Reinforcement Learning and Optimal Control

IFT6760C, Fall 2021

Pierre-Luc Bacon

December 6, 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:

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

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

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 $\{x^\star | t \in [0, t_f]\}$ 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.

Lemma 3.3.1 Let $F(x, u, t)$ be a continuously differentiable function of $t \in \mathbb{R}$, $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ and let $\mathcal{U}$ be a convex subset of $\mathbb{R}^m$. Assume that $\mu^\star(x, t)$ is a continuously differentiable function of u such that:

$$\mu^\star(x, t) = \arg\min_{u \in \mathcal{U}} F(x, u, t) \quad \text{for all } x, t \quad .$$

Lemma 3.3.1 continued

Let $G(x,t) \triangleq \min_{u\in\mathcal{U}} F(x,u,t)$, it follows that for all x, t:

1. $D_1G(x,t) = D_1F(x,\mu^\star(x,t),t)$
2. $D_2G(x,t) = D_3F(x,\mu^\star(x,t),t)$

Proof: Note that $G(x,t) = F(x,\mu^\star(x,t),t)$. Taking the (total) derivative:

$$D_1G(x,t) = D_1F(x,\mu^\star(x,t),t) + D_2F(x,\mu^\star(x,t),t)D_1\mu^\star(x,t)$$
$$D_2G(x,t) = D_3F(x,\mu^\star(x,t),t) + D_2F(x,\mu^\star(x,t),t)D_2\mu^\star(x,t)$$

Or more compactly with $y = (x,t)$ and
$G(y) = \min_{u\in\mathcal{U}} F(y,u) = F(y,\mu^\star(y))$:

$$DG(y) = D_1F(y,\mu^\star(y)) + D_2F(y,\mu^\star(y))D\mu^\star(y) \quad .$$

Lemma 3.3.1: Taylor Theorem

Let $\mu^\star(x, t) \triangleq \mu^\star(y)$, by Taylor's theorem:

$$\mu^\star(z) = \mu^\star(y) + D\mu^\star(y)(z - y) + h_1(z)(z - y) \quad ,$$

with $\lim_{z \to y} h_1(z) = 0$. If we evaluate the function at $y + \Delta_y$:

$$\mu^\star(y + \Delta_y) = \mu^\star(y) + D\mu^\star(y)\Delta_y + o(\|\Delta_y\|) \quad ,$$

or written differently:

$$\mu^\star(y + \Delta_y) - \mu^\star(y) = D\mu^\star(y)\Delta_y + o(\|\Delta_y\|) \quad .$$

(the difference in the output for a perturbation Δ_y)

Lemma 3.3.1: Optimality Conditions

The optimality conditions in the convex case are:

$$D_2F(y, \mu^\star(y))(u - \mu^\star(y)) \geq 0 \text{ for all } u \in \mathcal{U}$$

Which we can also write as:

$$D_2F(y, \mu^\star(y))(\mu^\star(y + \Delta_y) - \mu^\star(y)) \geq 0 \text{ for all } \Delta_y$$

Using Taylor's theorem from previous slide:

$$D_2F(x, \mu^\star(y), t)\left(D\mu^\star(y)\Delta_y + o(\|\Delta_y\|)\right) \geq 0 \text{ for all } \Delta_y$$

which means that:

$$D_2F(x, \mu^\star(y), t)D\mu^\star(y) = 0 \quad .$$

Lemma 3.3.1: Simplifying total derivatives

Going back to our problem

$$DG(y) = D_1F(y, \mu^\star(y)) + D_2F(y, \mu^\star(y))D\mu^\star(y) \quad .$$

But since

$$D_2F(x, \mu^\star(y), t)D\mu^\star(y) = 0 \quad .$$

It follows that:

$$DG(y) = D_1F(y, \mu^\star(y)) \quad .$$

Back to HJB

$$0 = \min_{u \in \mathcal{U}} \{c(x,u) + D_2J(x,t) + D_1J(x,t)f(x,u)\} \quad ,$$

which we express using the optimal control function $\mu^\star$:

$$0 = c(x, \mu^\star(x,t)) + D_2J(x,t) + D_1J(x,t)f(x, \mu^\star(x,t)) \triangleq G(x,t) \quad .$$

HJB

Differentiating G in both arguments and using lemma 3.3.1:

$$D_1G(x,t) = D_1c(x,\mu^\star(x,t)) + D_1D_2J(x,t) + D_1^2J(x,t)f(x,\mu^\star(x,t)) + D_1f(x,\mu^\star(x,t))D_1J(x,t)$$

Similarly:

$$0 = D_2^2J(x,t) + D_2D_1J(x,t)f(x,\mu^\star(x,t))$$

With some re-writing of the above along $\dot{x}(t) = f(x(t),u(t))$, we have:

$$D_t\left(D_1J(x(t),t)\right) = D_1D_2J(x(t),t) + D_1^2J(x(t),t)f(x(t),u)$$
$$D_t\left(D_2J(x(t),t)\right) = D_2^2J(x(t),t) + D_1D_2J(x(t),t)f(x(t),u)$$

Adjoint equation

Denoting:

$$\lambda(t) = D_1 J(x,t)$$
$$\lambda(t_0) = D_2 J(x,t)$$

We can then write:

$$D_1 G(x,t) = D_1 c(x, \mu^\star(x,t)) + D_1 D_2 J(x,t) + D_1^2 J(x,t) f(x, \mu^\star(x,t)) + D_1 f(x, \mu^\star(x,t)) D_1 J(x,t)$$

as:

$$\dot{\lambda}(t) = -D_1 f(x(t), u(t))\lambda(t) - D_1 c(x(t), u(t)) \quad .$$