OSC I

Eserciziario

LQ-Discreto

Problema. $$\min_u J(u) = \left\{ \sum_{i=0}^{N-1} \left(x_k^T Q\, x_k + u_k^T R\, u_k \right) + x_N^T S\, x_N \right\}, Q \succcurlyeq 0, S \succcurlyeq 0, R \succ 0$$

$$x_{k+1} = A\, x_k + B\, u_k$$

$$x_k(0) = x_0$$

Il costo ottimo è definito come:

$$V_k(x) = J(u)^* = {x_0}^T P_0 x_0$$

N.B:

Nel caso in cui si richiede il costo di una legge di controllo generica è data dal costo del testo del problema.

Si risolve effettuando iterazioni all'indietro per calcolare le P_k ed iterazioni in avanti per calcolare u_k e x_k.:

- Iterazioni all'indietro:

$$P_N = S$$

$$P_k = Q + A^T P_{k+1} A - A^T P_{k+1} B (R + B^T P_{k+1} B)^{-1} B^T P_{k+1} A$$

- Iterazioni in avanti:

$$u_k = -(R + B^T P_{k+1} B)^{-1} B^T P_{k+1} A\, x_k$$

Reinforcement Learning

Definizione 1. *Value Iteration*

Il value iteration è un algoritmo iterativo per determinare la funzione valore ottima, per ciascun valore dello stato, in un problema di Reinforcement Learning. Si ottiene considerando un algoritmo di Policy Improvement ed eseguendo la fase di policy evaluation troncata ad un singolo passo.

Il risultato complessivo di questa strategia è quella di trasformare la legge di Bellman in una legge di aggiornamento:

$$v_{i+1}(x) = \max_k \{ R_{k+1} + \gamma\, v_i(x') \}$$

Definizione 2. *Policy Improvement*

Il policy improvement è un algoritmo iterativo che, per un certo istante di tempo e per un certo stato,c'è un'azione tra quelle disponibili che fornisce una ricompensa maggiore per poi seguire la policy di partenza. Ripetendo questa operazione per ogni stato si ridefinisce l'intera policy per ottenerne una migliore aggiornando la policy π in maniera greedy rispetto alla funzione qualità $q_\pi(x, u)$. In particolare la funzione valore delle policy risultante:

$$v_{\pi'}(x) \geqslant v_\pi(x)$$

Definizione 3. *Policy Evaluation*

L'algoritmo di Policy Evaluation è un algoritmo iterativo che permette di determinare il valore $v_\pi(x)$ della policy π. In particolare, ogni passo di iterazione poniamo:

$$v_{i+1}(x) = R_{k+1} + \gamma\, v_i(x')$$

e all'infinito, dopo aver aggiornato tutti gli stati senza cambiarne il valore, si converge ad un punto fisso definito come:

$$v_\pi(x) = R_{k+1} + \gamma\, v_\pi(x')$$

Definizione 4. *Funzione Valore*

La funzione valore è una funzione che fornisce il valore di uno stato, ovvero quanto ricavo si potrebbe ottenere in futuro per il fatto di trovarsi in quel particolare stato in funzione delle azioni eseguite:

$$v_\pi(x) = G_k(x) = \sum_{i=0}^{\infty} \gamma^i\, R_{k+i+1} = R_{k+1} + \gamma\, v_\pi(x_{k+1})$$

In cui i R_{k+i+1} sono ottenuti ad ogni passo seguendo la policy π. Nel caso stocastico equivale a:

$$v_\pi(x) = \mathbb{E}[G_t|S_t] = \sum_u \pi(u|x) \sum_{x'} \sum_r p(x', r|x, u)\,(r + \gamma\, v_\pi(x')) = \sum_u \pi(u|x) q_\pi(x, u)$$

In particolare:

- $\pi(u|x) :=$ *prob. di prendere l'azione u dallo stato x;*
- $p(x', r|x, u) :=$ *prob. di finire in x' ottenendo il reward r partendo dallo stato x e scegliendo l'azione u;*
- $(r + \gamma\, v_\pi(x')) :=$ *funzione valore dello stato x;*
- $q_\pi(x, u) :=$ *funzione qualità.*

Definizione 5. *Funzione Qualità*

La funzione qualità è una funzione valore stato-azione in cui all'istante k si esegue l'azione u e dall'istante $k+1$ si esegue la policy π. Nel caso deterministico è definita come:

$$q_\pi(x, u) = R_{k+1} + \sum_{i=1}^{\infty} \gamma^i\, R_{k+i+1}$$

Invece, nel caso stocastico:

$$q_\pi(x, u) = \sum_{x'} \sum_r p(x', r|x, u)\,(r + \gamma\, v_\pi(x'))$$

Viene utilizzata nel policy Improvement per migliorare la policy π.

