CHAPTER 10

Analytical Optimal Control with the Hamilton-Jacobi-Bellman Sufficiency Theorem

10.1 INTRODUCTION

10.1.1 Dynamic Programming in Continuous Time

Discrete time problems permit a simple derivation of dynamic programming. Indeed, for the numerical studies in the next chapter, and for digital (sampled-data) control systems in real robotics applications, the discrete-time treatment might be enough. However, for analytical studies it is often easier, and even more compact, to work directly in continuous time.

The Hamilton-Jacobi-Bellman Equation.
Let's develop the continuous time form of the cost-to-go function recursion by taking the limit as the time between control updates goes to zero.

$$\begin{aligned} J^*(\mathbf{x},T) =& h(\mathbf{x}) \\ J^*(\mathbf{x},t) =& \min_{[\mathbf{u}(t)\ldots\mathbf{u}(T)]} \left[h(\mathbf{x}(T)) + \int_t^T g(\mathbf{x}(t),\mathbf{u}(t))dt \right], \quad \mathbf{x}(t) = \mathbf{x}, \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x},\mathbf{u}) \\ =& \lim_{dt\rightarrow 0} \min_{\mathbf{u}} \left[g(\mathbf{x},\mathbf{u})dt + J(\mathbf{x}(t+dt), t+dt) \right] \\ \approx& \lim_{dt\rightarrow 0} \min_{\mathbf{u}} \left[g(\mathbf{x},\mathbf{u})dt + J^*(\mathbf{x},t) + \frac{\partial J^*}{\partial \mathbf{x}}\dot{\mathbf{x}}dt + \frac{\partial J^*}{\partial t}dt \right] \end{aligned}$$

Simplifying, we are left with

$$\boxed{0 = \min_{\mathbf{u}} \left[g(\mathbf{x},\mathbf{u}) + \frac{\partial J^*}{\partial \mathbf{x}}\mathbf{f}(\mathbf{x},\mathbf{u}) + \frac{\partial J^*}{\partial t} \right].} \tag{10.1}$$

This equation is well-known as the Hamilton-Jacobi-Bellman (HJB) equation.

Sufficiency theorem. The HJB equation assumes that the cost-to-go function is continuously differentiable in $\mathbf{x}$ and t, which is not necessarily the case. It therefore cannot be satisfied in all optimal control problems. It does, however, provide a sufficient condition for optimality.

Suppose we have found a policy, $\pi(\mathbf{x}, t)$, and a cost-to-go function, $J^\pi(\mathbf{x}, t)$. Suppose that π minimizes the right-hand-side of the HJB for all $\mathbf{x}$ and all $t \in [0, T]$, and that this minimum is zero for all $\mathbf{x}$ and all $t \in [0, T]$. Furthermore, suppose that $J^\pi(\mathbf{x}, T) = h(\mathbf{x})$. Then we have that

$$J^\pi(\mathbf{x}, t) = J^*(\mathbf{x}, t), \quad \pi(\mathbf{x}, t) = \pi^*(\mathbf{x}, t).$$

A more formal treatment of this theorem, including it's proof, can be found in [13].

The HJB provides limited utility for constructing optimal policies and value functions, but does provide a relatively inexpensive way to verify optimality if one is able to "guess" a solution.

Examples. The best way to get started using the HJB sufficiency theorem is to work through some examples.

EXAMPLE 10.1 First-order Regulator

Let's reconsider on of the first problems from this chapter:

$$\dot{x} = u, \quad |u| \le 1,$$
$$h(x) = \frac{1}{2}x^2, \quad g(x, u, t) = 0.$$

Our candidate for an optimal policy was simply $\pi(x, t) = -\operatorname{sgn}(x)$. To prove that this policy is optimal, we must first evaluate the policy to obtain J^π. The dynamics here are particularly simple (x moves one unit per second with maximal actuation), and we can quickly see that executing policy π from $x(0)$ will produce:

$$x(T) = \begin{cases} x(0) - T & x(0) > T \\ 0 & -T \le x(0) \le T \\ x(0) + T & x(0) < -T. \end{cases}$$

