OPTIMAL CONTROL SYSTEMS

DYNAMIC PROGRAMMING

Harry G. Kwatny

Department of Mechanical Engineering & Mechanics
Drexel University

Outline

Discrete Time
- Problem Definition
- Recursive Solution
- Example

Continuous Time
- Problem Definition
- The HJB Equation
- Example
- Summary
- Simple Examples

Linear Quadratic Regulator
- Basic Results

Problem Setup

Consider the system

$$x_{k+1} = f\left(x_k, u_k\right), \quad x \in R^n, u \in U \subset R^m$$

on the discrete time interval $k = 0, 1, \ldots, N-1$. A feedback policy is a sequence of functions

$$\pi = \left\{\mu_0\left(x_0\right), \mu_1\left(x_1\right), \ldots, \mu_{N-1}\left(x_{N-1}\right)\right\}$$

such that $u_k = \mu_k\left(x_k\right)$. The problem is to find a policy that minimizes the cost

$$J_\pi\left(x_0\right) = g_N\left(x_N\right) + \sum_{k=0}^{N-1} g_k\left(x_k, \mu_k\left(x_k\right)\right)$$

Principle of Optimality

The optimal cost is

$$J_{\pi^*}\left(x_0\right) = \min_{\pi \in \Pi} J_\pi\left(x_0\right)$$

and the optimal policy π^* is one that satisfies

$$J_{\pi^*}\left(x_0\right) \leq J_\pi\left(x_0\right) \quad \forall \pi \in \Pi$$

Theorem (Principle of Optimality)

Suppose $\pi^* = \left\{\mu_1^*, \ldots, \mu_{N-1}^*\right\}$ *is an optimal policy. Then the subpolicy* $\pi_i^* = \left\{\mu_i^*, \ldots, \mu_{N-1}^*\right\}$, $1 \leq i \leq, N-1$, *is optimal with respect to the cost function*

$$J_{\pi_i}\left(x_i\right) = g_N\left(x_N\right) + \sum_{k=i}^{N-1} g_k\left(x_k, \mu_k\left(x_k\right)\right)$$

PRINCIPLE OF OPTIMALITY - PROOF

The argument is based on contradiction as follows.

- Suppose $\pi^* = \{\mu_1^*, \ldots, \mu_{N-1}^*\}$ is an optimal policy with cost $J_{\pi^*}(x_0)$
- Clearly, we can write the cost

$$J_{\pi^*}(x_0) = \sum_{k=0}^{i} g_k\left(x_k^*, \mu_k^*(x_k)\right) + J_{\pi_i^*}(x_i^*)$$

- Suppose there exists an alternate subpolicy $\pi_i = \{\mu_i, \ldots, \mu_{N-1}\}$ such that $J_{\pi_i}(x_i^*) < J_{\pi_i^*}(x_i^*)$.
- Then

$$J_{\pi^*}(x_0) = \sum_{k=0}^{i} g_k\left(x_k^*, \mu_k^*(x_k)\right) + J_{\pi_i^*}(x_i^*) > \sum_{k=0}^{i} g_k\left(x_k^*, \mu_k^*(x_k)\right) + J_{\pi_i}(x_i^*)$$

so π_i^* is not optimal'

THE DYNAMIC PROGRAMMING RECURSION

DEFINITION (COST TO GO)

Denote the optimal cost of a trajectory beginning in state x at time i as $V(x, i)$. $V(x, i)$ is the cost to go'

The *principle of optimality* implies the recursion

$$V(x_{i-1}, i-1) = \min_{\pi_{i-1} \in \Pi} J_{\pi_{i-1}}(x_{i-1}) = \min_{\pi_{i-1} \in \Pi_{i-1}} \left\{ g_N(x_N) + \sum_{k=i-1}^{N-1} g_k(x_k, \mu_k(x_k)) \right\}$$
$$= \min_{\mu_{i-1}} \{ g_{i-1}(x_{i-1}, \mu_{i-1}(x_{i-1})) + V(x_i, i) \}$$

This equation is the basis for a recursive computation of the optimal policy.