Problema. Dare la definizione di *Value Iteration.*Eseguire un passo di value iteration con γ e v_0 dati.

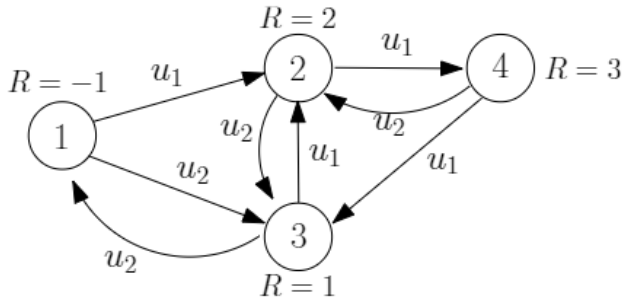

Figura 1.

$$v_{i+1}(x) = \max_u \{\rho_i(x,u) + \gamma\, v_i(x')\} \longrightarrow v_i = v_{i+1}$$

Problema. Problema di RL stocastico con funzioni valore date e probabilità di scelta delle azioni date. Calcolare $v_\pi(x)$

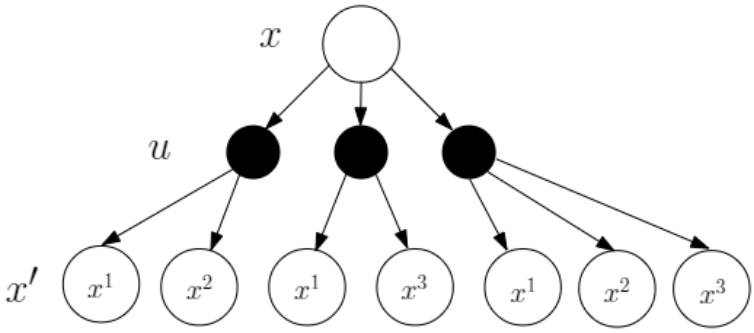

Figure 1: figura Domanda 3.

Funzione di massa di probabilità

$$(x',r) = \begin{cases} \text{Dati } x \text{ e } u_1 \\ (x^1, 0.5) \quad \text{prob} = 0.4 \\ (x^2, 0.5) \quad \text{prob} = 0.6 \end{cases}$$

$$(x',r) = \begin{cases} \text{Dati } x \text{ e } u_2 \\ (x^1, 1) \quad \text{prob} = 0.8 \\ (x^3, 4) \quad \text{prob} = 0.2 \end{cases}$$

$$(x',r) = \begin{cases} \text{Dati } x \text{ e } u_2 \\ (x^1, 0.5) \quad \text{prob} = 0.6 \\ (x^2, 4) \quad \text{prob} = 0.1 \\ (x^3, 1) \quad \text{prob} = 0.3 \end{cases}$$

Figura 2.

Applicare le formule per il caso stocastico:

- $q_\pi(x,u) = \sum_{x'} \sum_r p(x', r\,|x,u)\,(r + \gamma\, v_\pi(x'))$

- $v_\pi(x) = \sum_u \pi(u|x) q_\pi(x,u)$

Problema.

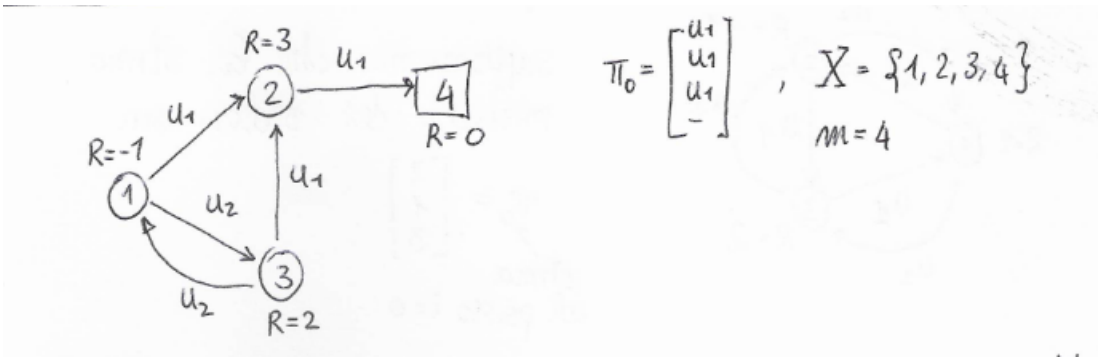

Figura 3.