A similar expression can be written, involving $T - t$, for all starting times $t < T$. Therefore, we have

$$J^\pi(x, t) = \begin{cases} \frac{1}{2}\left[x - (T - t)\right]^2 & x > (T - t) \\ 0 & -(T - t) \le x \le (T - t) \\ \frac{1}{2}\left[x + (T - t)\right]^2 & x < -(T - t). \end{cases}.$$

Notice that, although the expression is written piecewise, this value function is continuously differentiable in both x and t. It can be summarized as

$$J^\pi(x, t) = \max\{0, \frac{1}{2}\left[|x| - (T - t)\right]^2\},$$

and it's derivatives are ...

$$\frac{\partial J^\pi}{\partial x} = \operatorname{sgn}(x)\max\{0, |x| - (T - t)\}$$
$$\frac{\partial J^\pi}{\partial t} = \max\{0, |x| - (T - t)\}.$$

Substituting into the HJB, we have

$$\begin{aligned} 0 &= \min_u [0 + \text{sgn}(x) u \max\{0, |x| - (T-t)\} + \max\{0, |x| - (T-t)\}] \\ &= \min_u [(1 + \text{sgn}(x)u) \max\{0, |x| - (T-t)\}]. \end{aligned}$$

As we hoped, the policy $u = -\text{sgn}(x)$ does minimize this quantity, and the minimum is zero. By the sufficiency theorem, $\pi(x,t) = -\text{sgn}(x)$ is an optimal policy.

EXAMPLE 10.2 The Linear Quadratic Regulator (LQR)

The linear quadratic regulator is clearly the most important and influential result in optimal control theory to date. Consider systems governed by an LTI state-space equation of the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u},$$

and a finite-horizon cost function with

$$\begin{aligned} h(\mathbf{x}) &= \mathbf{x}^T \mathbf{Q}_f \mathbf{x}, \quad \mathbf{Q}_f = \mathbf{Q}_f^T \geq \mathbf{0} \\ g(\mathbf{x}, \mathbf{u}, t) &= \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{u}^T \mathbf{R} \mathbf{u}, \quad \mathbf{Q} = \mathbf{Q}^T \geq 0, \mathbf{R} = \mathbf{R}^T > 0 \end{aligned}$$

Writing the HJB, we have

$$0 = \min_{\mathbf{u}} \left[\mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{u}^T \mathbf{R} \mathbf{u} + \frac{\partial J^*}{\partial \mathbf{x}} (\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}) + \frac{\partial J^*}{\partial t} \right].$$

Due to the positive definite quadratic form on $\mathbf{u}$, we can find the minimum by setting the gradient to zero:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{u}} &= 2\mathbf{u}^T \mathbf{R} + \frac{\partial J^*}{\partial \mathbf{x}} \mathbf{B} = 0 \\ \mathbf{u}^* = \pi^*(\mathbf{x}, t) &= -\frac{1}{2} \mathbf{R}^{-1} \mathbf{B}^T \frac{\partial J^*}{\partial \mathbf{x}}^T \end{aligned}$$

In order to proceed, we need to investigate a particular form for the cost-to-go function, $J^*(\mathbf{x}, t)$. Let's try a solution of the form:

$$J^*(\mathbf{x}, t) = \mathbf{x}^T \mathbf{S}(t) \mathbf{x}, \quad \mathbf{S}(t) = \mathbf{S}^T(t) > \mathbf{0}.$$

In this case we have

$$\frac{\partial J^*}{\partial \mathbf{x}} = 2\mathbf{x}^T \mathbf{S}(t), \quad \frac{\partial J*}{\partial t} = \mathbf{x}^T \dot{\mathbf{S}}(t) \mathbf{x},$$

and therefore

$$\begin{aligned} \mathbf{u}^* = \pi^*(\mathbf{x}, t) &= -\mathbf{R}^{-1} \mathbf{B}^T \mathbf{S}(t) \mathbf{x} \\ 0 &= \mathbf{x}^T \left[\mathbf{Q} - \mathbf{S}(t) \mathbf{B} \mathbf{R}^{-1} \mathbf{B}^T \mathbf{S}(t) + 2\mathbf{S}(t) \mathbf{A} + \dot{\mathbf{S}}(t) \right] \mathbf{x}. \end{aligned}$$