- Step 1. Solve the single stage optimization problem with $i = N$

$$V(x_{N-1}, N-1) = \min_{\mu_{N-1}} \{ g_{N-1}(x_{N-1}, \mu_{N-1}(x_{N-1})) + V(x_N, N) \}$$

- Step 2. Apply the recursion successively for $i = N-1, i = N-2, \ldots$

Two Terminal Cases

1. Case 1. x_N is fixed and $g(x_N) \equiv 0$. In this case $V(x_N, N) = 0$. We also have the constraint
$$x_N = f(x_{N-1}, \mu_{N-1})$$
We must assume that with x_N specified there are solution pairs (x_{N-1}, μ_{N-1}). Otherwise, the problem is not well posed because x_N is not reachable. Then
$$V(x_{N-1}, N-1) = \min_{\mu_{N-1}} \{g_{N-1}(x_{N-1}, \mu_{N-1})\}$$
where the minimization is carried out with respect to the constraint.
2. Case 2. x_N is free and $V(x_N, N) = g_N(x_N)$. Now
$$\begin{aligned} V(x_{N-1}, N-1) &= \min_{\mu_{N-1}} \{g_{N-1}(x_{N-1}, \mu_{N-1}) + V(x_N, N)\} \\ &= \min_{\mu_{N-1}} \{g_{N-1}(x_{N-1}, \mu_{N-1}) + V(f(x_{N-1}, \mu_{N-1}), N)\} \end{aligned}$$

Continuing the Recursion

Once the pair $\mu_{N-1}, V\left(x_{N-1}, N-1\right)$ is obtained, compute the pair $\mu_{N-2}, V\left(x_{N-2}, N-2\right)$ from

$$
\begin{aligned}
V\left(x_{N-2}, N-2\right) &= \min_{\mu_{N-2}} \left\{g_{N-2}\left(x_{N-2}, \mu_{N-2}\left(x_{N-2}\right)\right) + V\left(x_{N-1}, N-1\right)\right\} \\
&= \min_{\mu_{N-2}} \left\{g_{N-2}\left(x_{N-2}, \mu_{N-2}\left(x_{N-2}\right)\right) + V\left(f\left(x_{N-2}, \mu_{N-2}\right), N-1\right)\right\}
\end{aligned}
$$

Continuing in this way:

$$
V\left(x_{N-i}, N-i\right) = \min_{\mu_{N-i}} \left\{g_{N-i}\left(x_{N-i}, \mu_{N-i}\left(x_{N-i}\right)\right) + V\left(f\left(x_{N-i}, \mu_{N-i}\right), N-i+1\right)\right\}
$$

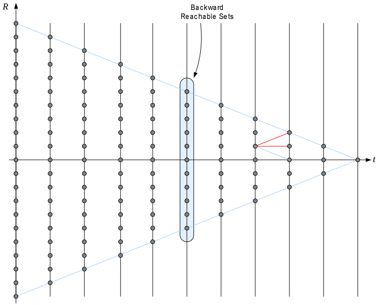
EXAMPLE

Consider a system

$$x_{i+1} = x_i + u, \quad u \in \{-1, 0, 1\}$$

on the time interval $0 \leq k \leq 10$. We wish to steer the system from arbitrary initial state to the origin in such a way as to minimize the cost

$$J = \sum_{k=0}^{10} x_k^2$$



Example, Cont'd

Values for 'cost to go'

