal{U}} \left\{ x^\top Q x + u^\top R u + x^\top \dot{K}(t) x + 2x^\top K(t) \left(Ax + Bu \right) \right\}$$

Setting the partial derivative of the inside quantity with respect to u to zero, we get:

$$2u^\top R + 2x^\top K(t) B = 0$$

Solving for u, we get:

$$u^\star = -R^{-1} B^\top K(t) x \quad .$$

Substituting $u^\star$

Plugging

$$u^\star = -R^{-1}B^\top K(t)x \quad ,$$

into the HJB:

$$\begin{aligned} 0 &= \min_{u \in \mathcal{U}} \left\{ x^\top Qx + u^\top Ru + D_2 J(x,t) + D_1 J(x,t)\,(Ax + Bu) \right\} \\ &= x^\top Qx + u^{\star\top} Ru^\star + D_2 J(x,t) + D_1 J(x,t) \left(Ax + Bu^{\star\top} \right) \\ &= x^\top \left(\dot{K}(t) + K(t)A + A^\top K(t) - K(t)BR^{-1}B^\top K(t) + Q \right) x \quad , \end{aligned}$$

for all x and t.

Continuous-Time Ricatti Equation

Matrix differential equation:

$$\dot{K}(t) = -K(t)QA - A^\top K(t) + K(t)BR^{-1}B^\top K(t) - Q \quad ,$$

with terminal condition $K(t) = Q_{t_f}$.

Solving this equation then allows us to find the cost-to-go function as $J(x,t) = x^\top K(t)x$. With this $K(t)$ in hand, we can also compute the optimal controls with:

$$\mu^\star(x,t) = -R^{-1}B^\top K(t)x \quad .$$

From HJB to PMP

Proposition 3.3.1 Let $\{u^\star(t) | t \in [0, t_f]\}$ be a an optimal control trajectory and let $\{x^\star | t \in [0, T]\}$ be the corresponding state trajectory. That is:

$$\dot{x}^\star(t) = f(x^\star(t), u^\star(t)), \quad x^\star(t_0) = x_0 \quad .$$

For all $t \in [0, t_f]$:

$$u^\star(t) \in \arg\min_{u \in \mathcal{U}} H(x^\star(t), u, \lambda(t)) \quad ,$$

where H is the Hamiltonian, ie:

$$H(x, u, \lambda) \triangleq c(x, u) + \lambda f(x, u) \quad .$$

From HJB to PMP

(continued) and λ satisfies the adjoint equation:

$$\dot{\lambda}(t) = -D_1 H(x^\star(t), u^\star(t), \lambda(t)) \quad ,$$

with boundary condition $\lambda(t_f) = Dc(x^\star(t_f))$. Furthermore, there is a constant C such that:

$$H(x^\star(t), u^\star(t), \lambda(t)) = C \quad ,$$

for all $t \in [0, t_f]$.

Derivation (informal)

Here, we use the approach in Bertsekas, which makes an assumption on the set of controls $\mathcal{U}$ being convex, in tandem with lemma 3.3.1 about differentiating through the min operator.

Let's start by inserting the optimal policy $\mu^\star$ into the HJB. We get:

$$c(x, \mu^\star(x,t)) + D_2 J(x,t) + D_1 J(x,t) f(x, \mu^\star(x,t)) = 0 \quad .$$

Differenting on both sides: