Indice

Descrivere brevemente l'equazione di Hamilton-Jacaobi Bellman. Dimostrare che HJB fornisce condizioni necessarie di ottimalità.

Answer. L'equazione di Hamilton Jacobi Bellman è:

$$\begin{cases} -\frac{\partial V^*}{\partial t}(x(t),t) = \min_{u(t)} \left\{ l(x(t),u(t),t) + \frac{\partial V^*}{\partial x}(x(t),t)\, f(x(t),u(t),t) \right\} \\ V^*(x,T) = m(x) \end{cases} \quad \forall x \in \mathbb{R}^n, t \in [t_0, T]$$

Permette di rappresentare la funzione valore ottima di un problema di controllo ottimo di Bolza:

$$\begin{cases} \min_u J(u) = \min_u \left\{ \int_{t_0}^{T} l(x(\tau), u(\tau), \tau) d\tau + m(x(T)) \right\} \\ \dot{x} = f(x, u, t) \\ x(t_0) = x_0 \end{cases}$$

Generalmente questa equazione non è lineare e la sua soluzione è proprio la funzione valore ottima che rispetta l'equazioene di Hellman del problema dato e, quindi, rappresentante le decisioni ottime da effettuare partendo da un determinato stato e tempo iniziale. In particolare, l'equazione HJB fornisce condizioni necessarie di ottimalità. A tale scopo definiamo la funzione valore per il problema di Bolza:

$$V^*(x(t), t) = \min_{u(\tau), \tau \in [\tau . T]} \left\{ \int_t^T l(x(s), u(s), s) ds + m(x(T)) \right\}$$

$$= \min_{u(\tau), \tau \in [t, T]} \left\{ \int_t^{t+\Delta t} l(x(s), u(s), s) ds + \int_{t+\Delta t}^{T} l(x(s), u(s), s) ds + m(x(T)) \right\}$$

$$= \min_{u(\tau), \tau \in [t, T]} \left\{ \int_t^{t+\Delta t} l(x(s), u(s), s) ds + V^*(x(t+\Delta t), t+\Delta t) \right\}$$

Da questa espressione ricaviamo la sua corrispondente espressione differenziale sviluppando al primo ordine la funzione valore:

$$\begin{cases} 0 = \min_{u(t)} \left\{ l(x(t), u(t), t) + -\frac{\partial V^*}{\partial t}(x(t), t) + \frac{\partial V^*}{\partial x}(x(t), t)\, f(x(t), u(t), t) \right\} \\ V^*(x, T) = m(x) \end{cases} \quad \forall x \in \mathbb{R}^n, t \in [t_0, T]$$

Dato che dobbiamo determinare la soluzione per ogni istante di tempo e per ogni stato, si ottiene l'equazione HJB:

$$\begin{cases} -\frac{\partial V^*}{\partial t}(x(t), t) = \min_{u(t)} \left\{ l(x(t), u(t), t) + \frac{\partial V^*}{\partial x}(x(t), t)\, f(x(t), u(t), t) \right\} \\ V^*(x, T) = m(x) \end{cases} \quad \forall x \in \mathbb{R}^n, t \in [t_0, T]$$

Discutere l'equazione differenziale di Riccati e dimostrare che la soluzione $P(t)$ esiste per ogni $t \in [0, T]$

Answer. Considerato un problerma di controllo ottimo su orizzone finito descritto da un sistema lineare ed indice di costo quadratico:

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

$$\dot{x}(t) = A\, x(t) + B\, u(t) \quad x(0) = x_0$$

Il relativo costo di questo problema è il seguente:

$$-\frac{1}{2} x^T \dot{P} x = \left(\frac{1}{2} x(t)^T Q\, x(t) + x(t)^T A\, x - \frac{1}{2}\, x^T P(t)\, B\, R^{-1}\, B^T P(t)\, x \right)$$

Dato che questa equazione deve valere per ongi sitante di tempo e eper ogni stato, la matrice $P(t)$ deve soddisfare l'equazione differenziale di Riccati:

$$-\dot{P}(t) = P(t)\,A + A^T P(t) - P(t)\,B\,R^{-1}\,B^T P(t) + Q \quad P(T) = M$$