$$\{181, 145, 113, 85, 61, 41, 25, 13, 5, 1, 0, 1, 5, 13, 25, 41, 61, 85, 113, 145, 181\}$$
$$\{145, 113, 85, 61, 41, 25, 13, 5, 1, 0, 1, 5, 13, 25, 41, 61, 85, 113, 145\}$$
$$\{113, 85, 61, 41, 25, 13, 5, 1, 0, 1, 5, 13, 25, 41, 61, 85, 113\}$$
$$\{85, 61, 41, 25, 13, 5, 1, 0, 1, 5, 13, 25, 41, 61, 85\}$$
$$\{61, 41, 25, 13, 5, 1, 0, 1, 5, 13, 25, 41, 61\}$$
$$\{41, 25, 13, 5, 1, 0, 1, 5, 13, 25, 41\}$$
$$\{25, 13, 5, 1, 0, 1, 5, 13, 25\}$$
$$\{13, 5, 1, 0, 1, 5, 13\}$$
$$\{5, 1, 0, 1, 5\}$$
$$\{1, 0, 1\}$$

Problem Setup

Consider the system

$$\dot{x} = f(x, u), \quad x \in R^n, u \in U \subset R^m$$
$$x(0) = x_0$$

over the time interval $t \in [0, T]$, where $T > 0$ is not necessarily fixed. The cost to be minimized is

$$J(u(\cdot)) = \ell(x(T)) + \int_0^T L(x(t), u(t), t)\, dt$$

Now, suppose that t is an arbitrary time in the interval $0 \leq t \leq T$, and consider the more general optimization problem starting at time t in arbitrary state x and cost function

$$J(u(\cdot); x, t) = \ell(x(T)) + \int_t^T L(x(t), u(t), t)\, dt$$

Problem Definition

Definition (Feedback Policy)

A feedback policy π is a family of functions $\mu_t(x)$, $0 \leq t \leq T$ such that $u(t) = \mu_t(x(t))$. An optimal policy is one that minimizes the cost

$$J_\pi(x_0) = \ell(x(T)) + \int_0^T L(x(t), \mu_t(x(t)), t)\, dt$$

The optimal policy satisfies

$$J_{\pi^*}(x_0) \leq J_\pi(x_0), \quad \forall \pi \in \Pi$$

Principle of Optimality

Theorem (Principle of Optimality)

Suppose $\pi^* = \{\mu_t^*(x) \,|0 \leq t \leq T\}$ *is an optimal policy. Then the subpolicy* $\pi_s^* = \{\mu_t^*(x) \,|0 < s \leq t \leq T\}$ *is optimal with respect to the cost*

$$J_{\pi_s}(x_s) = \ell(x(T)) + \int_s^T L(x(t), \mu_t(x(t)), t)\, dt$$

Definition (Cost to Go)

For $x \in R^n, 0 \leq t \leq T$ define the cost to go to be the minimum cost if we start in state x at time t:

$$V(x,t) \triangleq \min_{u \subset U} J(u; x, t)$$

The Hamilton-Jacobi-Bellman Equation

$$V(x,t) = \min_{\pi_t}\left\{\ell(x(T)) + \int_t^T L(x(\tau), \mu_\tau(x(\tau)), \tau)\, d\tau\right\}$$

Divide the interval $[t, T]$ to obtain

$$V(x,t) = \min_{\pi_t}\left\{\begin{array}{l} \int_t^{t+\Delta t} L(x(\tau), \mu_\tau(x(\tau)), \tau)\, d\tau \\ \quad +\ell(x(T)) + \int_{t+\Delta t}^T L(x(\tau), \mu_\tau(x(\tau)), \tau)\, d\tau \end{array}\right\}$$

The principle of optimality implies

$$V(x,t) = \min_{\substack{u(\tau) \\ t \le \tau \le t+\Delta t}} \left\{\int_t^{t+\Delta t} L(x(\tau), u(\tau), \tau)\, d\tau + V(x(t+\Delta t), t+\Delta t)\right\}$$

$$V(x,t) = \min_u \{L(x,u,t)\Delta t + V(x,t) + V_x(x,t) f(x,u)\Delta t + V_t(x,t)\Delta t + o(\Delta t)\}$$

$$\Downarrow$$

$$0 = V_t(x,t)\Delta t + \min_u \{L(x,u,t) + V_x(x,t) f(x,u)\}\Delta t + o(\Delta t)$$

