dynamic programming

Lewis et al 6

Stengel 3.4

Bertsekas 95 Ch 1

goal: solve multi-stage optimization problems using Bellman's principle of optimality

- suppose now we are given a DT DE

$$x^+ = F(x,u),\ x \in \mathbb{R}^n,\ u \in \mathbb{R}^m$$

and we wish to choose inputs over time $u : [0,t] \to \mathbb{R}^m$ to minimize cost

$$c(x,u) = \underbrace{\ell(t, x(t))}_{\text{"final" cost}} + \sum_{s=0}^{t-1} \underbrace{\mathcal{L}(s, x(s), u(s))}_{\text{"running" cost}}$$

- Richard Bellman published key insight in 1957:

* the optimal control at time s depends only on $x(s)$ (i.e. not on previous states/inputs)

– naturally leads to working backward from final time to determine optimal policy

- if we let $v_s^*(x(s))$ denote lowest (i.e. optimal) cost achievable from state $x(s)$ at time s,

i.e. $v_s^* : \mathbb{R}^n \to \mathbb{R}$ provides the value of each state at time s

$$v_s^*(x(s)) = \min_{u(s) \in \mathbb{R}^m} \left[\mathcal{L}(s, x(s), u(s)) + v_{s+1}^*(x(s+1)) \right]$$

*note: $x(s+1) = F(x(s), u(s))$ depends on $u(s)$

- this is referred to as the Bellman equation,

* Bellman equation enables us (in principle) to determine optimal control inputs by solving a sequence of NLP in backward time!

ex: $x^+ = ax + bu, \quad x, u \in \mathbb{R}$

$$c_\tau(x, u) = \sum_{s=\tau}^{t} q(s)\, x(s)^2 + r(s)\, u(s)^2$$

→ determine optimal input & value at $\tau = t$

- since $c_t(x, u) = q(t)\, x(t)^2 + r(t)\, u(t)^2$

$u^*(t) = 0$ is optimal input,

$V_t^*(x(t)) = q(t)\, x^*(t)^2$ is optimal value

→ determine optimal input & value at $\tau = t-1$ using Bellman's principle of optimality

- $$V_{t-1}^*(x(t-1)) = \min_{u(t-1)\in\mathbb{R}} \; q(t-1)\,x(t-1)^2 + r(t-1)\,u(t-1)^2 + q(t)\,x(t)^2$$

- $x(t) = a\,x(t-1) + b\,u(t-1)$, so cost is

$$q(t-1)\,x(t-1)^2 + r(t-1)\,u(t-1)^2 + q(t)\left[a\,x(t-1) + b\,u(t-1)\right]^2$$

- differentiating w.r.t. u_{t-1} to find stationary pt:

$$2 r_{t-1} u_{t-1} + 2 b\, q_t (a\, x_{t-1} + b\, u_{t-1}) = 0$$

$$\Longleftrightarrow \quad u_{t-1} = \frac{-ab\,q_t}{b+2r_t}\, x_{t-1} \quad (\text{\underline{linear} in } x\,!)$$

- in principle, this process could be repeated to determine optimal policy for all time

* we'll use Bellman's principle as both an analytical tool (to verify optimality of policies) & computational tool (to synthesize optimal policies)