AA203: Optimal and Learning-based Control
Course Notes

James Harrison*

April 18, 2019

3 The HJB and HJI Equations

In this section, we will extend the ideas of dynamic programming to the continuous time setting. Restating the continuous time optimal control problem, we assume dynamics

$$\dot{\boldsymbol{x}}(t) = f(\boldsymbol{x}(t), \boldsymbol{u}(t), t) \tag{1}$$

and cost

$$J(\boldsymbol{x}(0)) = c_f(\boldsymbol{x}(t_f), t_f) + \int_0^{t_f} c(\boldsymbol{x}(\tau), \boldsymbol{u}(\tau), \tau) d\tau. \tag{2}$$

where t_f is fixed.

3.1 The Principle of Optimality in Continuous Time

3.1.1 Hamilton-Jacobi-Bellman

As in the discrete time principle of optimality, consider the tail problem

$$J(\boldsymbol{x}(t), \{\boldsymbol{u}(\tau)\}_{\tau=t}^{t_f}, t) = c_f(\boldsymbol{x}(t_f), t_f) + \int_t^{t_f} c(\boldsymbol{x}(\tau), \boldsymbol{u}(\tau), \tau) d\tau \tag{3}$$

where $t \leq t_f$ and $\boldsymbol{x}(t)$ is an admissible state value. The optimal solution to this tail problem comes from the functional minimization

$$J^*(\boldsymbol{x}(t), t) = \min_{\{\boldsymbol{u}(\tau)\}_{\tau=t}^{t_f}} \left\{ c_f(\boldsymbol{x}(t_f), t_f) + \int_t^{t_f} c(\boldsymbol{x}(\tau), \boldsymbol{u}(\tau), \tau) d\tau \right\}. \tag{4}$$

*Contact: jharrison@stanford.edu

Note, then, that due to the additivity of cost we can split the problem up over time,

$$J^*(\boldsymbol{x}(t),t) = \min_{\{\boldsymbol{u}(\tau)\}_{\tau=t}^{t_f}} \left\{ \int_t^{t+\Delta t} c(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau),\tau)d\tau + c_f(\boldsymbol{x}(t_f),t_f) + \int_{t+\Delta t}^{t_f} c(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau),\tau)d\tau \right\} \tag{5}$$

which by applying the principle of optimality to the tail cost,

$$J^*(\boldsymbol{x}(t),t) = \min_{\{\boldsymbol{u}(\tau)\}_{\tau=t}^{t+\Delta t}} \left\{ \int_t^{t+\Delta t} c(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau),\tau)d\tau + J^*(\boldsymbol{x}(t+\Delta t),t+\Delta t) \right\}. \tag{6}$$

Let $J_t^*(\boldsymbol{x}(t),t) = \nabla_t J^*(\boldsymbol{x}(t),t)$ and $J_{\boldsymbol{x}}^*(\boldsymbol{x}(t),t) = \nabla_{\boldsymbol{x}} J^*(\boldsymbol{x}(t),t)$. Taylor expanding, we have

$$\begin{aligned} J^*(\boldsymbol{x}(t),t) = \min_{\{\boldsymbol{u}(\tau)\}_{\tau=t}^{t+\Delta t}} \{ & c(\boldsymbol{x}(t),\boldsymbol{u}(t),t)\Delta t + J^*(\boldsymbol{x}(t),t) + (J_t^*(\boldsymbol{x}(t),t))\Delta t \\ & + (J_{\boldsymbol{x}}^*(\boldsymbol{x}(t),t))^T(\boldsymbol{x}(t+\Delta t) - \boldsymbol{x}(t)) + o(\Delta t)\} \end{aligned} \tag{7}$$

for small Δt. The first term is a result of Taylor expanding the integral and applying the fundamental theorem of calculus. Note that we can pull $J^*(\boldsymbol{x}(t),t)$ out of the minimization over cost, as this quantity will not vary under different choices of future actions. Dividing through by Δt and taking the limit $\Delta t \to 0$, we obtain the *Hamilton-Jacobi-Bellman* equation

$$0 = J_t^*(\boldsymbol{x}(t),t) + \min_{\boldsymbol{u}(t)} \left\{ c(\boldsymbol{x}(t),\boldsymbol{u}(t),t) + (J^*(\boldsymbol{x}(t),t))^T f(\boldsymbol{x}(t),\boldsymbol{u}(t),t) \right\} \tag{8}$$