Per dimostrare che la soluzione della DRE esiste in tutto l'intervallo occorre dimostrare l'esistenza locale e globale della soluzione. In particolare, la soluzione locale è garantita integrando all'indietro; inoltre per l'esisteza della soluzione globale supponiamo per assurdo che esista un istante $\hat{t} < T$ è tale che $P(t)$ esiste sull'intervallo aperto $(\hat{t}, T]$, ma un suo elemento $p_{i,j}(t)$ diventa illimitato per t che converge a $\hat{t}$. Si possono distinguere due casi:

1. L'elemento di $p_{i,j}(t)$ si trova fuori diagonale,quindi $i \neq j$. Si considera il minore di ordine 2 ottenuto proprio dalle righe e colonne i, j. Ci calcoliamo il determinante:

   $$\det\begin{pmatrix} p_{i,i}(t) & p_{i,j}(t) \\ p_{i,j}(t) & p_{j,j}(t) \end{pmatrix} = p_{i,i}(t)p_{j,j}(t) - p_{i,j}(t)^2$$

   Il che è un assurdo poiché il minore diventa negativo per $t \to \hat{t}^+$ poiché $p_{i,j}(t)$ è fuori diagonale.

2. L'alemnto $p_{i,j}(t)$ appartiene alla diagonale $i = j$. Quindi si considerimaoa la base canonica come condizione:

   $$\eta_i = [0, \ldots, 1, 0, \ldots, 0]$$

   Che mi seleziona l'elemento $p_{i,i}(t)$ all'interno della diagonale. Quindi scrivendo l'indice di costo:

   $$V^*(\eta_i, t) = \frac{1}{2}\eta_i^T P(t)\,\eta_i = \frac{1}{2}P_{i,i}(t) \longrightarrow \infty \text{ per } t \longrightarrow \hat{t}^+$$

   A questo punto occorre considerare il costo ottenuto dal controllo ottimo che non deve essere maggiore rispetto al costo ottenuto da qualsiasi altro controllo, per esempio valutiamo il costo con $u = 0$. Allora:

   $$u = 0 \longrightarrow x = A\,x \longrightarrow \dot{x(\tau)} = e^{A(t-\tau)}\eta_i$$

   Quindi il costo diventa:

   $$J(0) = \frac{1}{2}\int_t^T \eta_i^T e^{A^T(t-\tau)} \mathrm{Q} \mathrm{e}^{A(t-\tau)}\eta_i d\tau + \frac{1}{2}\eta_i^T e^{A^T(T-t)} M\, e^{A^T(T-t)}\eta_i < \infty$$

   per ogni t, visto che si tratta di un integrale di una funzione continua su un intervallo limitato e quidi abbiamo un costo limitato minore del costo ottimo e quindi abbiamo un assurdo.

Enunciare il teorema di essitenza della soluzione del LQR ad orizzonte infinito. Dimostrare inoltre che la sequenza di soluzioni $P_T(t)$ del problema ad orizzone finito ha un limite per $T \longrightarrow \infty$.

Answer. Consideriamo il problema LQR su orizzonte finito con $M = 0, R > 0, Q = D^T D > 0$. Supponiamo che la coppia (A, D) è osservabile e (A, B) è controllabile. Allora:

- Esiste un'unica soluzione definita positiva $\overline{P}$ di ARE;
- Il sistema a ciclo chiuso:

  $$\dot{x} = (A - B\,R^{-1}B^T\overline{P})x$$

ha un equilibrio in zero asintoticamente stabile.

Per dimostrare l'esistenza della soluzioni occorre definire inizialmente una famiglia di successioni $\{P_{T_i}(t)\}_T$ di soluzioni in $[0,T_i]$ al variare di i. Questa famiglia di successione ha un limite per $T\to\infty$:

- è monoticamente non descrente;
- ogni elemento è limitato superiormente;

Per il primo punto prendiamo due istanti temrinali $T_1 < T_2$. Dalla definizione di funzione valore:

$$V_1^*(t,x) = \int_t^{T_1} l(x_1(\tau),u(\tau))d\tau \leqslant \int_t^{T_1} l(x_2(\tau),u_2(\tau))d\tau$$
$$\leqslant \int_t^{T_1} l(x_2(\tau),u_2(\tau))d\tau + \int_{T_1}^{T_2} l(x_2(\tau),u_2(\tau))d\tau = V_2^*(t,x)$$