The HJB Equation, Cont'd

Define

$$H^*(x, \lambda, t) = \min_u \{L(x, u, t) + \lambda f(x, u)\}$$

Then we obtain the Hamilton-Jacobi-Bellman equation

$$\frac{\partial}{\partial t} V(x, t) + H^* \left(x, \frac{\partial}{\partial x} V(x, t), t \right) = 0$$

with boundary condition

$$V(x, T) = \ell(x)$$

The optimal control is given by

$$u^*(x, t) = \arg\min_u \{L(x, u, t) + V_x(x, t) f(x, u)\}$$

Example 1

Consider the dynamics

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = u$$

on the time interval $0 \leq t \leq 50$, and cost

$$J = x_1^2(50) + \int_0^{50} \frac{1}{2}\left(x_1^2(t) + u^2(t)\right) dt$$

Now compute

$$H(x, \lambda, u, t) = \frac{1}{2}\left(x_1^2(t) + u^2(t)\right) + \lambda_1 x_2 + \lambda_2 u$$

which implies

$$u^*(x, \lambda) = -\lambda_2, \quad H^*(x, \lambda) = \frac{1}{2}\left(x_1^2 + 2\lambda_1 x_2 - \lambda_2^2\right)$$

Example 1, Cont'd

The HJB equation is

$$V_t(x,t) + \frac{1}{2}\left(x_1^2 + 2V_{x_1}(x,t)\,x_2 - V_{x_2}^2(x,t)\right) = 0$$

Assume a solution of the form

$$V(x,t) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} s_{11}(t) & s_{12}(t) \\ s_{12}(t) & s_{22}(t) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

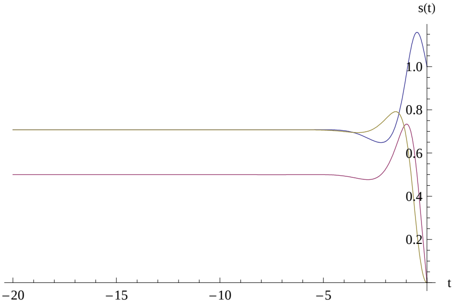
Substitution in HJB leads to

$$\begin{aligned} &\dot{s}_{11} - 2s_{12} + \tfrac{1}{2} = 0 \\ &\dot{s}_{22} - 2s_{22}^2 + 2s_{12} = 0 \\ &\dot{s}_{12} + \tfrac{1}{4}\left(4s_{11} - 8s_{12}s_{22}\right) = 0 \end{aligned}$$

and

$$u^*(x,t) = -\lambda_2 = \frac{\partial}{\partial x_2}V(x,t) = -2s_{12}(t)\,x_1 - 2s_{22}(t)\,x_2$$

Example 1, Cont'd



Notice that

$$\{s_{11}(-50), s_{12}(-50), s_{22}(-50)\} = \{0.707107, 0.5, 0.707107\}$$

and

$$eig\begin{bmatrix} 0 & 1 \\ -2\times 0.5 & -2\times 0.707107 \end{bmatrix} = -0.707107 \pm i\, 0.707107$$

THEOREM

Consider the continuous dynamical system

$$\dot{x} = f(x, u), \quad x \in R^n, u \in U \subset R^m$$
$$x(0) = x_0$$

with cost function $J(u(\cdot)) = \ell(x(T)) + \int_0^T L(x(t), u(t), t)\, dt$
The minimum cost optimal control is

$$u^*(x, t) = \arg\min_u \{L(x, u, t) + V_x(x, t) f(x, u)\}$$

where $V_x(x, t)$ is the *cost to go* which satisfies the HJB equation

$$\frac{\partial}{\partial t} V(x, t) + H^*\left(x, \frac{\partial}{\partial x} V(x, t), t\right) = 0, \quad V(x, T) = \ell(x)$$

and $H^*(x, \lambda, t) = \min_u \{L(x, u, t) + \lambda f(x, u)\}$

Linear Quadratic Regulator

Consider the linear dynamics

$$\dot{x} = Ax + Bu$$
$$y = Cx$$

and cost

$$J(u, w) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \left[y^T(t) y(t) + \rho u^T(t) u(t) \right] dt$$

