哈爾濱工業大學 深圳研究生院
Harbin Institute of Technology Shenzhen Graduate School

Optimization and Optimal Control

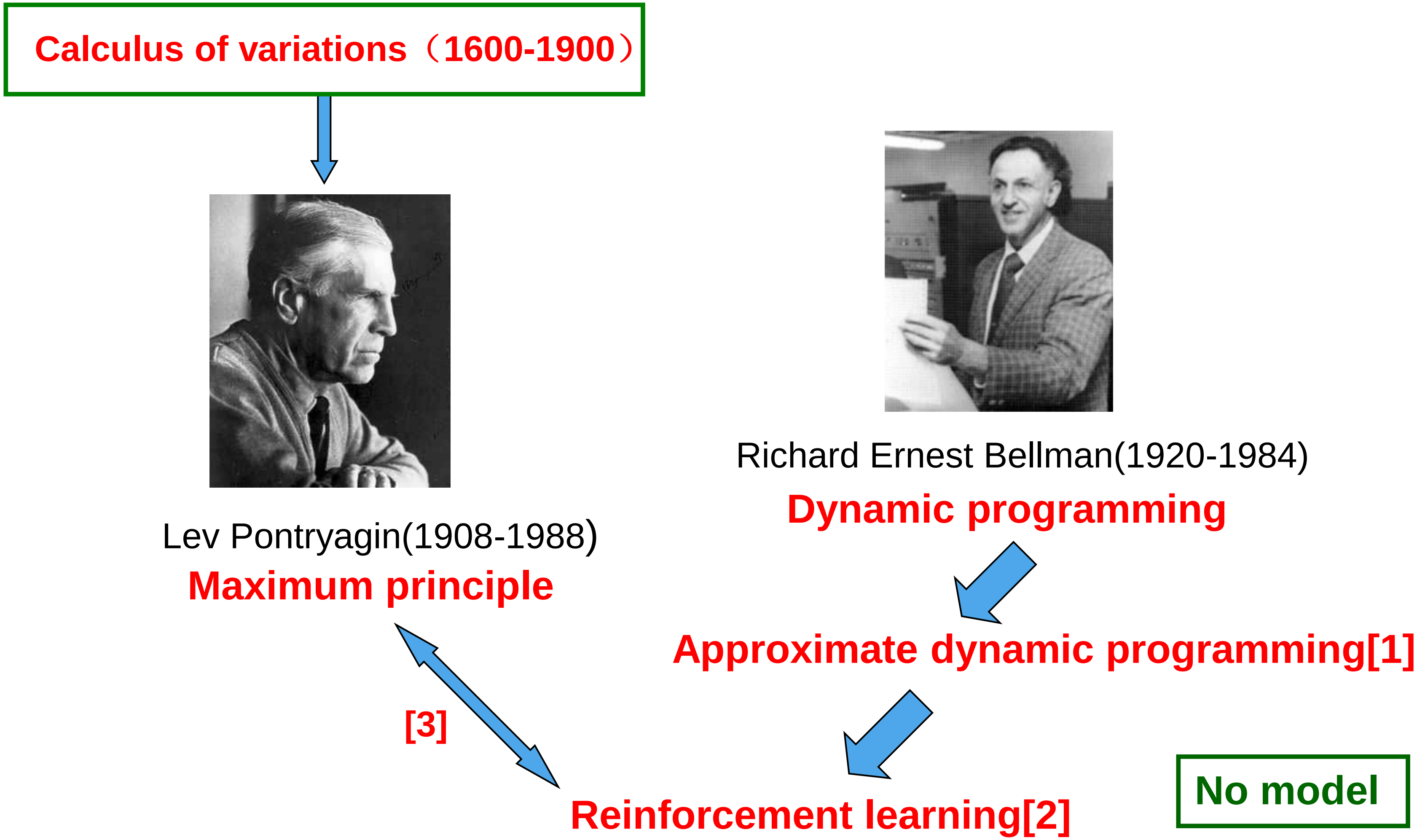
Introduction to Dynamic Programming

Instructor: Yanjie Li (李衍杰)
Office: G1011
Email: autolyj@hit.edu.cn

Content

- History and present of optimal control
- Principle of optimality
- Shorted path problem
- Solve optimal control by dynamic programming

History and Present



Calculus of variations（1600-1900）

Lev Pontryagin(1908-1988)

Maximum principle

Richard Ernest Bellman(1920-1984)

Dynamic programming

Approximate dynamic programming[1]

[3]

Reinforcement learning[2]

No model

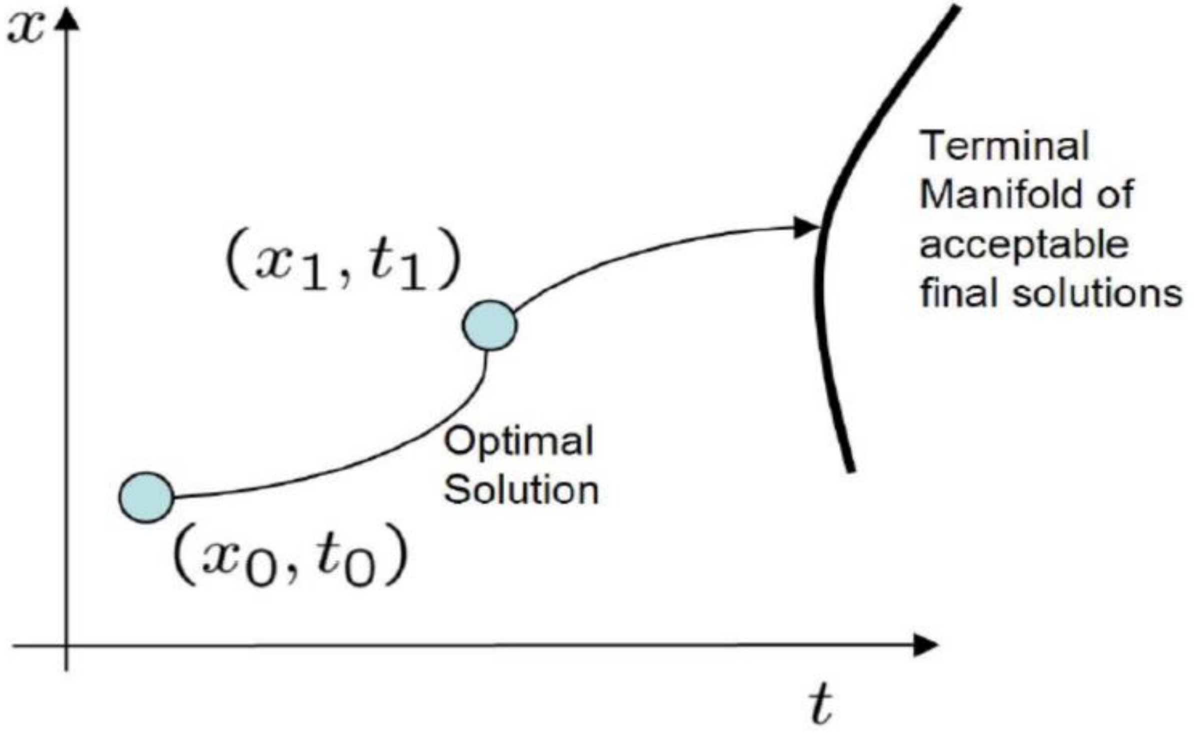
[1] D. P. Bestsekas, Dynamic Programming and Optimal Control, Athena Scientific, 2011.

[2] R. S. Sutton and A. G. Barto, Reinforcement learning: an Introduction, MIT Press, 1998.

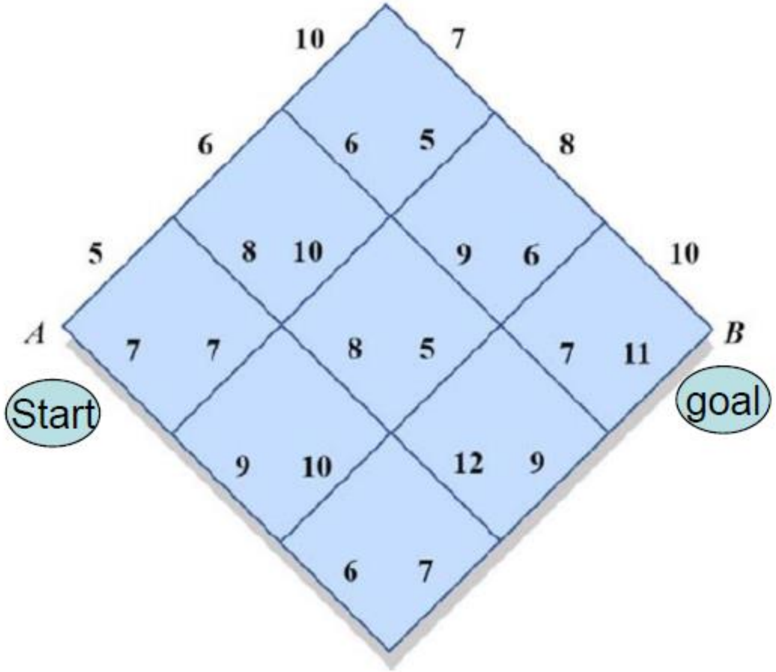
[3]P. Mehta and S. Meyn, Q-learning and Pontryagin's Minimum Principle, CDC 2009.

Principle of Optimality

Principle of Optimality: Suppose the optimal solution for a problem passes through some intermediate point (x_1, t_1), then the optimal solution to the same problem starting at (x_1, t_1) must be the continuation of the same path.



Shortest Path Problem



- There are 20 options to get from A to B – could evaluate each and compute travel time, but that would be pretty tedious

Shortest Path Problem

- **Alternative approach:** Start at B and work backwards, invoking the principle of optimality along the way.
  - First step backward can be either up (10) or down (11)

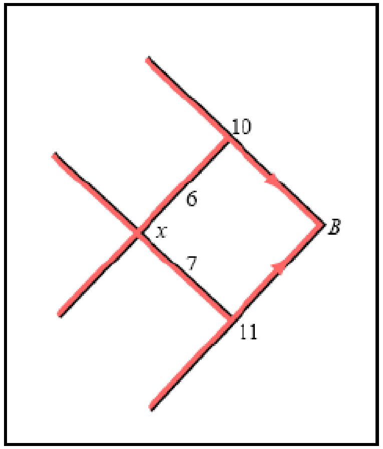
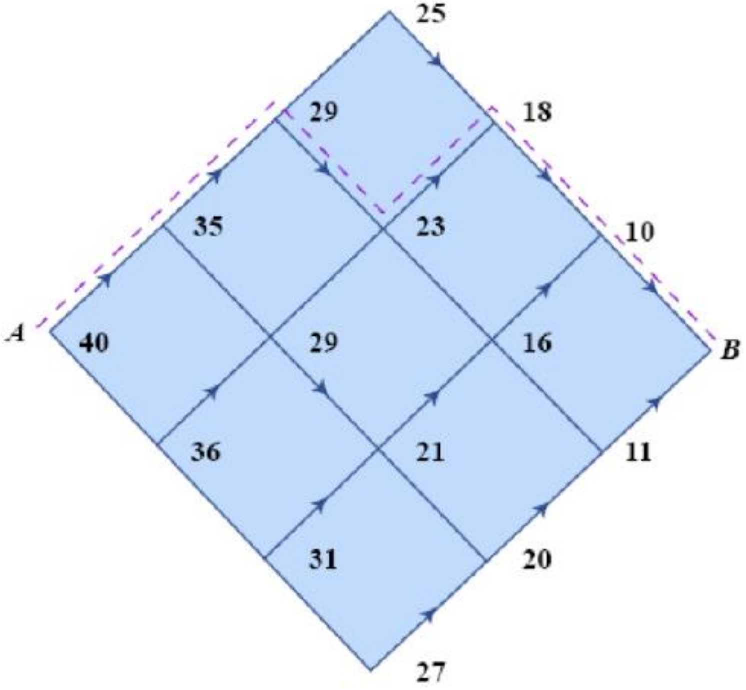


Figure by MIT OpenCourseWare.

- Consider the travel time from point x
  - Can go up and then down $6 + 10 = 16$

Shortest Path Problem

- Repeat process for all other points, until finally get to initial point ⇒ shortest path traced by moving in the directions of the arrows.



- **Key advantage** is that only had to find 15 numbers to solve this problem this way rather than evaluate the travel time for 20 paths

Basic Idea

- Examples show the basis of dynamic programming and use of principle of optimality.
  - In general, if there are numerous options at location α that next lead to locations $x_1, \ldots, x_n$, choose the action that leads to

$$J^\star_{\alpha h} = \min_{x_i} \left\{ [J_{\alpha x_1} + J^\star_{x_1 h}], [J_{\alpha x_2} + J^\star_{x_2 h}], \ldots, [J_{\alpha x_n} + J^\star_{x_n h}] \right\}$$

- Can apply the same process to more general control problems. Typically have to assume something about the system state (and possible control inputs), e.g., bounded, but also discretized.

Optimal Control

- Consider the problem of minimizing:

$$\min J = h(x(t_f)) + \int_{t_0}^{t_f} g(x(t), u(t), t)\, dt$$

subject to

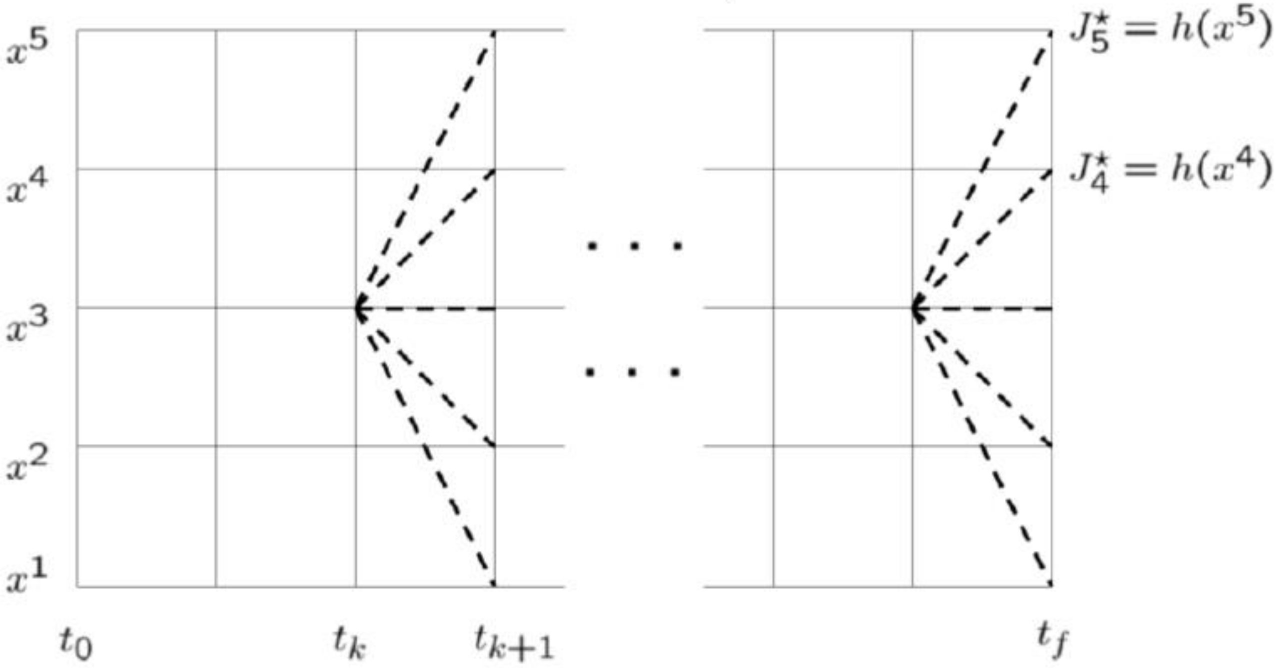
$$\begin{aligned} \dot{x} &= a(x, u, t) \\ x(t_0) &= \text{fixed} \\ t_f &= \text{fixed} \end{aligned}$$

– Other constraints on $x(t)$ and $u(t)$ can also be included.

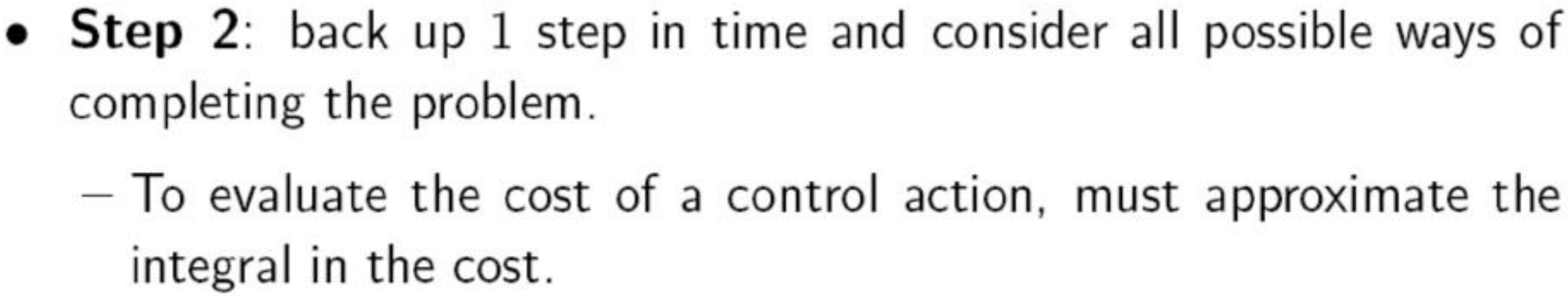
Optimal Control

- **Step 1** of solution approach is to develop a grid over space/time.
  - Then look at possible final states $x_i(t_f)$ and evaluate final costs
  - For example, can discretize the state into 5 possible cases $x^1, \ldots, x^5$

$$J_i^\star = h(x_{t_f}^i) \ , \ \forall \ i$$



Optimal Control



- **Step 2**: back up 1 step in time and consider all possible ways of completing the problem.
  - To evaluate the cost of a control action, must approximate the integral in the cost.