All of the matrices here are symmetric, except for $\mathbf{S}(t)\mathbf{A}$. But since $\mathbf{x}^T \mathbf{S}(t) \mathbf{A} \mathbf{x} = \mathbf{x}^T \mathbf{A}^T \mathbf{S}(t) \mathbf{x}$, we can equivalently write the symmetric form (which we assumed):

$$0 = \mathbf{x}^T \left[\mathbf{Q} - \mathbf{S}(t) \mathbf{B} \mathbf{R}^{-1} \mathbf{B}^T \mathbf{S}(t) + \mathbf{S}(t) \mathbf{A} + \mathbf{A}^T \mathbf{S}(t) + \dot{\mathbf{S}}(t) \right] \mathbf{x}.$$

Therefore, $\mathbf{S}(t)$ must satisfy the condition (known as the continuous time Riccati equation):

$$\dot{\mathbf{S}}(t) = -\mathbf{S}(t)\mathbf{A} - \mathbf{A}^T\mathbf{S}(t) + \mathbf{S}(t)\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}(t) - \mathbf{Q},$$

and the terminal condition

$$\mathbf{S}(T) = \mathbf{Q}_f.$$

Since we were able to satisfy the HJB with the minimizing policy, we have met the sufficiency condition, and have found the optimal policy and optimal cost-to-go function.

The dependence on time in both the value function and the feedback policy might surprise readers who have used LQR before. The more common usage is actually the infinite-horizon version of the problem, which we will develop in Example 6.

EXAMPLE 10.3 Linear Quadratic Optimal Tracking

For completeness, we consider a slightly more general form of the linear quadratic regulator. The standard LQR derivation attempts to drive the system to zero. Consider now the problem:

$$\begin{gathered}
\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \\
h(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^d(T))^T\mathbf{Q}_f(\mathbf{x} - \mathbf{x}^d(T)), \quad \mathbf{Q}_f = \mathbf{Q}_f^T \geq 0 \\
g(\mathbf{x},\mathbf{u},t) = (\mathbf{x} - \mathbf{x}^d(t))^T\mathbf{Q}(\mathbf{x} - \mathbf{x}^d(t)) + (\mathbf{u} - \mathbf{u}^d(t))^T\mathbf{R}(\mathbf{u} - \mathbf{u}^d(t)), \\
\mathbf{Q} = \mathbf{Q}^T \geq 0, \mathbf{R} = \mathbf{R}^T > 0
\end{gathered}$$

Now, guess a solution

$$J^*(\mathbf{x},t) = \mathbf{x}^T\mathbf{S_2}(t)\mathbf{x} + \mathbf{x}^T\mathbf{s_1}(t) + s_0(t), \quad \mathbf{S_2}(t) = \mathbf{S_2}^T(t) > \mathbf{0}.$$

In this case, we have

$$\frac{\partial J^*}{\partial \mathbf{x}} = 2\mathbf{x}^T\mathbf{S_2}(t) + \mathbf{s_1}^T(t), \quad \frac{\partial J^*}{\partial t} = \mathbf{x}^T\dot{\mathbf{S_2}}(t)\mathbf{x} + \mathbf{x}^T\dot{\mathbf{s_1}}(t) + \dot{s_0}(t).$$

