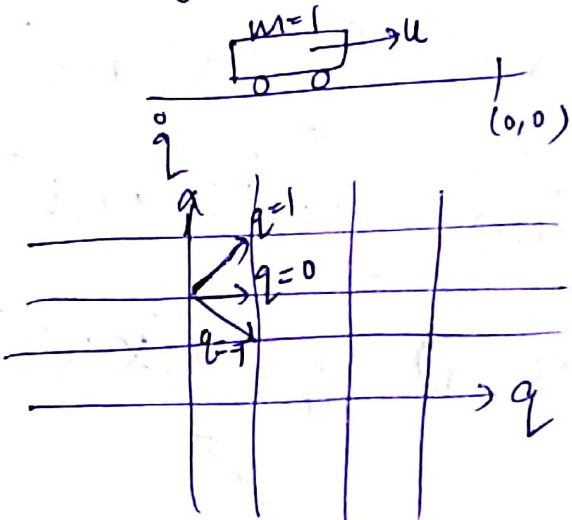
Lecture - 4

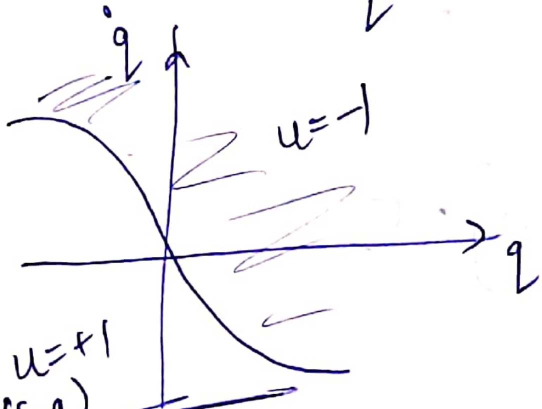
Continuous ~~[illegible]~~ Dynamic Programming

Previously (discrete case) — Minimum-time Double Integrator

$\ddot{q} = u \quad |u| \leq 1$



(finite set of states)



$l(s,a)$

min time

$$l(s) = \begin{cases} 1 & s \neq s_{goal} \\ 0 & s = s_{goal} \end{cases}$$

quadratic cost

$$\boxed{l(x,u) = x^T x + u^T u}$$

Value Iteration Algorithm (discrete setting)

one state update

$$\forall s_i \quad J^*(s_i) = \min_a \left[l(s,a) + J^*(f(s,a)) \right]$$

min over action

J^* — optimal cost-to-go (aka value function)

$$\sum_{n=0}^{N=\infty} l(s[n], a[n])$$

estimate notation $\hat{J}^*$ — vector with one element for each s_i (state)

$$\pi^*(s) = \underbrace{\operatorname{argmin}_a}_{} \left[l(s,a) + J^*(f(s,a)) \right]$$

π^* — optimal policy

argmin — outputs 'a' which obtains the minimum value

~~[illegible]~~

Controllers

$u = \pi(x)$

$a = \pi(s)$

Discrete

$$s[n+1] = f_d(s[n], a[n])$$

$$\min_{a[\cdot]} \sum_{n=0}^{\infty} l_d(s[n], a[n])$$

$$\forall s_i \quad J^*(s_i) = \min_a \left[l_d(s,a) + J^*(f_d(s,a)) \right]$$

Continuous

$$\dot{x}(t) = f_c(x(t), u(t))$$

$$\min_{u(\cdot)} \int_0^{\infty} l_c(x(t), u(t))\, dt$$

partial diff eqn

$$\forall x, \quad \min_u \left[l(x,u) + \frac{\partial J}{\partial x} f(x,u) \right] = 0$$

Hamilton-Jacobi-Bellman Equation (HJB)

$$x[n+1] \approx x[n] + h f_c(x[n], u[n])$$

(approximation) time-step

Euler integration

$$L_d(x,u) \approx h\, l_c(x,u)$$

$$J^*(x) = \min_u \left[h\, l_c(x,u) + J^*(x + h f_c(x,u)) \right]$$

(Euler approx.) $J^*(x) + h \frac{\partial J}{\partial x} f_c(x,u)$

doesn't depend on u

$$J^*(x) = \min_u \left[J^*(x) + h\, l_c(x,u) + h \frac{\partial J}{\partial x} f_c(x,u) \right]$$

$$\Rightarrow \cancel{J^*(x)} = \cancel{J^*(x)} + \min_u \left[h\, l_c(x,u) + h \frac{\partial J}{\partial x} f_c(x,u) \right]$$

$$\Rightarrow \min_u \left[\cancel{h}\, l_c(x,u) + \cancel{h} \frac{\partial J}{\partial x} f_c(x,u) \right] = 0$$

$u^* = \pi^*(x)$

optimal action for given x

negative of the current cost.

$$\underbrace{\frac{\partial J}{\partial x} f_c(x, u^*)} = -l_c(x, u^*)$$

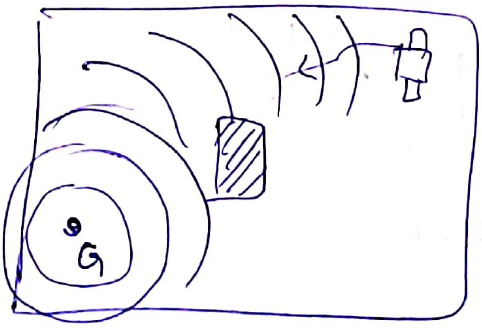
$$\frac{dJ}{dt} = \frac{\partial J}{\partial x}\frac{\partial x}{\partial t}$$

rate of change of the cost to go

The way the cost to go changes as we are moving is at the rate we are incurring cost.