Optimal control

- Consider the scenario where you are at state x^i at time t_k, and apply control u_k^{ij} to move to state x^j at time $t_{k+1} = t_k + \Delta t$.
  - Approximate cost is

$$\int_{t_k}^{t_{k+1}} g(x(t), u(t), t))\, dt \approx g(x_k^i, u_k^{ij}, t_k)\Delta t$$

  - Can solve for control inputs directly from system model:

$$x_{k+1}^j \approx x_k^i + a(x_k^i, u_k^{ij}, t_k)\Delta t \qquad \Rightarrow a(x_k^i, u_k^{ij}, t_k) = \frac{x_{k+1}^j - x_k^i}{\Delta t}$$

    which can be solved to find u_k^{ij}.

  - Process is especially simple if the control inputs are affine:

$$\dot{x} = f(x,t) + q(x,t)u$$

    which gives

$$u_k^{ij} = q(x_k^i, t_k)^{-1} \left[\frac{x_{k+1}^j - x_k^i}{\Delta t} - f(x_k^i, t_k) \right]$$

Optimal control

- So for any combination of x_k^i and x_{k+1}^j can evaluate the incremental cost $\Delta J(x_k^i, x_{k+1}^j)$ of making this state transition
- Assuming already know the optimal path from each new terminal point (x_{k+1}^j), can establish optimal path to take from x_k^i using

$$J^\star(x_k^i, t_k) = \min_{x_{k+1}^j} \left[\Delta J(x_k^i, x_{k+1}^j) + J^\star(x_{k+1}^j) \right]$$

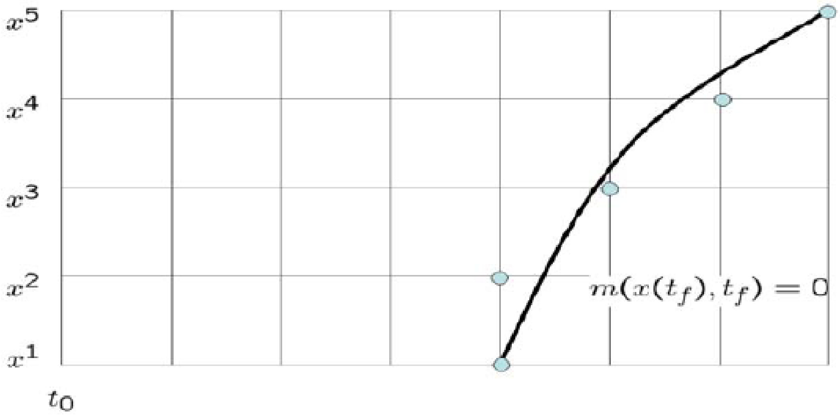
  - Then for each x_k^i, output is:
    - ◇ Best x_{k+1}^i to pick, because it gives lowest cost
    - ◇ Control input required to achieve this best cost.
- Then work backwards in time until you reach x_{t_0}, when only one value of x is allowed because of the given initial condition.

Optimal control

- Extends to free end time problems, where

$$\min J = h(x(t_f), t_f) + \int_{t_0}^{t_f} g(x(t), u(t), t))\,dt$$

with some additional constraint on the final state $m(x(t_f), t_f) = 0$.



  - Gives group of points that (approximately) satisfy the terminal constraint
  - Can evaluate cost for each, and work backwards from there.

Optimal control

- Process extends to higher dimensional problems where the state is a vector.
  - Just have to define a grid of points in $\mathbf{x}$ and t, which for two dimensions would look like:

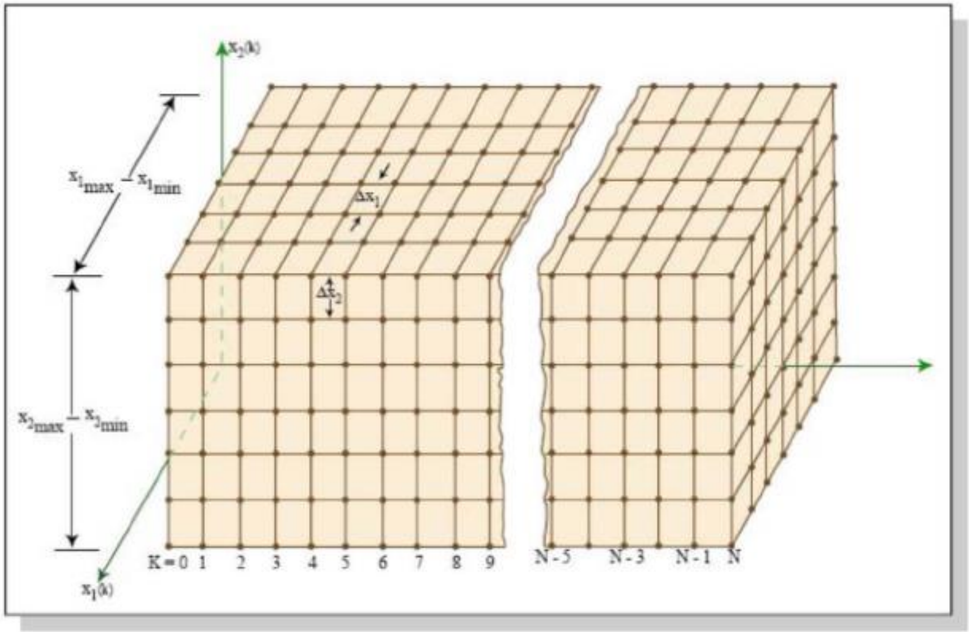


Figure by MIT OpenCourseWare.

Optimal control

- Previous formulation picked x's and used those to determine the u's.
  - For more general problems, might be better off picking the u's and using those to determine the propagated x's

$$\boxed{\begin{aligned} J^{\star}(x_k^i, t_k) &= \min_{u_k^{ij}} \left[\Delta J(x_k^i, u_k^{ij}) + J^{\star}(x_{k+1}^j, t_{k+1}) \right] \\ &= \min_{u_k^{ij}} \left[g(x_k^i, u_k^{ij}, t_k)\Delta t + J^{\star}(x_{k+1}^j, t_{k+1}) \right] \end{aligned}}$$

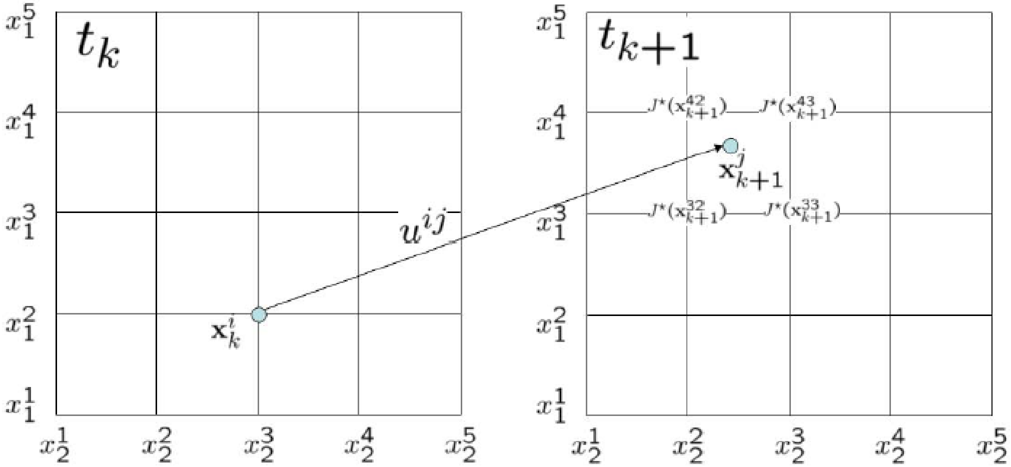
  - To do this, must quantize the control inputs as well.
  - But then likely that terminal points from one time step to the next will not lie on the state discrete points ⇒ must interpolate the cost to go between them.

Optimal control

- **Option 1:** find the control that moves the state from a point on one grid to a point on another.
- **Option 2:** quantize the control inputs, and then evaluate the resulting state for all possible inputs

$$\mathbf{x}_{k+1}^{j} = \mathbf{x}_{k}^{i} + \mathrm{a}(\mathbf{x}_{k}^{i}, \mathbf{u}_{k}^{ij}, t_k)\Delta t$$

  – Issue at that point is that $\mathbf{x}_{k+1}^{j}$ probably will not agree with the t_{k+1} grid points ⇒ must interpolate the available $J^{\star}$.