Scrivere il sitema per le funzioni valore, trovare la funzione qualità e scegliere la policy π' in maniera greedy rispetto ad essa.

N.B.: Bisogna valutare tutte le u possibili in ogni stato x.

LQR Tempo Finito-Hamilton

Il problema è nella forma:

$$\min_u J(u) = \left\{ \frac{1}{2} \int_{t_0}^{T} (x^T Q\, x + u^T R\, u) \mathrm{d}t + x^T M\, x \right\}$$

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

$$J(u^*) = \frac{1}{2} x_0^T P x_0$$

Controllo ottimo:u*=$-R^{-1}\, B^T P(t)\, x$

a) Determinare la soluzione ottima del problema$\longrightarrow$**Sistema Hamiltoniano**

b) Determinare il costo ottimo dalla condizione iniziale per un controllo dato/già calcolato. Calcolo $P(t)$ a seconda dei dati dati.

c) Determinare T.

Sistema Hamiltoniano

Identificare matrice Q, R, A, B, M

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

Trovare autovalori $\lambda_s = \{\ldots\}, \lambda_u = \{\ldots\}$e verificare $m_a = m_g$ (m_g=$n - \mathrm{rg}(A - \lambda I)$):

- $m_a = m_g$

Calcolo autovettori $(H-\lambda_i I)v=0$. Le relativi matrici:

$$\Lambda_s=\begin{pmatrix}\lambda_s & 0 & \dots & 0\\ 0 & \lambda_s & \dots & 0\\ \vdots & \cdots & \lambda_s & \vdots\\ 0 & \dots & 0 & \lambda_s\end{pmatrix};\Lambda_u=\begin{pmatrix}\lambda_u & 0 & \dots & 0\\ 0 & \lambda_u & \dots & 0\\ \vdots & \cdots & \lambda_u & \vdots\\ 0 & \dots & 0 & \lambda_u\end{pmatrix}$$

Le relative matrici esponenziali:

$$e^{\Lambda_s(T-t_0)}=\begin{pmatrix}e^{\lambda_s(T-t)}\dots & \dots\\ \dots & \end{pmatrix}$$

Stessa cosa per u

- $m_a \neq m_g$

  N.B.: $m_a \rightarrow$ dimensione del blocco di Jordan in Λ_s

  1. Calcolo autovettore per trovare la catena degli autovettori generalizzati imponendo:

     $$(H-\lambda_i I)\,v_s=0$$

  2. Ora calcolo autovettore generalizzato tante volte quante m_g-m_a:

     $$(H-\lambda_i I)v=v_s$$

  3. Le relative forme di Jordan sono:

     $$\Lambda_s=\begin{pmatrix}\lambda_s & 1 & 0\\ 0 & \lambda_s & 1\\ 0 & 0 & \lambda_s\end{pmatrix}$$

  4. Calcolo matrice esponenziale:

     $$e^{\Lambda_s(T-t)}=\begin{pmatrix}e^{\lambda_s(T-t)} & (T-t)\,e^{\lambda_s(T-t)} & \frac{(T-t)^2}{2!}e^{\lambda_s(T-t)}\\ 0 & e^{\lambda_s(T-t)} & (T-t)\,e^{\lambda_s(T-t)}\\ 0 & 0 & e^{\lambda_s(T-t)}\end{pmatrix}$$

Calcolo:

$$U=\begin{pmatrix}\vdots & \vdots\\ v_s & v_u\\ \vdots & \vdots\end{pmatrix}$$

Calcolore:

$$G=(U_{22}-MU_{12})^{-1}\,(U_{21}-MU_{11})$$

Calcolare:

$$\Lambda(t)=e^{-\Lambda_u(T-t)}\,G\,e^{\Lambda_s(T-t)}$$

Calcolare:

$$P(t)=(U_{21}+U_{22}\,\Lambda(t))(U_{11}+U_{12}\,\Lambda(t))^{-1}$$

LQR Infinito - HJB/Kleinman

$$\min_u J(u) = \left\{ \frac{1}{2} \int_{t_0}^{\infty} (x^T Q\, x + u^T R\, u) \mathrm{dt} \right\}$$

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

$$J(u^*) = \frac{1}{2}\, x_0^T P x_0$$

a) Verificare se la legge di controllo in retroazione u e la matrice P_0 è valida per Kleinman;