Using the HJB,

$$0 = \min_{\mathbf{u}}\left[(\mathbf{x} - \mathbf{x}^d(t))^T\mathbf{Q}(\mathbf{x} - \mathbf{x}^d(t)) + (\mathbf{u} - \mathbf{u}^d(t))^T\mathbf{R}(\mathbf{u} - \mathbf{u}^d(t)) + \frac{\partial J^*}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}) + \frac{\partial J^*}{\partial t}\right],$$

we have

$$\begin{gathered}
\frac{\partial}{\partial \mathbf{u}} = 2(\mathbf{u} - \mathbf{u}^d(t))^T\mathbf{R} + (2\mathbf{x}^T\mathbf{S_2}(t) + \mathbf{s_1}^T(t))\mathbf{B} = 0, \\
\mathbf{u}^*(t) = \mathbf{u}^d(t) - \mathbf{R}^{-1}\mathbf{B}^T\left[\mathbf{S_2}(t)\mathbf{x} + \frac{1}{2}\mathbf{s_1}(t)\right]
\end{gathered}$$

The HJB can be satisfied by integrating backwards

$$\begin{aligned}
-\dot{\mathbf{S_2}}(t) =& \mathbf{Q} - \mathbf{S_2}(t)\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S_2}(t) + \mathbf{S_2}(t)\mathbf{A} + \mathbf{A}^T\mathbf{S_2}(t) \\
-\dot{\mathbf{s_1}}(t) =& -2\mathbf{Q}\mathbf{x}^d(t) + \left[\mathbf{A}^T - \mathbf{S}_2\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\right]\mathbf{s_1}(t) + 2\mathbf{S_2}(t)\mathbf{B}\mathbf{u}^d(t) \\
-\dot{s_0}(t) =& \mathbf{x}^d(t)^T\mathbf{Q}\mathbf{x}^d(t) - \frac{1}{4}\mathbf{s_1}^T(t)\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{s_1}(t) + \mathbf{s_1}(t)^T\mathbf{B}\mathbf{u}^d(t),
\end{aligned}$$

from the final conditions

$$\begin{aligned}\mathbf{S_2}(T) =& \mathbf{Q}_f \\ \mathbf{s_1}(T) =& -2\mathbf{Q}_f \mathbf{x}^d(T) \\ s_0(T) =& \left[\mathbf{x}^d(T)\right]^T \mathbf{Q}_f \left[\mathbf{x}^d(T)\right].\end{aligned}$$

Notice that the solution for $\mathbf{S_2}$ is the same as the simpler LQR derivation, and is symmetric (as we assumed). Note also that $s_0(t)$ has no effect on control (even indirectly), and so can often be ignored.

A quick observation about the quadratic form, which might be helpful in debugging. We know that $J(x,t)$ must be uniformly positive. This is true iff $\mathbf{S}_2 > 0$ and $s_0 > \frac{1}{4}\mathbf{s}_1^T \mathbf{S}_2^{-1} \mathbf{s}_1$, which comes from evaluating the function at x_{min} defined by $\frac{\partial}{\partial x} = 0$.

EXAMPLE 10.4 Minimum-time Linear Quadratic Regulator

$$\begin{gathered}\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \\ h(\mathbf{x}) = \mathbf{x}^T \mathbf{Q}_f \mathbf{x}, \quad \mathbf{Q}_f = \mathbf{Q}_f^T \geq 0 \\ g(\mathbf{x}, \mathbf{u}, t) = 1 + \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{u}\mathbf{R}\mathbf{u}, \\ \mathbf{Q} = \mathbf{Q}^T \geq 0, \mathbf{R} = \mathbf{R}^T > 0\end{gathered}$$

EXAMPLE 10.5 Linear Final Boundary Value Problems

The finite-horizon LQR formulation can be used to impose a strict final boundary value condition by setting an infinite $\mathbf{Q}_f$. However, integrating the Riccati equation backwards from an infinite initial condition isn't very practical. To get around this, let us consider solving for $\mathbf{P}(t) = \mathbf{S}(t)^{-1}$. Using the matrix relation $\frac{d\mathbf{S}^{-1}}{dt} = -\mathbf{S}^{-1}\frac{d\mathbf{S}}{dt}\mathbf{S}^{-1}$, we have:

$$-\dot{\mathbf{P}}(t) = -\mathbf{P}(t)\mathbf{Q}\mathbf{P}(t) + \mathbf{B}\mathbf{R}^{-1}\mathbf{B} - \mathbf{A}\mathbf{P}(t) - \mathbf{P}(t)\mathbf{A}^T,$$

with the final conditions

$$\mathbf{P}(T) = 0.$$

This Riccati equation can be integrated backwards in time for a solution.