with terminal condition

$$J^*(\boldsymbol{x}(t_f),t_f) = c_f(\boldsymbol{x}(t_f),t_f). \tag{9}$$

For convenience, we will define the Hamiltonian

$$\mathcal{H}(\boldsymbol{x}(t),\boldsymbol{u}(t),J_{\boldsymbol{x}}^*,t) := c(\boldsymbol{x}(t),\boldsymbol{u}(t),t) + (J_{\boldsymbol{x}}^*(\boldsymbol{x}(t),t))^T f(\boldsymbol{x}(t),\boldsymbol{u}(t),t) \tag{10}$$

which allow us to compactly write the HJB equation as

$$0 = J_t^*(\boldsymbol{x}(t),t) + \min_{\boldsymbol{u}(t)} \{\mathcal{H}(\boldsymbol{x}(t),\boldsymbol{u}(t),J_{\boldsymbol{x}}^*,t)\}. \tag{11}$$

The HJB equation is a partial differential equation that, for cost-to-go $J^*(\boldsymbol{x}(t),t)$, will satisfy all time-state pairs $(\boldsymbol{x}(t),t)$. The previous informal derivation assumed differentiability of $J^*(\boldsymbol{x}(t),t)$, which we do not know a priori. This assumption is rectified by the following theorem on solutions to the HJB equation.

Theorem 3.1 (Sufficiency Theorem). *Suppose $V(\boldsymbol{x},t)$ is a solution to the HJB equation, that $V(\boldsymbol{x},t)$ is C^1 in $\boldsymbol{x}$ and t, and that*

$$\begin{aligned} 0 &= V_t(\boldsymbol{x},t) + \min_{\boldsymbol{u}\in\mathcal{U}} \left\{ c(\boldsymbol{x},\boldsymbol{u},t) + (V_{\boldsymbol{x}}(\boldsymbol{x},t))^T f(\boldsymbol{x},\boldsymbol{u},t) \right\} \\ V(\boldsymbol{x},t_f) &= c_f(\boldsymbol{x},t_f) \ \forall \boldsymbol{x} \end{aligned}$$

Suppose also that $\pi^*(\boldsymbol{x}, t)$ attains the minimum in this equation for all t and $\boldsymbol{x}$. Let $\{\boldsymbol{x}^*(t) \mid t \in [t_0, t_f]\}$ be the state trajectory obtained from the given initial condition $\boldsymbol{x}(0)$ when the control trajectory $\boldsymbol{u}^*(t) = \pi^*(\boldsymbol{x}^*(t), t), t \in [t_0, t_f]$ is used. Then V is equal to the optimal cost-to-go function, i.e.,

$$V(\boldsymbol{x}, t) = J^*(\boldsymbol{x}, t) \; \forall \, \boldsymbol{x}, t. \tag{12}$$

Furthermore, the control trajectory $\{\boldsymbol{u}^*(t) \mid t \in [t_0, t_f]\}$ is optimal..

Proof. [Ber12], Volume 1, Section 3.2. □

3.1.2 Continuous-Time LQR

As a useful result of the HJB equations, we will derive LQR in continuous time. We aim to minimize

$$J(\boldsymbol{x}(0)) = \frac{1}{2}\boldsymbol{x}^T(t_f)Q_f\boldsymbol{x}(t_f) + \frac{1}{2}\int_0^{t_f} \boldsymbol{x}^T(t)Q(t)\boldsymbol{x}(t) + \boldsymbol{u}^T(t)R(t)\boldsymbol{u}(t)dt \tag{13}$$

subject to dynamics

$$\dot{\boldsymbol{x}}(t) = A(t)\boldsymbol{x}(t) + B(t)\boldsymbol{u}(t). \tag{14}$$

As in discrete LQR, we will assume $Q_f, Q(t)$ are positive semidefinite, and $R(t)$ is positive definite. We will also assume t_f is fixed, and the state and action are unconstrained.

We will write the Hamiltonian,

$$\mathcal{H} = \frac{1}{2}\boldsymbol{x}^T(t)Q(t)\boldsymbol{x}(t) + \frac{1}{2}\boldsymbol{u}^T(t)R(t)\boldsymbol{u}(t) + J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)^T(A(t)\boldsymbol{x}(t) + B(t)\boldsymbol{u}(t)) \tag{15}$$

which yeilds necessary optimality conditions

$$0 = \nabla_{\boldsymbol{u}}\mathcal{H} = R(t)\boldsymbol{u}(t) + B^T(t)J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t). \tag{16}$$