Quindi sussiste una relazione d'ordinare del tipo:

$$P_{T_1}(t) \leqslant P_{T_2}(t)$$

Quindi $P(t)$ è monotonicamente non decrescente. A questo punto per dimostrare la limitatezza di ogni suo elemento facciamo riferimento all'ipotesi di controllabilità di (A,B). Quindi, esiste una matrice k t.c. $A+BK$ abbia tutit gli autovalori a parte reale negativa:

$$\sigma(A+BK)\subset\mathbb{C}^-$$

Il sistema a ciclo chiuso con $u=Kx$ diventa $\dot{x}=(A+BK)x \Rightarrow x(t)=e^{(A+BK)t}x_0$. Il relativo costo:

$$J(u)=\frac{1}{2}\int_0^T (x(\tau)^T Q\, x(\tau)+u(\tau)^T R\, u(\tau))d\tau = \frac{1}{2}\int_0^T x(\tau)^T\,(Q+K^T R\,K)\,x(\tau)d\tau$$
$$= \frac{1}{2}x_0\left(\int_0^T e^{(A+BK)^T\tau}\,(Q+K^T R\,K)\,e^{(A+BK)}d\tau\right)x_0$$

Per dimostrare che è limitato dobbiamo effettuare il limite per $T\to\infty$, questo limite lo chiamiamo $\hat{P}$.Quindi $\forall x_0\in\mathbb{R}^n$:

$$\frac{1}{2}\,x_0^T P_T(0)x_0 \leqslant \frac{1}{2}\,x_0^T\left(\int_0^T e^{(A+BK)^T\tau}(Q+K^T R\,K)\,e^{(A+BK)\tau}d\tau\right)x_0$$
$$\leqslant \frac{1}{2}x_0^T\left(\int_0^\infty e^{(A+BK)^T\tau}(Q+K^T R\,K)\,e^{(A+BK)\tau}d\tau\right)x_0 = \frac{1}{2}x_0^T\hat{P}x_0$$
$$\Longrightarrow P_T(0)\leqslant\hat{P}$$

Enunciare il teorema di esistenza della soluzione del LQR ad orizzonte infinito. Dimostrare che il sistema a ciclo chiuso con la soluzione di LQR risulta asintoticamente stabile.

Answer. Il teorema di esistenza della soluzione di un problema LQR ad orizzonte infinito fonrisce condizioni sufficienti di essitenza della soluzione. In particolare, considerato il problema LQR su orizzonte infinito con $M=0, R\succ 0, Q=D^TD\succcurlyeq 0$. Supponiamo che la coppia (A,D) sia osservabile e (A,B) controllabile. Allora:

- Esiste un'unica soluzione definita positiva $\overline{P}$ dell'equazione algebrica di riccati;

- Il sistema a ciclo chiuso:

$$\dot{x} = (A - B\,R^{-1}B^T\,\overline{P})x$$

  ha un equilibrio in zero asintoticamente stabile.

Per dimostrare che il sistema a ciclo chiuso è asintoticamente stabile occorre dimostrare perliminarmente la stabilità del sistema a ciclo chiuso e poi l'attrattività dei moti tramite il teorema di LaSalle, prendiamo come funzione di Lyapunov la funzione valore $V(x) = \frac{1}{2}x^T Px$ e quest'ultima deve risultate:

1. La matrice $\overline{P}$ è definita positva;
2. $\dot{V}$ è definita negativa.

Per il primo punto si suppone per assurdo che $\overline{P}$ sia semidefinita positiva. Quindi essiste una condizione iniziale non nulla per cui:

$$\overline{P}x_0 = 0 \Longrightarrow x_0^T\,\overline{P}x_0 = 0 \Longrightarrow \int_0^\infty (x(t)^T D^T D\,e^{At} + u(t)^T R\,u(t))dt = 0$$

Dato che $R \succ 0 \Longrightarrow u(t) = 0 \quad \forall t \Longrightarrow \dot{x} = A\,x \Longrightarrow x(t) = e^{At}x_0 = 0$. Essendo (A, D) osservabile, la matrice Gramiana di osservabilità è definita positiva e l'unica condizione iniziale possibile è $x_0 = 0$. Ciò è un'assurdo.