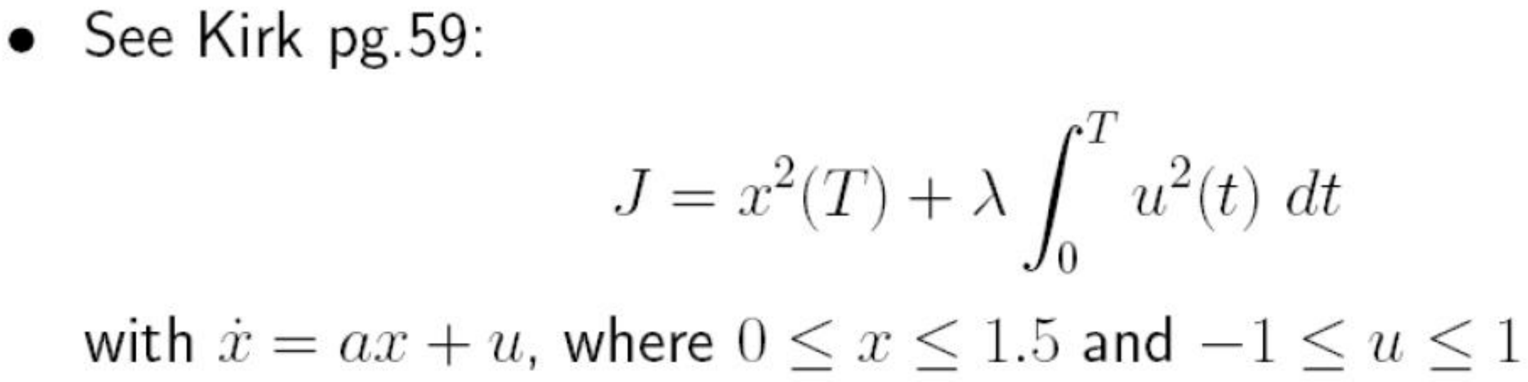


- In the expression:

$$J^\star(x_k^i, t_k) = \min_{u_k^{ij}} \left[g(x_k^i, u_k^{ij}, t_k)\Delta t + J^\star(x_{k+1}^j, t_{k+1}) \right]$$

  the term $J^\star(x_{k+1}^j, t_{k+1})$ plays the role of a "**cost-to-go**", which is a key concept in DP and other control problems.

Example



- See Kirk pg.59:

$$J = x^2(T) + \lambda \int_0^T u^2(t)\, dt$$

with $\dot{x} = ax + u$, where $0 \leq x \leq 1.5$ and $-1 \leq u \leq 1$

T=2.

Example

- Must quantize the state within the allowable values and time within the range $t \in [0, 2]$ using N=2, $\Delta t = T/N = 1$.
  - Approximate the continuous system as:

    $$\dot{x} \approx \frac{x(t + \Delta t) - x(t)}{\Delta t} = ax(t) + u(t)$$

    which gives that

    $$x_{k+1} = (1 + a\Delta t)x_k + (\Delta t)u_k$$

  - Very common discretization process (Euler integration approximation) that works well if Δt is small

Example

- Use approximate calculation from previous section – cost becomes

$$J = x^2(T) + \lambda \sum_{k=0}^{N-1} u_k^2 \Delta t$$

- Take $\lambda = 2$ and $a = 0$ to simplify things a bit.
  - With $0 \leq x(k) \leq 1.5$, take x quantized into four possible values $x_k \in \{0, 0.5, 1.0, 1.5\}$
  - With control bounded $|u(k)| \leq 1$, assume it is quantized into five possible values: $u_k \in \{-1, -0.5, 0, 0.5, 1\}$

Example

- Start – evaluate cost associated with all possible terminal states

x_2^j	$J_2^\star = h(x_2^j) = (x_2^j)^2$
0	0
0.5	0.25
1	1
1.5	2.25

- Given x_1 and possible x_2, can evaluate the control effort required to make that transition:

$u(1)$	$x_2^j = x_1^i + u(1)$			
x_1^i	**0**	**0.5**	**1**	**1.5**
0	0	0.5	1	**1.5**
0.5	-0.5	0	0.5	1
1	-1	-0.5	0	0.5
1.5	**-1.5**	-1	-0.5	0

Example

which can be used to compute the cost increments:

ΔJ_{12}^{ij}	x_2^j			
x_1^i	**0**	**0.5**	**1**	**1.5**
0	0	0.5	2	XX
0.5	0.5	0	0.5	2
1	2	0.5	0	0.5
1.5	XX	2	0.5	0

and costs at time $t = 1$ given by $J_1 = \Delta J_{12}^{ij} + J_2^\star(x_2^j)$

J_1	x_2^j			
x_1^i	**0**	**0.5**	**1**	**1.5**
0	**0**	0.75	3	XX
0.5	0.5	**0.25**	1.5	4.25
1	2	**0.75**	1	2.75
1.5	XX	2.25	**1.5**	2.25

Example

Take min across each row to determine best action at each possible $x_1 \Rightarrow J_1^\star(x_1^j)$

x_1^i		x_2^j
0	$\longrightarrow$	0
0.5	$\longrightarrow$	0.5
1	$\longrightarrow$	0.5
1.5	$\longrightarrow$	1

- Can repeat the process to find the costs at time $t = 0$ which are $J_0 = \Delta J_{01}^{ij} + J_1^\star(x_1^j)$

J_0	x_1^j			
x_0^i	**0**	**0.5**	**1**	**1.5**
0	**0**	0.75	2.75	XX
0.5	0.5	**0.25**	1.25	3.5
1	2	**0.75**	0.75	2
1.5	XX	2.25	**1.25**	1.5

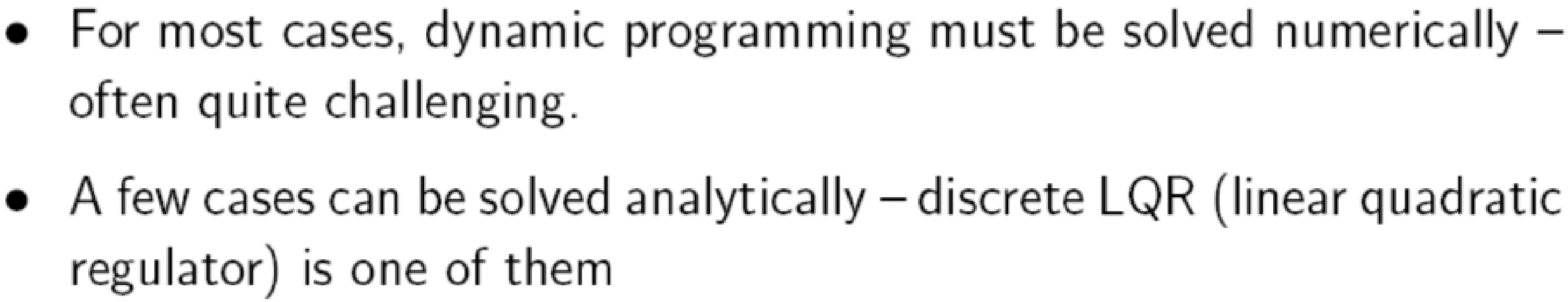
Example

and again, taking min across the rows gives the best actions:

x_0^i	$\rightarrow$	x_1^j
0	$\rightarrow$	0
0.5	$\rightarrow$	0.5
1	$\rightarrow$	0.5
1.5	$\rightarrow$	1

- So now we have a complete strategy for how to get from any x_0^i to the best x_2 to minimize the cost
  - This process can be highly automated, and this clumsy presentation is typically not needed.

Curse of Dimensionality



- For most cases, dynamic programming must be solved numerically – often quite challenging.
- A few cases can be solved analytically – discrete LQR (linear quadratic regulator) is one of them

Discrete LQR

- **Goal:** select control inputs to minimize

$$J = \frac{1}{2}\mathbf{x}_N^T H \mathbf{x}_N + \frac{1}{2}\sum_{k=0}^{N-1}[\mathbf{x}_k^T Q_k \mathbf{x}_k + \mathbf{u}_k^T R_k \mathbf{u}_k]$$

so that

$$g_d(\mathbf{x}_k, \mathbf{u}_k) = \frac{1}{2}\left(\mathbf{x}_k^T Q_k \mathbf{x}_k + \mathbf{u}_k^T R_k \mathbf{u}_k\right)$$

subject to the dynamics

$$\mathbf{x}_{k+1} = A_k \mathbf{x}_k + B_k \mathbf{u}_k$$

  – Assume that $H = H^T \geq 0$, $Q = Q^T \geq 0$, and $R = R^T > 0$

  – Including any other constraints greatly complicates problem

- Clearly $J_N^\star[\mathbf{x}_N] = \frac{1}{2}\mathbf{x}_N^T H \mathbf{x}_N \Rightarrow$ now need to find $J_{N-1}^\star[\mathbf{x}_{N-1}]$

$$\begin{aligned} J_{N-1}^\star[\mathbf{x}_{N-1}] &= \min_{\mathbf{u}_{N-1}} \{g_d(\mathbf{x}_{N-1}, \mathbf{u}_{N-1}) + J_N^\star[\mathbf{x}_N]\} \\ &= \min_{\mathbf{u}_{N-1}} \frac{1}{2} \left\{\mathbf{x}_{N-1}^T Q_{N-1} \mathbf{x}_{N-1} + \mathbf{u}_{N-1}^T R_{N-1} \mathbf{u}_{N-1} + \mathbf{x}_N^T H \mathbf{x}_N]\right\} \end{aligned}$$

- Note that $\mathbf{x}_N = A_{N-1}\mathbf{x}_{N-1} + B_{N-1}\mathbf{u}_{N-1}$, so that

$$J_{N-1}^\star[\mathbf{x}_{N-1}] = \min_{\mathbf{u}_{N-1}} \frac{1}{2} \left\{\mathbf{x}_{N-1}^T Q_{N-1} \mathbf{x}_{N-1} + \mathbf{u}_{N-1}^T R_{N-1} \mathbf{u}_{N-1} \right.$$
$$\left. + \{A_{N-1}\mathbf{x}_{N-1} + B_{N-1}\mathbf{u}_{N-1}\}^T H \{A_{N-1}\mathbf{x}_{N-1} + B_{N-1}\mathbf{u}_{N-1}\} \right\}$$