Kleinman

b) Iterare l'algoritmo fino a quando la legge di controllo $u_i = K_i\, x$ fornisce un costo $\gtrless =$ di un certo valore con condizione iniziale;

Kleinman

c) Verificare se la legge di controllo sia ottima;

HJB

d) Trovare costo significato se u non ottima

HJB $\longrightarrow l(x,u) + c(x_1, x_2) \geqslant 0$

e) Trovare funzione Γ (presente nei vincoli dell'es.) tale che la legge di controllo sia ottima

HJB

f) Trovare funzioni valore;

a occhio vedo quella che è definita positiva. (Se metto dentro HJB DEVE tornare tutto).

g) Determinare controllo affinché $J(u)<$costante

HJB/Kleinman

h) Confrontare costi;

HJB/Kleinman

i) Determinare condizioni iniziali per avere un costo fissato con controllo dato.

Kleinman

HJB

Equazione di HJB generica

$$-\frac{\partial V^*}{\partial t} = \min_u \left\{ l(x,u,t) + \frac{\partial V^*}{\partial x}(x,t) f(x,u,t) \right\}$$

$$V^*(x,T) = m(x)$$

$$u^* = \arg\min_u \left\{ l(x,u,t) + \frac{\partial V^*}{\partial x}(x,t) f(x,u,t) \right\}$$

a tempo infinito: diventa stazionario (sparisce dipendenza da t), inoltre $-\frac{\partial V^*}{\partial t} = 0$, $V^*(0) = 0$
A tempo infinito vale ARE:

$$0 = A^T \overline{P} + \overline{P} A + Q - \overline{P} B R^{-1} B^T \overline{P}$$

$$u = -R^{-1} B^T \overline{P} x$$

SOLUZIONE PER ORIZZONTE INFINITO

1. impongo condizioni SUFFICIENTI (potrebbero essere sufficienti) di controllo ottimo

$$0 = \min_u \left\{ l(x,u) + \frac{\partial V^*}{\partial x}(x) f(x,u) \right\} = \min_u \{ g(x,u) \}$$

2. Impongo

$$\frac{\partial g}{\partial u} = 0 \longrightarrow \text{mi fornisce espressione per trovare } \frac{\partial V^*}{\partial x}(x)$$

3. Per testare se una soluzione è ottima

   a. sostituisco $u(t)$ nell'espressione precedente e posso determinare $V(x) = f(x) + h(x)$, dove $f(x)$ è nota e $h(x)$ costante di integrazione nella variable su cui NON ho integrato.

   b. sostituisco dentro $g(x,u)$ sia u data che la $V(x)$ trovata. Ottengo espressione in funzione di $h(x)$.

   c. Cerco $h(x)$ per azzerare la g. Generalmente metto $\frac{\partial h}{\partial x} = \alpha x$ in quanto termini omogenei.

   d. Se trovo h che soddisfa per qualche α, allora u ottima.

   NB h deve essere in funzione della SOLA variabile in cui non ho integrato.

4. h trovato non soddisfa, aggiungo costo significativo.

   a. aggiungo $c(x)$ all'equazione nel passo 3.c. Trovo $c(x)$ in funzione di α e x.

   b. Provo a scrivere $l(x,u) = \tilde{l}(x) + u^2$

   c. verifico $\tilde{l} + c \geqslant 0$ cercando se la rispettiva forma quadratica è semidefinita positiva e trovo α che soddisfa la condizione.

   NB potrebbe non esistere costo.

5. Se voglio calcolare il costo data x_0 estraggo $V(x)$ e sostituisco condizione iniziale.

Algoritmo di Kleinman-SOLO SE vincoli lineari

La u trovata è ottima solo all'infinito.
INIT:

1. $S_0 = A + B K_0 \longrightarrow \sigma(S_0) \subset \mathbb{C}^-$
2. trovare $P_0 \succeq 0$ tale che: $S_i \overline{P_i} + \overline{P_i} S_i = -Q - K_i^T R K_i$
3. $K_1 = -R^{-1} B^T P_0 \longrightarrow u_1 = K_1$

LOOP

1. Ripeti 1 (dovrebbe essere assicurato), 2,3 finché non soddisfi condizioni richieste. ricordando che $S_{i+1} = A + B K_i$

Se voglio determinare il costo calcolo

$$V(x,u) = J(u) = \frac{1}{2} x_0^T P_{0,T} x_0$$