So, if we are to follow a trajectory along the system, our cost to go is going downhill at the rate we would be incurring cost.

Continuous version of grid world



Double integrator (w/) with quadratic cost

$\ddot{q} = u \qquad l(q, \dot{q}, u) = q^2 + \dot{q}^2 + u^2$

Linear quadratic regulator

claim: $u^* = \pi^*(x) = -q - \sqrt{3}\dot{q}$

optimal policy

proof: $J^*(x) = \sqrt{3} q^2 + 2q\dot{q} + \sqrt{3}\dot{q}^2$

$x = \begin{bmatrix} q \\ \dot{q} \end{bmatrix}$

$\frac{\partial J}{\partial q} = 2\sqrt{3} q + 2\dot{q}$

$\frac{\partial J}{\partial \dot{q}} = 2\sqrt{3}\dot{q} + 2q$

$$0 = \min_u \Big[q^2 + \dot{q}^2 + u^2 + \underbrace{\frac{\partial J}{\partial q}\dot{q} + \frac{\partial J}{\partial \dot{q}} u} \Big]$$

$\frac{dJ}{dt} = \frac{\partial J}{\partial x} f(x, u)$

$= \left[\frac{\partial J}{\partial q}, \frac{\partial J}{\partial \dot{q}}\right]\begin{bmatrix} \dot{q} \\ u \end{bmatrix}$

$$0 = \min_u \left[q^2 + \dot{q}^2 + u^2 + (2\sqrt{3}q + 2\dot{q})\dot{q} + (2\sqrt{3}\dot{q} + 2q)u \right]$$

$$\Rightarrow \frac{\partial}{\partial u} = 0$$

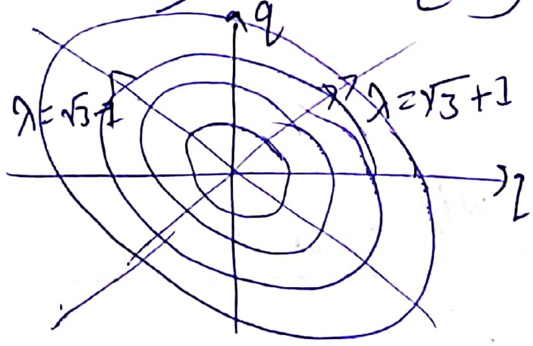
$$\frac{\partial}{\partial u}\left[\quad \right] = 2u + 2\sqrt{3}\dot{q} + 2q = 0$$

$$u = -q - \sqrt{3}\dot{q}$$

Can we draw $J^*(x)$? $\Rightarrow$ YES

$$J(x) = x^T \begin{bmatrix} \sqrt{3} & 1 \\ 1 & \sqrt{3} \end{bmatrix} x = x^T S x$$

Eigen vector $V = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}$



General Case (Linear Quadratic Regulator)

$\dot{x} = Ax + Bu$
(Linear dynamical system)

Cost function
$\ell(x,u) = x^T Q x + u^T R u$ (general quadratic form)

Given:
$J^*(x) = x^T S x \ , S > 0$

$Q \geq 0$ (positive semi-definite matrix)

only
$R > 0$
positive definite matrix
$\Downarrow$
$R = R^T$ (symmetric)

$$\min_u \left[x^TQx + u^TRu + 2x^TS(Ax+Bu) \right]$$

gradient of cost function $\frac{\partial J}{\partial x} = \cancel{2}x^TS + x^TS^T = 2x^TS$ ($\because S$ is symmetric)

$$\frac{\partial}{\partial u}[\dots] = \cancel{2}u^TR + \cancel{2}x^TSB = 0$$

go downhill but project me in the actuator space

$$\boxed{u^* = \underbrace{R^{-1}B^TS}_{K}x = -Kx}$$

$$\forall x \quad x^T[Q - SBR^{-1}B^TS + SA + A^TS]x = 0$$

if this is going to be true for all x, then it's sufficient for the ~~quantity~~ expression in the middle to be 0.

$$0 = Q - SBR^{-1}B^TS + SA + A^TS$$

(Differential Riccati eqn)

CARE — Continuous Algebraic Riccati Equation

Matlab

Solver

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

$\boxed{-Sx \text{ go down the gradient of } S}$

$-B^TSx$

(B maps from state space to actuator space from $u \to \dot{x}$)

Note — If we have cost-to-go function, then that captures everything we need to know about the long-term behavior.

→ Being greedy w.r.t. the cost-to-go function is equivalent to taking the long-term decisions on the full dynamics.

→ Linear systems tend to live well on quadratic form.