- Take derivative with respect to the control inputs

$$\frac{\partial J_{N-1}^\star[\mathbf{x}_{N-1}]}{\partial \mathbf{u}_{N-1}} = \mathbf{u}_{N-1}^T R_{N-1} + \{A_{N-1}\mathbf{x}_{N-1} + B_{N-1}\mathbf{u}_{N-1}\}^T H B_{N-1}$$

- Take transpose and set equal to 0, yields

$$\left[R_{N-1} + B_{N-1}^T H B_{N-1}\right] \mathbf{u}_{N-1} + B_{N-1}^T H A_{N-1} \mathbf{x}_{N-1} = 0$$

- Which suggests a couple of key things:
  - The best control action at time $N-1$, is a linear state feedback on the state at time $N-1$:

$$\begin{aligned}\mathbf{u}^\star_{N-1} &= -\left[R_{N-1} + B^T_{N-1} H B_{N-1}\right]^{-1} B^T_{N-1} H A_{N-1} \mathbf{x}_{N-1} \\ &\equiv -F_{N-1}\mathbf{x}_{N-1}\end{aligned}$$

  - Furthermore, can show that

$$\frac{\partial^2 J^\star_{N-1}[\mathbf{x}_{N-1}]}{\partial \mathbf{u}^2_{N-1}} = R_{N-1} + B^T_{N-1} H B_{N-1} > 0$$

    so that the stationary point is a minimum

- With this control decision, take another look at

$$
\begin{aligned}
J^\star_{N-1}[\mathbf{x}_{N-1}] &= \frac{1}{2}\mathbf{x}^T_{N-1}\left\{Q_{N-1} + F^T_{N-1}R_{N-1}F_{N-1} + \right. \\
&\quad \left. \{A_{N-1} - B_{N-1}F_{N-1}\}^T H \{A_{N-1} - B_{N-1}F_{N-1}\}\right\}\mathbf{x}_{N-1} \\
&\equiv \frac{1}{2}\mathbf{x}^T_{N-1}P_{N-1}\mathbf{x}_{N-1}
\end{aligned}
$$

  - Note that $P_N = H$, which suggests a convenient form for gain F:

$$F_{N-1} = \left[R_{N-1} + B^T_{N-1}P_N B_{N-1}\right]^{-1} B^T_{N-1}P_N A_{N-1} \qquad (3.20)$$

- Now can continue using induction – assume that at time k the control will be of the form $\mathbf{u}^\star_k = -F_k\mathbf{x}_k$ where

$$F_k = \left[R_k + B^T_k P_{k+1}B_k\right]^{-1} B^T_k P_{k+1}A_k$$

and $J^\star_k[\mathbf{x}_k] = \frac{1}{2}\mathbf{x}^T_k P_k \mathbf{x}_k$ where

$$P_k = Q_k + F^T_k R_k F_k + \{A_k - B_k F_k\}^T P_{k+1}\{A_k - B_k F_k\}$$

  - Recall that both equations are solved backwards from $k+1$ to k.

- Now consider time $k-1$, with

$$J^{\star}_{k-1}[\mathbf{x}_{k-1}] = \min_{\mathbf{u}_{k-1}} \left\{ \frac{1}{2}\mathbf{x}^T_{k-1} Q_{k-1}\mathbf{x}_{k-1} + \mathbf{u}^T_{k-1} R_{k-1}\mathbf{u}_{k-1} + J^{\star}_k[\mathbf{x}_k] \right\}$$

- Taking derivative with respect to $\mathbf{u}_{k-1}$ gives,

$$\frac{\partial J^{\star}_{k-1}[\mathbf{x}_{k-1}]}{\partial \mathbf{u}_{k-1}} = \mathbf{u}^T_{k-1} R_{k-1} + \left\{ A_{k-1}\mathbf{x}_{k-1} + B_{k-1}\mathbf{u}_{k-1} \right\}^T P_k B_{k-1}$$

so that the best control input is

$$\begin{aligned} \mathbf{u}^{\star}_{k-1} &= -\left[R_{k-1} + B^T_{k-1} P_k B_{k-1} \right]^{-1} B^T_{k-1} P_k A_{k-1}\mathbf{x}_{k-1} \\ &= -F_{k-1}\mathbf{x}_{k-1} \end{aligned}$$

- Substitute this control into the expression for $J^\star_{k-1}[\mathbf{x}_{k-1}]$ to show that

  $$J^\star_{k-1}[\mathbf{x}_{k-1}] = \frac{1}{2}\mathbf{x}^T_{k-1}P_{k-1}\mathbf{x}_{k-1}$$

  and

  $$\begin{aligned} P_{k-1} &= Q_{k-1} + F^T_{k-1}R_{k-1}F_{k-1} + \\ &\quad \{A_{k-1} - B_{k-1}F_{k-1}\}^T P_k \{A_{k-1} - B_{k-1}F_{k-1}\} \end{aligned}$$

- Thus the same properties hold at time $k-1$ and k, and N and $N-1$ in particular, so they will always be true.

- Can summarize the above in the algorithm:

$$\begin{aligned}
(i) &\quad P_N = H \\
(ii) &\quad F_k = \left[R_k + B_k^T P_{k+1} B_k\right]^{-1} B_k^T P_{k+1} A_k \\
(iii) &\quad P_k = Q_k + F_k^T R_k F_k + \{A_k - B_k F_k\}^T P_{k+1} \{A_k - B_k F_k\}
\end{aligned}$$

  cycle through steps (ii) and (iii) from $N - 1 \rightarrow 0$.

- Notes:
  - The result is a control schedule that is time varying, even if A, B, Q, and R are constant.
  - Clear that P_k and F_k are independent of the state and can be computed ahead of time, off-line.
  - Possible to eliminate the F_k part of the cycle, and just cycle through P_k

$$P_k = Q_k + A_k^T \left\{P_{k+1} - P_{k+1} B_k \left[R_k + B_k^T P_{k+1} B_k\right]^{-1} B_k^T P_{k+1}\right\} A_k$$

Steady State

- Assume[3]
  - Time invariant problem (LTI) – i.e., A, B, Q, R are constant
  - System $[A, B]$ stabilizable – uncontrollable modes are stable.

- For any H, then as $N \rightarrow \infty$, the recursion for P tends to a constant solution with $P_{ss} \geq 0$ that is bounded and satisfies (set $P_k \equiv P_{k+1}$)

$$P_{ss} = Q + A^T \left\{ P_{ss} - P_{ss} B \left[R + B^T P_{ss} B \right]^{-1} B^T P_{ss} \right\} A \quad (3.21)$$

  - Discrete form of the famous **Algebraic Riccati Equation**
  - Typically hard to solve analytically, but easy to solve numerically.
  - Can be many PSD solutions of (3.21), recursive solution will be one.

- Let $Q = C^T C \geq 0$, which is equivalent to having cost measurements $\mathbf{z} = C\mathbf{x}$ and state penalty $\mathbf{z}^T\mathbf{z} = \mathbf{x}^T C^T C\mathbf{x} = \mathbf{x}^T Q\mathbf{x}$. If $[A, C]$ detectable, then:
  - Independent of H, recursion for P has a **unique** steady state solution $P_{ss} \geq 0$ that is the unique PSD solution of (3.21).
  - The associated steady state gain is

    $$F_{ss} = \left[R + B^T P_{ss} B\right]^{-1} B^T P_{ss} A$$

    and using F_{ss}, the closed-loop system $\mathbf{x}_{k+1} = (A - BF_{ss})\mathbf{x}_k$ is **asymptotically stable**, i.e.,

    $$|\lambda(A - BF_{ss})| < 1$$

- If, in addition, $[A, C]$ observable[4], then there is a unique $P_{ss} > 0$

DP in Continuous Time

- Have analyzed a couple of approximate solutions to the classic control problem of minimizing:

$$\min J = h(\mathbf{x}(t_f), t_f) + \int_{t_0}^{t_f} g(\mathbf{x}(t), \mathbf{u}(t), t)\, dt$$

subject to

$$\begin{aligned}
\dot{\mathbf{x}} &= \mathbf{a}(\mathbf{x}, \mathbf{u}, t) \\
\mathbf{x}(t_0) &= \text{given} \\
\mathbf{m}(\mathbf{x}(t_f), t_f) &= 0 \text{ set of terminal conditions} \\
\mathbf{u}(t) &\in \mathcal{U} \text{ set of possible constraints}
\end{aligned}$$

- Previous approaches discretized in time, state, and control actions
  - Useful for implementation on a computer, but now want to consider the exact solution in continuous time
  - Result will be a nonlinear partial differential equation called the **Hamilton-Jacobi-Bellman** equation (**HJB**) – a key result.