Per dimostrare il secondo punto,effettuiamo la derivata lungo le traiettorie del sistema a ciclo chiuso:

$$\dot{V} = x^T\overline{P}(A - B\,R^{-1}\,B^T\overline{P})x = \frac{1}{2}x^T(\overline{P}A + A^T\overline{P} - 2\overline{P}B\,R^{-1}B^T\overline{P})x = -\frac{1}{2}x^T Q\,x - \frac{1}{2}x^T\overline{P}B\,R^{-1}B^T\overline{P}x$$

Per ottenere la definita negatività, occorre avere per ogni x almeno uno dei due termini positivi:

- $Q \succ 0$, ma per ipotesi $Q \succcurlyeq 0$;
- $\tilde{P}B\,R^{-1}\,B^T\tilde{P} \succ 0$,ma $B\,R^{-1}\,B^T$ ha rango pieno solo se m=n

Quindi possiamo concludere che $\dot{V} \preccurlyeq 0$.Quindi la stabilità semplice del sistema. Per affermare la stabilità asintotica del sistema dobbiamo utilizzare il teorema di LaSalle sull'attrattività dei moti di un sistema autonomo.

Teorema di LaSalle

Consideriamo un sistema autonomo $\dot{x} = A\,x$.

- Sia $\Omega \subset \mathbb{R}^n$ un insieme compatto positivamente invariante per il sistema
- Sia $V: \mathbb{R}^n \longrightarrow \mathbb{R}$ una funzione C^1 tale che $\dot{V} \leqslant 0$
- Sia E insieme dei punti in Ω tale che $\dot{V} = 0$
- Sia M il più grande insieme invariante contenuto in E

Allora ogni soluzione converge a M per $t \to \infty$.

Quindi,Sappiamo che:

- $V(x)$ è quadratica e i suoi insiemi di livello sono ellissoidi compatti;

- $V(x(t)) \leqslant V(x_0)\, \forall t \geqslant 0$

Dunque, possiamo scegliere un insieme compatto positivamente invariante per il sistema:

$$\Omega = \{x \in \mathbb{R}^n : V(x) \leqslant V(x_0)\}$$

Quindi, per caratterizzare l'insieme invariante $M \subset E = \{x : \dot{V} = 0\}$

$$\dot{V} = 0 \Longrightarrow x(t)^T D^T D\, x(t) = 0 \Longrightarrow D\, x(t) = 0 \quad \forall t$$

$$M = \{x \in \mathbb{R}^n : D\, e^{At}\, x = 0, \forall t\}$$

Grazie all'ipotesi di osservabilità, è possibile distinguere due uscite di due qualsiasi condizioni iniziali diverse. Quindi:

- La condizione iniziale $x_0 = 0$ forsnisce $D\, e^{At}\, x_0 = 0 \quad \forall t$;
- è l'unico stato per cui vale questa proprietà;

Allora $M = \{0\}$. Quindi abbiamo dimostrato l'attrività di tutte le soluzioni che, unita alla stabilià, ci forniscono stabilità asintotica.

Enunciare e dimostrare il Principio di Ottimalità.

Answer. Considerato un problema di Bolza. Se u^* è ottima sull'intervallo $[t, T]$ a partire dallo stato $x(t)$, allora $u^*(\tau), \tau[t + \Delta t, T]$ è necessariamente ottima per un problema di Bolza ristretto all'intervallo $[t + \Delta t, T]$ a partire da $x^*(t + \Delta t)\, \forall \Delta t$ tale che $0 < \Delta t \leqslant T - t$.

Per dimostrare il Principio di Ottimalità, supponiamo per assurdo che esista u^{**} che fornisce un valore minore dell'indice di costo del problema di Bolza ristretto:

$$\overline{J}(u) = \int_{t+\Delta t}^{T} l(x(\tau), u(\tau), \tau) d\tau + m(x(T))$$

rispetto ad u^* in $[t + \Delta T, T]$. Quindi $\overline{J}(u^{**}) < \overline{J}(u^*)$. Se ciò è vero, possiamo definire il nuovo controllo $\hat{u}$:

$$\hat{u}(\tau) = \begin{cases} u^*(\tau) & \text{se } t \leqslant \tau \leqslant t + \Delta t \\ u^{**}(\tau) & \text{se } t + \Delta t \leqslant \tau \leqslant T \end{cases}$$