Suppose

- (A, B) is stabilizable
- (A, C) is detectable

Then

$$u(t) = -\frac{1}{\rho} B^T S x(t)$$

$$A^T S + SA - S \left(\frac{1}{\rho} BB^T \right) S = -C^T C$$

EXAMPLE 2

Consider, once again, the system

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = u$$

However, suppose u is constrained

$$|u| \leq 1$$

and our goal is to reach the origin in minimum time

$$J = \int_0^T dt = T$$

In this case, the 'cost to go' from any initial state x is independent of t, Thus, the HJB equations is simply

$$H^*\left(x, \frac{\partial}{\partial x}V\left(x\right), t\right) = 0$$

Example 2, Cont'd

Now,

$$H = 1 + \lambda_1 x_2 + \lambda_2 u$$

and

$$u^* = \arg\min_u \{1 + \lambda_1 x_2 + \lambda_2 u\}$$
$$\Rightarrow u^* = -\text{sgn}\lambda_2 = -\text{sgn}\frac{\partial}{\partial x_2} V(x_1, x_2)$$

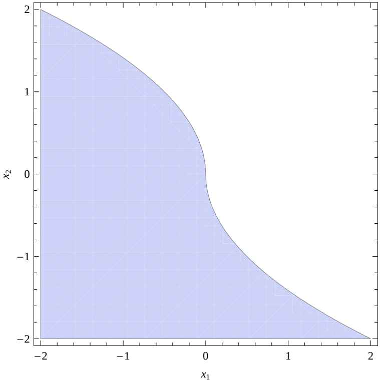
Thus, the HJB equation is

$$x_2 V_{x_1}(x_1, x_2) - |V_{x_2}(x_1, x_2)| + 1 = 0, \quad V(0,0) = 0$$

We can verify the solution

$$V(x_1, x_2) = \begin{cases} -x_2 - 2\left(x_1 + \frac{1}{2}x_2^2\right)^{1/2} & x_1 \geq -\frac{1}{2}x_2|x_2| \\ x_2 - 2\left(-x_1 + \frac{1}{2}x_2^2\right)^{1/2} & x_1 \leq -\frac{1}{2}x_2|x_2| \end{cases}$$

Example 2, Cont'd



$$u^* = \begin{cases} -1 & white \\ 1 & blue \end{cases}$$

LQR Problem Setup

Consider the linear time-invariant system:

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x \in R^n, u \in R^m$$

with cost function

$$J = \frac{1}{2} x^T(t_f) Q_f x(t_f) + \frac{1}{2}\int_0^{t_f} \left\{ x^T Q x + u^T R u \right\} dt$$

Q and Q_f are symmetric and positive semi-definite. R is symmetric positive definitite.

OPEN LOOP SOLUTION

The Hamiltonian is

$$H(x,u,\lambda) = \frac{1}{2}\left(x^T Q x + u^T R u\right) + \lambda^T (Ax + Bu)$$

from which

$$u^* = -R^{-1}B^T\lambda, \quad H^*(x,\lambda) = \frac{1}{2}x^T Q x + \lambda^T A x - \frac{1}{2}\lambda^T B R^{-1} B^T \lambda$$

Combining the system and adjoint equations yields

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix}, \quad x(0) = x_0, \lambda(t_f) = Q_f x(t_f)$$

solving the two-point boundary value problem provides the open loop optimal control

$$u^*(t) = -R^{-1}B^T\lambda(t)$$

Closed Loop Solution

To obtain the closed loop control, solve the HJB equation for the cost-to-go function $V(x,t)$

$$\frac{\partial V}{\partial t} + \frac{1}{2}x^T Q x + \frac{\partial V}{\partial x} A x - \frac{1}{2}\frac{\partial V}{\partial x} B R^{-1} B^T \left(\frac{\partial V}{\partial x}\right)^T = 0, \quad V(x, t_f) = \frac{1}{2}x^T Q_f x$$