- First step: consider cost over the interval $[t, t_f]$, where $t \leq t_f$ of any control sequence $\mathbf{u}(\tau)$, $t \leq \tau \leq t_f$

$$J(\mathbf{x}(t), t, \mathbf{u}(\tau)) = h(\mathbf{x}(t_f), t_f) + \int_t^{t_f} g(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau)\, d\tau$$

  - Clearly the goal is to pick $\mathbf{u}(\tau)$, $t \leq \tau \leq t_f$ to minimize this cost.

$$J^\star(\mathbf{x}(t), t) = \min_{\substack{\mathbf{u}(\tau) \in \mathcal{U} \\ t \leq \tau \leq t_f}} J(\mathbf{x}(t), t, \mathbf{u}(\tau))$$

- Approach:
  - Split time interval $[t, t_f]$ into $[t, t + \Delta t]$ and $[t + \Delta t, t_f]$, and are specifically interested in the case where $\Delta t \to 0$
  - Identify the optimal cost-to-go $J^\star(\mathbf{x}(t + \Delta t), t + \Delta t)$
  - Determine the "stage cost" in time $[t, t + \Delta t]$
  - Combine above to find best strategy from time t.
  - Manipulate result into HJB equation.

- Split:

$$J^\star(\mathbf{x}(t), t) = \min_{\substack{\mathbf{u}(\tau)\in\mathcal{U} \\ t\le\tau\le t_f}} \left\{ h(\mathbf{x}(t_f), t_f) + \int_t^{t_f} g(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau))\, d\tau \right\}$$

$$= \min_{\substack{\mathbf{u}(\tau)\in\mathcal{U} \\ t\le\tau\le t_f}} \left\{ h(\mathbf{x}(t_f), t_f) + \int_t^{t+\Delta t} g(\mathbf{x}, \mathbf{u}, \tau)\, d\tau + \int_{t+\Delta t}^{t_f} g(\mathbf{x}, \mathbf{u}, \tau)\, d\tau \right\}$$

- Implicit here that at time $t+\Delta t$, the system will be at state $\mathbf{x}(t+\Delta t)$.
  - But from the **principle of optimality**, we can write that the optimal cost-to-go from this state is:

$$J^{\star}(\mathbf{x}(t+\Delta t), t+\Delta t)$$

- Thus can rewrite the cost calculation as:

$$J^{\star}(\mathbf{x}(t), t) \;=\; \min_{\substack{\mathbf{u}(\tau)\in\mathcal{U} \\ t\leq\tau\leq t+\Delta t}} \left\{ \int_{t}^{t+\Delta t} g(\mathbf{x}, \mathbf{u}, \tau)\, d\tau + J^{\star}(\mathbf{x}(t+\Delta t), t+\Delta t) \right\}$$

- Assuming $J^\star(\mathbf{x}(t+\Delta t), t+\Delta t)$ has bounded second derivatives in both arguments, can expand this cost as a Taylor series about $\mathbf{x}(t), t$

$$\begin{aligned} J^\star(\mathbf{x}(t+\Delta t), t+\Delta t) \;&\approx\; J^\star(\mathbf{x}(t),t) + \left[\frac{\partial J^\star}{\partial t}(\mathbf{x}(t),t)\right]\Delta t \\ &\quad + \left[\frac{\partial J^\star}{\partial \mathbf{x}}(\mathbf{x}(t),t)\right](\mathbf{x}(t+\Delta t) - \mathbf{x}(t)) \end{aligned}$$

  - Which for small Δt can be compactly written as:

$$\begin{aligned} J^\star(\mathbf{x}(t+\Delta t), t+\Delta t) \;&\approx\; J^\star(\mathbf{x}(t),t) + J^\star_t(\mathbf{x}(t),t)\Delta t \\ &\quad + J^\star_{\mathbf{x}}(\mathbf{x}(t),t)\mathbf{a}(\mathbf{x}(t),\mathbf{u}(t),t)\Delta t \end{aligned}$$

- Substitute this into the cost calculation with a small Δt to get

$$\begin{aligned} J^\star(\mathbf{x}(t),t) \;=\; & \min_{\mathbf{u}(t)\in\mathcal{U}} \{g(\mathbf{x}(t),\mathbf{u}(t),t)\Delta t + J^\star(\mathbf{x}(t),t) \\ & + J^\star_t(\mathbf{x}(t),t)\Delta t + J^\star_{\mathbf{x}}(\mathbf{x}(t),t)\mathbf{a}(\mathbf{x}(t),\mathbf{u}(t),t)\Delta t\} \end{aligned}$$

- Extract the terms that are independent of $\mathbf{u}(t)$ and cancel

$$0 = J^\star_t(\mathbf{x}(t),t) + \min_{\mathbf{u}(t)\in\mathcal{U}} \{g(\mathbf{x}(t),\mathbf{u}(t),t) + J^\star_{\mathbf{x}}(\mathbf{x}(t),t)\mathbf{a}(\mathbf{x}(t),\mathbf{u}(t),t)\}$$

  - This is a partial differential equation in $J^\star(\mathbf{x}(t),t)$ that is solved backwards in time with an initial condition that

$$J^\star(\mathbf{x}(t_f),t_f) = h(\mathbf{x}(t_f))$$

  for $\mathbf{x}(t_f)$ and t_f combinations that satisfy $m(\mathbf{x}(t_f),t_f) = 0$

HJB Equation

- For simplicity, define the **Hamiltonian**

$$\mathcal{H}(\mathbf{x}, \mathbf{u}, J^\star_{\mathbf{x}}, t) = g(\mathbf{x}(t), \mathbf{u}(t), t) + J^\star_{\mathbf{x}}(\mathbf{x}(t), t)\mathbf{a}(\mathbf{x}(t), \mathbf{u}(t), t)$$

then the **HJB equation** is

$$\boxed{-J^\star_t(\mathbf{x}(t), t) \;=\; \min_{\mathbf{u}(t)\in\mathcal{U}} \mathcal{H}(\mathbf{x}(t), \mathbf{u}(t), J^\star_{\mathbf{x}}(\mathbf{x}(t), t), t)}$$

  - A very powerful result, that is both a **necessary and sufficient** condition for optimality
  - But one that is hard to solve/use in general.

Continuous LQR

- Specialize to a linear system model and a quadratic cost function

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + B(t)\mathbf{u}(t)$$

$$J = \frac{1}{2}\mathbf{x}(t_f)^T H \mathbf{x}(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left\{\mathbf{x}(t)^T R_{\mathrm{xx}}(t)\mathbf{x}(t) + \mathbf{u}(t)^T R_{\mathrm{uu}}(t)\mathbf{u}(t)\right\} dt$$

  – Assume that t_f fixed and there are no bounds on $\mathbf{u}$,

  – Assume $H, R_{\mathrm{xx}}(t) \geq 0$ and $R_{\mathrm{uu}}(t) > 0$, then

$$\begin{aligned}\mathcal{H}(\mathbf{x}, \mathbf{u}, J^\star_{\mathbf{x}}, t) \;=\; & \frac{1}{2}\left[\mathbf{x}(t)^T R_{\mathrm{xx}}(t)\mathbf{x}(t) + \mathbf{u}(t)^T R_{\mathrm{uu}}(t)\mathbf{u}(t)\right] \\ & + J^\star_{\mathbf{x}}(\mathbf{x}(t), t)\left[A(t)\mathbf{x}(t) + B(t)\mathbf{u}(t)\right]\end{aligned}$$

- Now need to find the minimum of $\mathcal{H}$ with respect to $\mathbf{u}$, which will occur at a stationary point that we can find using (no constraints)

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}} = \mathbf{u}(t)^T R_{\mathrm{uu}}(t) + J^\star_{\mathbf{x}}(\mathbf{x}(t), t) B(t) = 0$$

  – Which gives the **optimal control law:**

$$\mathbf{u}^\star(t) = -R_{\mathrm{uu}}^{-1}(t) B(t)^T J^\star_{\mathbf{x}}(\mathbf{x}(t), t)^T$$

  – Since

$$\frac{\partial^2 \mathcal{H}}{\partial \mathbf{u}^2} = R_{\mathrm{uu}}(t) > 0$$

  then this defines a global minimum.

- Given this control law, can rewrite the Hamiltonian as:

$$\mathcal{H}(\mathbf{x},\mathbf{u}^\star,J^\star_{\mathbf{x}},t) \;=$$

$$\frac{1}{2}\left[\mathbf{x}(t)^T R_{\mathrm{xx}}(t)\mathbf{x}(t) + J^\star_{\mathbf{x}}(\mathbf{x}(t),t)B(t)R_{\mathrm{uu}}^{-1}(t)R_{\mathrm{uu}}(t)R_{\mathrm{uu}}^{-1}(t)B(t)^T J^\star_{\mathbf{x}}(\mathbf{x}(t),t)^T\right]$$
$$+J^\star_{\mathbf{x}}(\mathbf{x}(t),t)\left[A(t)\mathbf{x}(t) - B(t)R_{\mathrm{uu}}^{-1}(t)B(t)^T J^\star_{\mathbf{x}}(\mathbf{x}(t),t)^T\right]$$

$$= \frac{1}{2}\mathbf{x}(t)^T R_{\mathrm{xx}}(t)\mathbf{x}(t) + J^\star_{\mathbf{x}}(\mathbf{x}(t),t)A(t)\mathbf{x}(t)$$
$$-\frac{1}{2}J^\star_{\mathbf{x}}(\mathbf{x}(t),t)B(t)R_{\mathrm{uu}}^{-1}(t)B(t)^T J^\star_{\mathbf{x}}(\mathbf{x}(t),t)^T$$

- Might be difficult to see where this is heading, but note that the boundary condition for this PDE is:

$$J^{\star}(\mathbf{x}(t_f), t_f) = \frac{1}{2}\mathbf{x}^T(t_f) H \mathbf{x}(t_f)$$

  - So a candidate solution to investigate is to maintain a quadratic form for this cost for all time t. So could assume that

$$J^{\star}(\mathbf{x}(t), t) = \frac{1}{2}\mathbf{x}^T(t) P(t) \mathbf{x}(t), \qquad P(t) = P^T(t)$$

    and see what conditions we must impose on $P(t)$.