Allora, sull'intero intervallo $[t, T]$ abbiamo:

$$\begin{aligned} J(\hat{u}) &= \int_t^T l(\hat{x}(s), \hat{u}(s), s) ds + m(\hat{x}(T)) \\ &= \int_t^{t+\Delta t} l(x^*(s), u^*(s), s) ds + \int_{t+\Delta t}^{T} l(x^{**}(s), u^{**}(s), s) ds + m(x^{**}(T)) \\ &< \int_t^{t+\Delta t} l(x^*(s), u^*(s), s) ds + \int_{t+\Delta t}^{T} l(x^*(s), u^*(s), s) ds + m(x^*(T)) = J(u^*) \end{aligned}$$

Di conseguenza abbiamo contraddetto l'ipotesi per cui u^* sia la soluzione ottima su $[t, T]$.

Discutere il problema del Tracking e della reiezione di disturbi noti.

Answer. Il problema di Tracking è un problema di minimizzazione del tipo:

$$\min_u J(u) = \frac{1}{2}\int_0^T ((\xi(t) - \dot{\xi}(t))^T Q(\xi(t) - \dot{\xi}(t)) + u(t)^T R\, u(t))dt$$

$$\dot{\xi}(t) = A\xi(t) + B\, u(t), \xi(0) = \xi_0$$

In cui si vuole ottenere lo stato del sistema più vicino possibile ad un segnale nel tempo in un intervallo di tempo fissato. Alcune tipologie di questo problema possono essere ricondotte ad un LQR con l'ipotesi che, fissando il segnale desiderato $\tilde{\xi}$:

$$A\,\tilde{\xi} - \dot{\tilde{\xi}} = 0 \longrightarrow \tilde{\xi} = \sum_{i=1}^{\nu}\sum_{j=0}^{\mu_i} c_{ij}\, e^{\lambda_i t}\, t^j$$

con condizioni iniziale c_{ij}, ν autovalori di A,μ_i molteplicità algebrica di λ_i,λ_i autovalore i-esmo di A.L'indice di costo è descritto da:

$$J(u) = \frac{1}{2}\int_0^T (x(t)^T Q\, x(t) + u^T(t)\, R\, u(t))d\tau$$

e il sistema dinamico:

$$\begin{aligned} \dot{x} &= \dot{\xi} - \dot{\tilde{\xi}} = A\,\xi + B\,u - \dot{\tilde{\xi}} \\ &= A\,(x + \tilde{\xi}) + B\,u - \dot{\tilde{\xi}} = A\,x + B\,u + \left(A\tilde{\xi} - \dot{\tilde{\xi}}\right) \\ &= A\,x + B\,u \end{aligned}$$

Per il principio del modello interno, si può avere tracking perfetto se il processo contiente una copia dei modi del segnale di riferimento. Nel caso in cui questo principio non venga rispettato si può considerare la dinamica:

$$\dot{x} = A\,x + B\,u + w(t) \quad w(t) := \text{disturbo}$$

Così facendo, si ottiene u sistema affine e la minimizzazione dell'ìndice di costo:

$$J(u) = \frac{1}{2}\int_0^T (x(t)^T Q\, x(t) + u^T(t)\, R\, u(t))d\tau$$

viene detto problema di reiezione dei disturbi noti in cui vale:

- HJB:

  $$-\frac{\partial V}{\partial t} = \min_u \left\{\frac{\partial V}{\partial x}(A\,x + B\,u + w(t)) + \frac{1}{2}x^T Q\,x + \frac{1}{2}u^T(t)\,R\,u(t)\right\}, V(x,T) = 0 \quad \forall x \in \mathbb{R}^n, t \in [0, T]$$

- La funzione valore:

  $$V(x,t) = \frac{1}{2}x^T P(t)\,x + b(t)^T x + c(t), P(t) \in \mathbb{R}^{n\times n}, b(t) \in \mathbb{R}^{n\times 1}, c(t) \in \mathbb{R}$$

e il controllo ottimo:

$$u^* = -R^{-1}\,B^T(P(t)\,x + b(t)) = -R^{-1}B^T P(t)\,x - R^{-1}\,B^T b(t)$$

Infine la soluzione di HJB, si ottiene sostituendo il controllo ottimo nell'equazione HJB.

Dimostrare che l'equazione di Hamilton Jacobi Bellman fornisce condizioni sufficienti di ottimalità.