Assume $V(x,t)$ takes the form

$$V(x,t) = \frac{1}{2}x^T S(t) x \Rightarrow \frac{\partial V}{\partial t} = \frac{1}{2}x^T \dot{S}(t) x, \quad \frac{\partial V}{\partial x} = x^T S(t)$$

with symmatric S. Then the HJB reduces to

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

Closed Loop – Summary

The optimal feedback control is

$$u^*(x,t) = -R^{-1}B^T\lambda = -R^{-1}B^T S(t)\,x$$

where $S(t)$ satisfies the Riccati equation

$$\dot{S}(t) + Q + \left(S(t)A + A^T S(t)\right) - S(t)BR^{-1}B^T S(t) = 0$$

In the event that $T \to \infty$, $S(t) \to S_0$, a constant, symmetric matrix that satisfies the algebraic Riccati equation (ARE)

$$S_0 A + A^T S_0 - S_0 B R^{-1} B^T S_0 = -Q$$

The optimal feedback control is

$$u^*(x) = -Kx, \quad K = R^{-1}B^T S_0$$

Hamiltonian Matrix

Consider the Hamiltonian matrix:

$$H = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix}$$

Notice that the ARE can be written

$$-Q - A^T S_0 = S_0 \left[A - BR^{-1}B^T S_0\right]$$

Direct computation leads to

$$H \begin{bmatrix} I & 0 \\ S_0 & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ S_0 & I \end{bmatrix} \begin{bmatrix} A - BR^{-1}B^T S_0 & -BR^{-1}B^T \\ 0 & -\left(A - BR^{-1}B^T S_0\right)^T \end{bmatrix}$$

Thus, H is similar to a block diagonal matrix. The $2n$ eigenvalues of H composed of the n eigenvalues of $\left(A - BR^{-1}B^T S_0\right)$ and the n eigenvalues of $-\left(A - BR^{-1}B^T S_0\right)$

Hamiltonian Matrix, Cont'd

- If S_0 is a stabilizing solution of the ARE, then all eigenvalues of $\left(A - BR^{-1}B^T S_0\right)$ are in the open left half plane and all eigenvalues of $-\left(A - BR^{-1}B^T S_0\right)$ are their mirror image in the open right half plane.
- In this case H has no eigenvalues on the imaginary axis.

Algebraic Riccati Equation

Recall the system and adjoint equations in Hamilton's form

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = H \begin{bmatrix} x \\ \lambda \end{bmatrix}, \quad H = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix}$$

Let, $X_1, X_2 \in R^{n \times n}$ be two matrices with X_1 invertible, such that the n subspace $V \subset R^{2n}$

$$V = \mathrm{Im} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

is H-invariant. Invariance implies that there exists a matrix Λ such that

$$H \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \Lambda$$

$$H \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} X_1^{-1} = \begin{bmatrix} I \\ X \end{bmatrix} X_1 \Lambda X_1^{-1}, \quad X = X_2 X_1^{-1}$$

ARE Cont'd

Premultiply by $\begin{bmatrix} -X & I \end{bmatrix}$ to obtain

$$\begin{bmatrix} -X & I \end{bmatrix} \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix} \begin{bmatrix} I \\ X \end{bmatrix} = \begin{bmatrix} -X & I \end{bmatrix} \begin{bmatrix} I \\ X \end{bmatrix} X_1 \Lambda X_1^{-1} = 0$$

Expand and rearrange to obtain

$$XA + A^T X - XBR^{-1}B^T X = -Q$$

Consequently, X as constructed above satisfies the ARE. It all begins with constructing V. To do this select a set of n eigenvalue-eigenvector pairs. While there are $\begin{pmatrix} 2n \\ n \end{pmatrix}$ ways of doing this, there is only one way with all eigenvalues in the left half plane. Thus, only one stabilizing solution to the ARE.