  - Note that in this case, $J^{\star}$ is a function of the variables $\mathbf{x}$ and t

$$\begin{aligned} \frac{\partial J^{\star}}{\partial \mathbf{x}} &= \mathbf{x}^T(t) P(t) \\ \frac{\partial J^{\star}}{\partial t} &= \frac{1}{2}\mathbf{x}^T(t) \dot{P}(t) \mathbf{x}(t) \end{aligned}$$

- To use HJB equation need to evaluate:

$$-J_t^\star(\mathbf{x}(t),t) = \min_{\mathbf{u}(t)\in\mathcal{U}} \mathcal{H}(\mathbf{x}(t),\mathbf{u}(t),J_{\mathbf{x}}^\star,t)$$

- Substitute candidate solution into HJB:

$$\begin{aligned}-\frac{1}{2}\mathbf{x}(t)^T\dot{P}(t)\mathbf{x}(t) &= \frac{1}{2}\mathbf{x}(t)^T R_{\mathrm{xx}}(t)\mathbf{x}(t) + \mathbf{x}^T P(t)A(t)\mathbf{x}(t)\\ &\quad -\frac{1}{2}\mathbf{x}^T(t)P(t)B(t)R_{\mathrm{uu}}^{-1}(t)B(t)^T P(t)\mathbf{x}(t)\end{aligned}$$

$$\begin{aligned}&= \frac{1}{2}\mathbf{x}(t)^T R_{\mathrm{xx}}(t)\mathbf{x}(t) + \frac{1}{2}\mathbf{x}^T(t)\{P(t)A(t) + A(t)^T P(t)\}\mathbf{x}(t)\\ &\quad -\frac{1}{2}\mathbf{x}^T(t)P(t)B(t)R_{\mathrm{uu}}^{-1}(t)B(t)^T P(t)\mathbf{x}(t)\end{aligned}$$

which must be true for all $\mathbf{x}(t)$, so we require that $P(t)$ solve

$$\boxed{\begin{aligned}-\dot{P}(t) &= P(t)A(t) + A(t)^T P(t) + R_{\mathrm{xx}}(t) - P(t)B(t)R_{\mathrm{uu}}^{-1}(t)B(t)^T P(t)\\ P(t_f) &= H\end{aligned}}$$

- If $P(t)$ solves this **Differential Riccati Equation**, then the HJB equation is satisfied by the candidate $J^\star(\mathbf{x}(t), t)$ and the resulting control is **optimal**.

- Key thing about this $J^\star$ solution is that, since $J^\star_{\mathbf{x}} = \mathbf{x}^T(t)P(t)$, then

$$\begin{aligned} \mathbf{u}^\star(t) &= -R_{\mathrm{uu}}^{-1}(t)B(t)^T J^\star_{\mathbf{x}}(\mathbf{x}(t), t)^T \\ &= -R_{\mathrm{uu}}^{-1}(t)B(t)^T P(t)\mathbf{x}(t) \end{aligned}$$

  - Thus **optimal feedback control is a linear state feedback** with gain

$$F(t) = R_{\mathrm{uu}}^{-1}(t)B(t)^T P(t) \Rightarrow \mathbf{u}(t) = -F(t)\mathbf{x}(t)$$

    ◇ Can be solved for ahead of time.

- If we assume LTI dynamics and let $t_f \rightarrow \infty$, then at any finite time t, would expect the Differential Riccati Equation to settle down to a steady state value (if it exists) which is the solution of

$$PA + A^T P + R_{\text{xx}} - PBR_{\text{uu}}^{-1} B^T P = 0$$

  - Called the **(Control) Algebraic Riccati Equation (CARE)**
  - Typically assume that $R_{\text{xx}} = C_z^T R_{\text{zz}} C_z$, $R_{\text{zz}} > 0$ associated with performance output variable $\mathbf{z}(t) = C_z \mathbf{x}(t)$

- A scalar system with dynamics $\dot{x} = ax + bu$ and with cost $(R_{\rm xx} > 0$ and $R_{\rm uu} > 0)$

$$J = \int_0^{t_f} \left(R_{\rm xx}x^2(t) + R_{\rm uu}u^2(t)\right)\, dt$$

- This simple system represents one of the few cases for which the differential Riccati equation can be solved analytically:

$$P(\tau) = \frac{(aP_{t_f} + R_{\rm xx})\sinh(\beta\tau) + \beta P_{t_f}\cosh(\beta\tau)}{(b^2P_{t_f}/R_{\rm uu} - a)\sinh(\beta\tau) + \beta\cosh(\beta\tau)}$$

where $\tau = t_f - t$, $\beta = \sqrt{a^2 + b^2(R_{\rm xx}/R_{\rm uu})}$.

HJB Equation and Eular equation

Consider the fixed endpoint problem:

$$J=\int_{t_0}^{t_f} M\left(x,\dot{x},t\right)dt$$

From the Hamilton-Jacobi-Bellman equation:

$$-\frac{\partial V\left[x,t\right]}{\partial t}=\min_{x(t)}\left\{M\left(x,\dot{x},t\right)+\frac{\partial V\left[x,t\right]}{\partial x^T}\dot{x}\right\}$$

The optimal solution should satisfy (notice that left side has no $\dot{x}$)

$$\frac{\partial}{\partial \dot{x}}\left\{M+\frac{\partial V}{\partial x^T}\dot{x}\right\}=0$$

that is $\frac{\partial M}{\partial \dot{x}}+\frac{\partial V}{\partial x}=0 \quad \rightarrow \quad \frac{d}{dt}\left\{\frac{\partial M}{\partial \dot{x}}+\frac{\partial V}{\partial x}\right\}=0 \quad \rightarrow \quad \frac{d}{dt}\frac{\partial M}{\partial \dot{x}}+\frac{\partial^2 V}{\partial x \partial t}+\frac{\partial^2 V}{\partial x^2}\dot{x}=0 \quad \cdots(1)$

From the Hamilton-Jacobi-Bellman equation, the optimal solution should also satisfy

$$M+\frac{\partial V}{\partial t}+\frac{\partial V}{\partial x^T}\dot{x}=0 \quad \rightarrow \quad \frac{\partial}{\partial x}\left\{M+\frac{\partial V}{\partial t}+\frac{\partial V}{\partial x^T}\dot{x}\right\}=0 \quad \rightarrow \quad \frac{\partial M}{\partial x}+\frac{\partial^2 V}{\partial t \partial x}+\frac{\partial^2 V}{\partial x^2}\dot{x}=0 \quad \cdots(2)$$

Let $(2)-(1)$:

$$\frac{\partial M}{\partial x}-\frac{d}{dt}\frac{\partial M}{\partial \dot{x}}=0 \quad \rightarrow \quad \text{Euler equation}$$

Dynamic programming and minimum principle

- Dynamic programming

Optimal objective function

Twice continuously differentiable

Viscous solution

- Minimum principle

$$\dot{X} = f(X, U, t) \qquad\qquad X(t_0) = X_0$$

Choose $U(t)$ **to minimize**

$$J = \phi\left[X(t_f), t_f\right] + \int_{t_0}^{t_f} F(X, U, t)dt$$

where t_0, t_f **is fixed,** $X(t_f)$ **is free,** U**may be constrained or not.**

1、 $\dot{X} = \frac{\partial H}{\partial \lambda}$ （**State equation**）

2、 $\dot{\lambda} = -\frac{\partial H}{\partial X}$ （**Costate equation**）

3、 $X(t_0) = X_0$ （**Boundary equation**）

4、 $\lambda(t_f) = \frac{\partial \phi}{\partial X(t_f)}$ （**Transversality condition**）

5、 $\min_{u \in \Omega} H(X^*, \lambda^*, U, t) = H(X^*, \lambda^*, U^*, t)$

（**Extremal condition**）

From the results of dynamic programming, co-state λ satisfies

$$\lambda = \frac{\partial J}{\partial X}$$

HJB equation indicates the extremal condition of Hamilton function is same as condition (5) in the minimal principle. Moreover,

$$H^* = -\frac{\partial J}{\partial t}$$

From dynamic programming to co-state equation

$$\dot{\lambda} = \frac{d\lambda}{dt} = \frac{d}{dt}\left[\frac{\partial J(X,t)}{\partial X}\right] = \frac{\partial}{\partial t}\left[\frac{\partial J(X,t)}{\partial X}\right] + \frac{\partial^2 J(X,t)}{\partial X^2}\frac{dX}{dt}$$