Answer. Vogliamo dimostrare che se la funzione $\hat{V}$ risolve HJB, allora questa funzione è la funzione valore. A tale scopo, se $\hat{V}$ risolve HJB e $\hat{u}$ raggiunge il minmo del termine di destra di HJB:

$$\hat{u}(t) = \arg\min_u \left\{ l(x,u,t) + \frac{\partial \hat{V}}{\partial x}(x,t)\; f(x,u,t) \right\} \quad \forall x, \forall t$$

Allora lungo la traiettoria $\hat{x}(t)$ a partire dalla condizione iniziale $\hat{x}(t_0 = x_0)$, si ha:

$$-\frac{\partial \hat{V}}{\partial t}(\hat{x}(t),t) = l(\hat{x}(t),\hat{u}(t),t) + \frac{\partial \hat{V}}{\partial x}\,(\hat{x}(t),t) f(\hat{x}(t),\hat{u}(t),t)$$

Ricordando che;

$$\frac{d}{dt}\hat{V}(\hat{x}(t),t) = \frac{\partial}{\partial x}\hat{V}(\hat{x}(t),t)\; f(\hat{x}(t),\hat{u}(t),t) + \frac{\partial}{\partial t}\,\hat{V}(\hat{x}(t),t)$$

Si ottiene:

$$0 = l(\hat{x}(t),\hat{u}(t),t) + \frac{d}{dt}\hat{V}(\hat{x}(t),t)$$

Integrando:

$$0 = \int_{t_0}^{T} l(\hat{x}(t),\hat{u}(t),t)dt + \int_{t_0}^{T} \frac{d}{dt}\hat{V}(\hat{x}(t),t)dt$$

$$0 = J(\hat{u}) - \hat{V}(\hat{x}(t_0),t_0) \rightarrow J(\hat{u}) = \hat{V}(\hat{x}(t_0),t_0)$$

Ora dobbiamo dimostrare che questo costo è minimo e quindi confrontarlo con quello ottenuto da una qualsiasi altra u:

$$-\frac{\partial \hat{V}}{\partial t}(x(t),t) \leqslant l(x(t),u(t),t) + \frac{\partial \hat{V}}{\partial x}(x(t),u(t),t)$$

$$0 \leqslant l(x(t),u(t),t) + \frac{d}{dt}\hat{V}(x(t),t)$$

Integrando:

$$0 \;\leqslant\; \int_{t_0}^{T} l(x(t),u(t),t)dt + \int_{t_0}^{T} \frac{d}{dt}\hat{V}(x(t),t)dt = \int_{t_0}^{T} l(x(t),u(t),t)dt + m(x(T)) - \hat{V}(x_0,t_0)$$

Quindi:

$$\hat{V}(x_0,t_0) \leqslant J(u)$$

$$J(\hat{u}) \leqslant J(u) \quad \forall u$$

Dare definizione di Value Iteration

Answer. 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:

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

Dare definizione di Policy Improvement

Answer. Il policy improvement è un algoritmo iterativo che, per un certo istante di tempo e per un certo stato,ci dice che 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:

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

Dare definizione di Policy Evaluation

Answer. 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:

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

Dare definizione di Funzione Valore in un contesto di Reinforcement Learning

Answer. 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.

Dare definizione di Funzione Qualità

Answer. 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

Discutere di MPC e dimostrare la ammissibilità,stabilità.

Answer. Nel model predictive control l'obiettivo è quello di minimizzare un indice di costo $J(u) = \sum_{t=0}^{\infty} (x^T Q\, x + u_t^T R u_t)$ in cui, ad ogni istante di tempo, si misura il valore dello stato corrente $x(t)$, si risolve un problema di ottimizzazione dinamica a batch in una finestra di N_u passi di cui si implementa solo la prima azione ottima u_0^*. Questo problema è di ottimizzazione statica a cui si possono aggiungere eventualmente dei vincoli.