Since $\nabla^2_{\boldsymbol{uu}}\mathcal{H} = R(t) > 0$, the control that satisfies the necessary conditions is the global minimizer. Rearranging, we have

$$\boldsymbol{u}^*(t) = -R^{-1}(t)B^T(t)J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t) \tag{17}$$

which we can plug back into the Hamiltonian to yield

$$\begin{aligned}
\mathcal{H} = {} & \frac{1}{2}\boldsymbol{x}^T(t)Q(t)\boldsymbol{x}(t) + \frac{1}{2}J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)^T B(t)R^{-1}(t)B^T(t)J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t) && (18)\\
& + J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)^T A(t)\boldsymbol{x}(t) - J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)^T B(t)R^{-1}(t)B^T(t)J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t) \\
= {} & \frac{1}{2}\boldsymbol{x}^T(t)Q(t)\boldsymbol{x}(t) - \frac{1}{2}J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)^T B(t)R^{-1}(t)B^T(t)J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t) + J^*_{\boldsymbol{x}}(\boldsymbol{x}(t), t)^T A(t)\boldsymbol{x}(t). && (19)
\end{aligned}$$

This gives the HJB equation

$$
\begin{aligned}
0 = J_t^*(\boldsymbol{x}(t), t) + \frac{1}{2}\boldsymbol{x}^T(t)Q(t)\boldsymbol{x}(t) - \frac{1}{2}J_{\boldsymbol{x}}^*(\boldsymbol{x}(t), t)^T B(t)R^{-1}(t)B^T(t)J_{\boldsymbol{x}}^*(\boldsymbol{x}(t), t) \\
+ J_{\boldsymbol{x}}^*(\boldsymbol{x}(t), t)^T A(t)\boldsymbol{x}(t)
\end{aligned}
\tag{20}
$$

with boundary condition

$$
J^*(\boldsymbol{x}(t_f), t_f) = \frac{1}{2}\boldsymbol{x}^T(t_f)Q_f\boldsymbol{x}(t_f). \tag{21}
$$

It may appear as if we are stuck here, as this form of the HJB doesn't immediately yield $J^*(\boldsymbol{x}(t), t)$. Armed with the knowledge that the discrete time LQR problem has a quadratic cost-to-go, we will cross our fingers and guess a solution of the form

$$
J^*(\boldsymbol{x}(t), t) = \frac{1}{2}\boldsymbol{x}^T(t)V(t)\boldsymbol{x}(t). \tag{22}
$$

Substituting, we have

$$
\begin{aligned}
0 = \frac{1}{2}\boldsymbol{x}^T(t)\dot{V}(t)\boldsymbol{x}(t) + \frac{1}{2}\boldsymbol{x}^T(t)Q(t)\boldsymbol{x}(t) \\
- \frac{1}{2}\boldsymbol{x}^T(t)V(t)B(t)R^{-1}(t)B^T(t)V(t)\boldsymbol{x}(t) + \boldsymbol{x}^T(t)V(t)A(t)\boldsymbol{x}(t)
\end{aligned}
\tag{23}
$$

Note that we will decompose

$$
\boldsymbol{x}^T(t)V(t)A(t)\boldsymbol{x}(t) = \frac{1}{2}\boldsymbol{x}^T(t)V(t)A(t)\boldsymbol{x}(t) + \frac{1}{2}\boldsymbol{x}^T(t)A^T(t)V(t)\boldsymbol{x}(t) \tag{24}
$$

which yields

$$
0 = \frac{1}{2}\boldsymbol{x}^T(t)\left(\dot{V}(t) + Q(t) - V(t)B(t)R^{-1}(t)B^T(t)V(t) + V(t)A(t) + A^T(t)V(t)\right)\boldsymbol{x}(t). \tag{25}
$$

This equation must hold for all $\boldsymbol{x}(t)$, so

$$
-\dot{V}(t) = Q(t) - V(t)B(t)R^{-1}(t)B^T(t)V(t) + V(t)A(t) + A^T(t)V(t) \tag{26}
$$

with boundary condition $V(t_f) = Q_f$.

Therefore, the HJB PDE has been reduced to a set of matrix ordinary differential equations (the Riccati equation). This is integrated backwards in time to find the full control policy as a function of time. One we have found $V(t)$, the control policy is

$$
\boldsymbol{u}^*(t) = -R^{-1}(t)B^T(t)V(t)\boldsymbol{x}(t). \tag{27}
$$

Similarly to the discrete case, the feedback gains tend toward constant in the limit of the infinite horizon problem, under some technical assumptions.

3.2 Differential Games

3.2.1 Differential Games and Information Patterns

3.2.2 Hamilton-Jacobi-Isaacs

3.2.3 Reachability

3.3 Further Reading

References

[Ber12] Dimitri P Bertsekas. *Dynamic programming and optimal control*. Number 1. 4 edition, 2012.