From the twice continuously differentiability, we can exchange the order of derivative，

$$
\begin{aligned}
\dot{\lambda} &= \frac{\partial}{\partial X}\left[\frac{\partial J(X,t)}{\partial t}\right] + \frac{\partial^2 J(X,t)}{\partial X^2} f(X,U,t) \\
&= -\frac{\partial}{\partial X}\left[F(X,U,t) + (\frac{\partial J}{\partial X})^T \cdot f(X,U,t)\right] + \frac{\partial^2 J(X,t)}{\partial X^2} f(X,U,t) \\
&= -\left[\frac{\partial F}{\partial X} + (\frac{\partial J}{\partial X})^T \frac{\partial f}{\partial X}\right] \\
&= -\frac{\partial H}{\partial X}
\end{aligned}
$$

since $H = F(X,U,t) + \lambda(t)^T f(X,U,t)$

$$-\frac{\partial}{\partial X}\left[F + (\frac{\partial J}{\partial X})^T f\right]$$

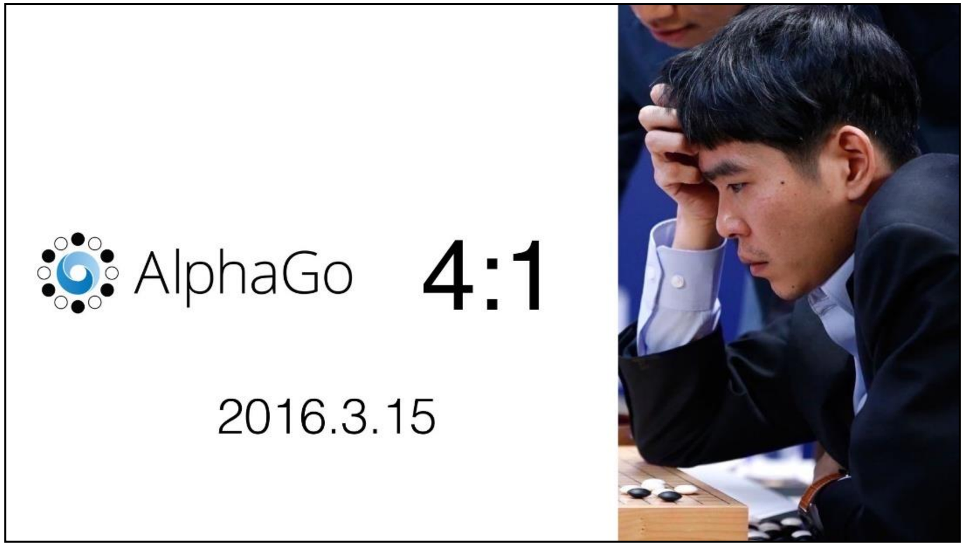
Transversality condition：

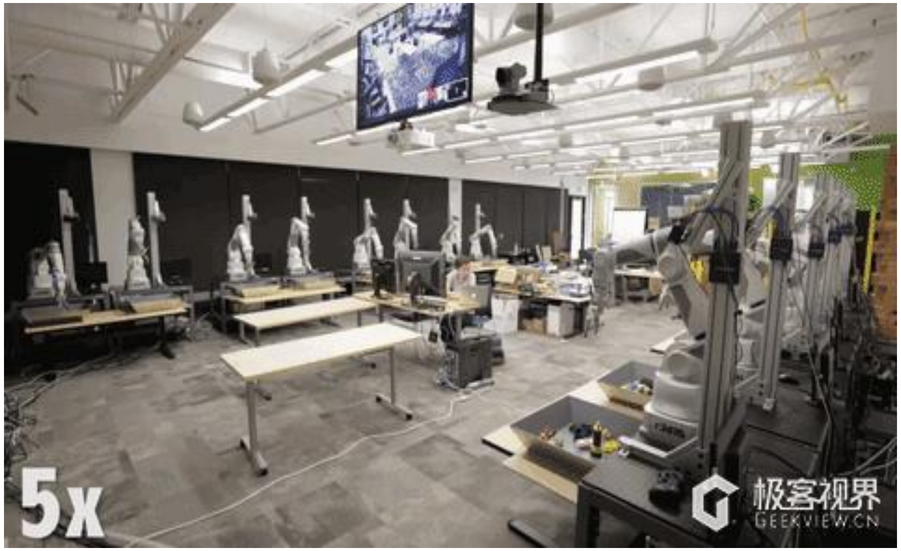
$$\lambda(t_f) = \frac{\partial J\left[X(t_f), t_f\right]}{\partial X(t_f)}$$

$$= \frac{\partial \varphi\left[X(t_f), t_f\right]}{\partial X(t_f)}$$

Other conditions, such as state equation and initial condition, are given。So, we obtain the minimal principle from HJB equation。This is not to say that minimal principle can be proved by dynamic programming. We assume that $V(X,t)$ is *twice diferentible* with respect to x and t

Deep Reinforcement Learning





Dynamic programming and Reinforcement learning

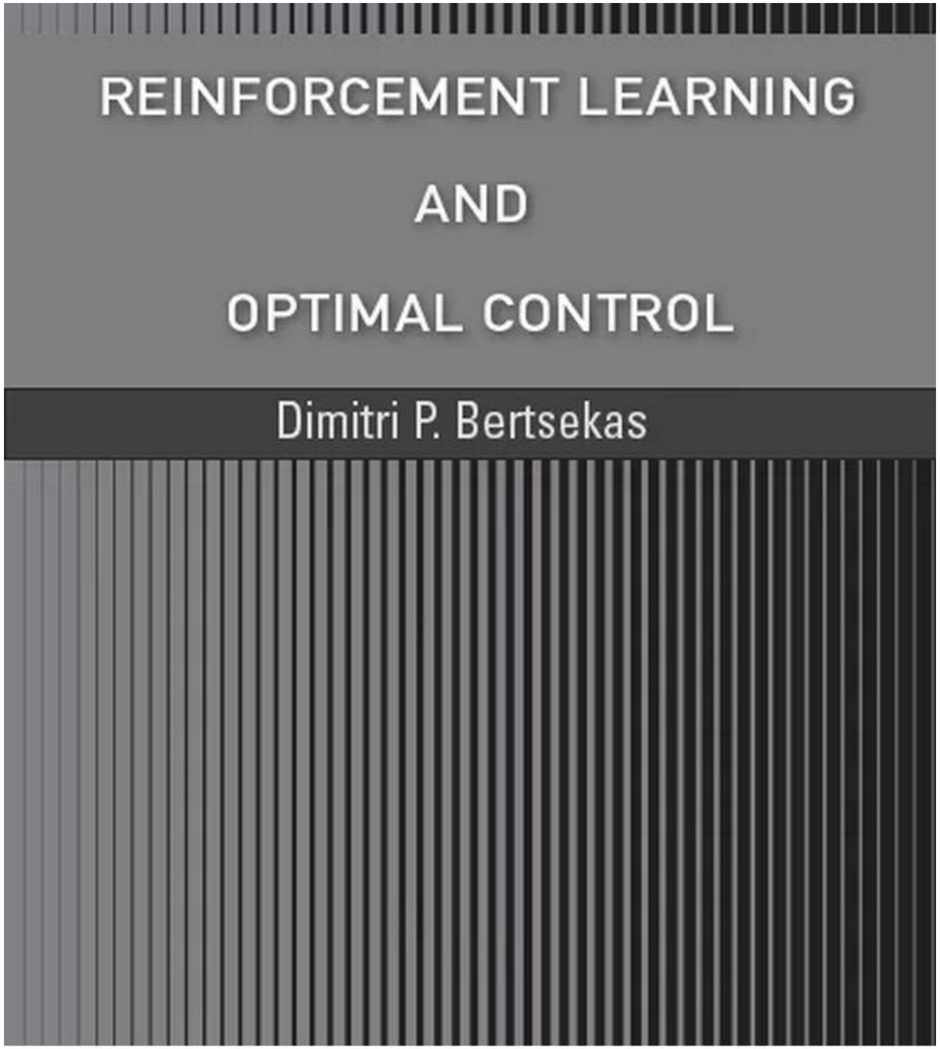
- Optimality equation

$$J^\star(x_k^i, t_k) = \min_{u_k^{ij}} \left[g(x_k^i, u_k^{ij}, t_k)\Delta t + J^\star(x_{k+1}^j, t_{k+1}) \right]$$

$$J^\star(\mathbf{x}(t), t) \;=\; \min_{\substack{\mathbf{u}(\tau)\in\mathcal{U}\\ t\le\tau\le t+\Delta t}} \left\{ \int_t^{t+\Delta t} g(\mathbf{x}, \mathbf{u}, \tau)\, d\tau + J^\star(\mathbf{x}(t+\Delta t), t+\Delta t) \right\}$$

Value function

Q-function



REINFORCEMENT LEARNING

AND

OPTIMAL CONTROL

Dimitri P. Bertsekas

REINFORCEMENT LEARNING AND OPTIMAL CONTROL
by Dimitri P. Bertsekas
Athena Scientific, 2019

https://mp.weixin.qq.com/s/Y9DfxQJ-w23QXxKV0z26ag

Thank You