Per dimostrare l'ammissibilità e la stabilità, consideriamo il sistema lineare $x_{t+1} = A\, x_t + B\, u_t$, $t = 0, 1, \ldots n$ e supponiamo che la strategia di MPC sia basata sul seguente problema QP:

$$V^*(x(t)) = \min_u \sum_{k=0}^{N_u - 1} x_k^T Q\, x_k + u_k^T\, R\, u_k$$

$$s.t.$$

$$x_{k+1} = A\, x_k + B\, u_k, x_0 = x(t)$$

$$u_{\min} \leqslant u_k \leqslant u_{\max}$$

$$y_{\min} \leqslant C\, x_k \leqslant y_{\max}$$

$$x_{N_u} = 0 \longrightarrow \textbf{vincolo terminale}$$

con $Q > 0, R > 0, u_{\min} < 0 < u_{\max}$, $y_{\min} < 0 < y_{\max}$.Allora, **se il problema QP è ammissibile al tempo** t=0:

$$\lim_{t \to \infty} x_t = 0 \qquad \lim_{t \to \infty} u_t = 0$$

ed i vincoli sono soddisfatti per ogni t.

Dimostrazione.

Supponiamo che $U_t^* = [u_0^t, \ldots, u_{N-1}^t]$ sia la soluzione ottima di QP al tempo t con valore $V^*(x(t))$. Dal momento che U_t^* è ammissibile, sappiamo che $x_{t+N_U} = 0$;

Di conseguenza, la sequenza $U_{t+1} = [u_1^t, \ldots, u_{N-1}^t, 0]$ è ammissibile per QP al tempo t+1. Infatti:

- $x(t+1) = x(t+1|t)$ dato che abbiamo utilizzato u_0^t;
- La sequenza U_{t+1} è tale che $x(t+N_u|t+1) = x(t+N_u+1|t+1) = 0$

Ora, il costo U_{t+1} è:

$$V^*(x(t)) - x(t)^T Q\, x(t) - u(t)^T R\, u(t) \geqslant V^*(x(t+1))$$

è maggiore del valore ottimo a $t+1$:

$$\Longrightarrow V^*(x(t)) \text{ è monoticamente descrente } e \text{ lmitata inferiormente}$$

$$\exists \lim_{t\to\infty} V^*(x(t)) =: V_\infty$$

Dunque $0 \leqslant x(t)^T Q\, x(t) + u(t)^T R\, u(t) \leqslant V^*(x(t)) - V^*(x(t+1)) \to 0 \quad$ per $t \longrightarrow \infty$ e la tesi segue dal fatto che Q ed R sono matrici definite positive.

□

Enunciare il teorema di essitenza della soluzione del LQR ad orizzonte infinito. L'unicità della soluzione stabilizzante di ARE.

Answer. Il teorema di esistenza della soluzione di un problema LQR ad orizzonte infinito fonrisce condizioni sufficienti di essitenza della soluzione. In particolare, considerato il problema LQR su orizzonte infinito con $M = 0, R \succ 0, Q = D^T D \succcurlyeq 0$. Supponiamo che la coppia (A, D) sia osservabile e (A, B) controllabile. Allora:

- Esiste un'unica soluzione definita positiva $\overline{P}$ dell'equazione algebrica di riccati;
- Il sistema a ciclo chiuso:

  $$\dot{x} = (A - B\, R^{-1} B^T \overline{P}) x$$

  ha un equilibrio in zero asintoticamente stabile.

Per dimostrare l'unicità della soluzione stabilizzante di ARE, supponiamo che esistano due matrici $\tilde{P}_1$ e $\tilde{P}_2$ che soddisfano:

$$\begin{aligned} 0 &= \tilde{P}_1 A + A^T \tilde{P}_1 + Q - \tilde{P}_1 B\, R^{-1} B^T \tilde{P}_1 \quad A_{\text{cl},1} = A - B\, R^{-1} B^T \tilde{P}_1 \quad \sigma(A_{\text{cl},1}) \subset \mathbb{C}^- \\ 0 &= \tilde{P}_2 A + A^T \tilde{P}_2 + Q - \tilde{P}_2 B\, R^{-1} B^T \tilde{P}_2 \quad A_{\text{cl},2} = A - B\, R^{-1} B^T \tilde{P}_2 \quad \sigma(A_{\text{cl},2}) \subset \mathbb{C}^- \end{aligned}$$

Sottraiedno la seconda equazione dalla prima e sommando e sottraendo $\tilde{P}_2 B\, R^{-1} B^T \tilde{P}_2$ otteniamo:

$$0 = (\tilde{P}_1 - \tilde{P}_2)\, A_{\text{cl},1} + A_{\text{cl},2}^T (\tilde{P}_1 - \tilde{P}_2) \quad \textbf{Equazione di Sylvester}$$