It is very interesting, and powerful, to note that, if one chooses $\mathbf{Q} = 0$, therefore imposing no position cost on the trajectory before time T, then this inverse Riccati equation becomes a linear ODE which can be solved explicitly. These relationships are used in the derivation of the controllability Grammian, but here we use them to design a feedback law.

10.2 INFINITE-HORIZON PROBLEMS

Limit as $T \rightarrow \infty$. Value function converges (or blows up).

Examples and motivations.

10.2.1 The Hamilton-Jacobi-Bellman

For infinite-horizon, all dependence of J on t drops out, and the HBJ reduces to:

$$0 = \min_{\mathbf{u}} \left[g(\mathbf{x}, \mathbf{u}) + \frac{\partial J}{\partial \mathbf{x}} \mathbf{f}(\mathbf{x}, \mathbf{u}) \right].$$

10.2.2 Examples

EXAMPLE 10.6 The Infinite-Horizon Linear Quadratic Regulator

Consider again a system in LTI state space form,

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u},$$

but now consider the infinite-horizon cost function given by

$$g(\mathbf{x}, \mathbf{u}) = \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{u}^T \mathbf{R} \mathbf{u}, \quad \mathbf{Q} = \mathbf{Q}^T \geq \mathbf{0}, \mathbf{R} = \mathbf{R}^T > 0.$$

Since we know that the optimal value function cannot depend on time in the infinite-horizon case, we will guess the form:

$$J^*(\mathbf{x}) = \mathbf{x}^T \mathbf{S} \mathbf{x}.$$

This yields the optimal policy

$$\mathbf{u}^* = \pi^*(\mathbf{x}) = -\mathbf{R}^{-1} \mathbf{B}^T \mathbf{S} \mathbf{x} = -\mathbf{K}\mathbf{x}.$$

In fact, this form can also be marched through the HJB and verified. The solution, not surprisingly, is the steady-state solution of the Riccati equation described in the finite-horizon problem. Setting $\dot{\mathbf{S}}(t) = 0$, we have

$$0 = \mathbf{S}\mathbf{A} + \mathbf{A}^T \mathbf{S} - \mathbf{S}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T \mathbf{S} + \mathbf{Q}.$$

Note that this equation is not linear in $\mathbf{S}$, and therefore solving the equation is non-trivial (but well understood). Both the optimal policy and optimal value function are available from MATLAB by calling

```
[K,S]=lqr(A,B,Q,R).
```

It is worth examining the form of the optimal policy more closely. Since the value function represents cost-to-go, it would be sensible to move down this landscape as quickly as possible. Indeed, $-\mathbf{S}\mathbf{x}$ is in the direction of steepest descent of the value function. However, not all directions are possible to achieve in state-space. $-\mathbf{B}^T\mathbf{S}\mathbf{x}$ represents precisely the projection of the steepest descent onto the control space, and is the steepest descent achievable with the control inputs $\mathbf{u}$. Finally, the pre-scaling by the matrix $\mathbf{R}^{-1}$ biases the direction of descent to account for relative weightings that we have placed on the different control inputs. Note that although this interpretation is straight-forward, the slope that we are descending (in the value function, $\mathbf{S}$) is a complicated function of the dynamics and cost.

EXAMPLE 10.7 Second-order Quadratic Regulator

Let's use the LQR result to solve for the optimal regulator of our trivial linear system:

$$\ddot{q} = u.$$

EXAMPLE 10.8 The Overdamped Pendulum

PROBLEMS

10.1. (CHALLENGE) *Optimal control of the simple pendulum.*

Find the optimal policy for the minimum-time problem on the simple pendulum described by the dynamics:

$$\ddot{q} = u - \sin(q).$$