Equazione di Sylvester

Un'equazione di Sylvester è un'equazione matriciale lineare nell'incognita X della forma

$$X\,A + B\,X = 0$$

A, B, C matrici note di coefficienti. Quale che sia C, l'equazione ammette un'unica soluzione se e solo se $\sigma(A) \cap \sigma(-B) = \varnothing, \sigma(-A_{\mathrm{cl},2}^{T}) \subseteq \mathbb{C}^{+}$

L'unica soluzione è dunque :

$$(\tilde{P}_1 - \tilde{P}_2) = 0 \longrightarrow \tilde{P}_1 = \tilde{P}_2$$

Principio Ottimalità (Programmazione Quadratica)

Answer. Una politica decisionale ottima ha la proprietà che, qualche che sia la configurazione iniziale e le decisioni iniziali, se consideriamo un punto intermedio, le decisioni rimanenti della polica devono costituire una soluzione ottima rispetto alla configurazione raggiunta dalla prima parte delle decisione.

Equazione di Bellman e Soluzione ricorsia di Bellman

Answer. $\begin{cases} v_k(x_k) = \min_{u_k \in U_k} \{g_k(x_k, u_k) + V_{k+1}(f(x_k, u_k))\} \quad \forall x_k \in X \quad k = \{0, 1, \ldots, N\} \\ V_N(x_N) = g_N(x_N) \quad \forall x_N \in X \end{cases}$

Questa equazione si risolve ricorsivamente all'indietor, cioè si parte dal punto finale e tornando indietro, si ricostruiscono gli ingressi ottimali. In altre parole:

Dopo aver calcolato il valore di lungo periodo, l'**azione ottima** è quella che trasferisce il sistema nello stato a maggior valore:

$$u_k^*(x_k) = \arg \min_{u_k \in U_k} \{g_k(x_k, u_k) + V_{k+1}(x_{k+1})\}$$

Come abbiamo già detto l'equazione di Bellman si può risolvere ricorsivamente all'indietro, quindi consideriamo l'istante terminale N, il costo migliore a partire da x^1 al tempo N è pari al costo terminale g_N valutato in x^1. Possiamo ora assegnare il costo ottimo al tempo N a ciascuno stato in X, calcolando $V_N(\cdot)$.Facendo un passo indietro a N-1, il costo della scelta 1 è $c_1 + V_N(x^1)$.

Conosciamo tutti i costi $V_N(x^i)$, quindi possiamo determinare la migliore scelta da x^1 al tempo N-1:

$$V_{N-1}(x^1) = \min_i \{c_i + V_N(x^i)\}$$

Ora, ragionando al generico istante k, per il principio di ottimalità, la coda della soluzione ottima da k deve necessariamente coincidere con $V_{k+1}(x^1)$. Scelte greedy basate su considerazioni istantanee rispetto ad una funzione che racchiude anche le conseguenze future.

Processo Decisionale di Markov Stocastico

Answer. Un processo Markoviano stocastico è un processo Markoviano in cui il comportamento dell'ambiente (che possiede la proprietà markoviana di stato), il coomportamento dell'agente e la ricompensa è definita come segue:

- lo stato successivo sia x' e la ricompensa r, specificando la **funzione di massi di probabilità:**

$$p(x', r|x, u) = \Pr[R_{k+1} = r, X_{k+1} = x'|X_k, U_k]$$

In cui R, X, U rappresentano le variabili aleatorie discrete da cui possiamo calcolare la **probabilità di transizione di stato**:

$$p(x',|x,u) = \sum_r p(x', r\,|x,u)$$

- si descriva una **policy** come una distribuzione di probabilità $\pi(u|x)$
- l'**obbiettivo** sia di massimizzazione del valore atteso del ricavo:

$$E_\pi[G_k|X_k = x] = E_\pi\left[\sum_{i=0}^{\infty} R_{k+i+1}|X_k = x\right]$$

Ottenendo la corrispettiva funzione valore v_π:

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

e la funzione valore stato:

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

Nel caso ottimo v_*:

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

Discutere l'equazione differenziale di Riccati e dimostrare esistenza soluzione tramite matrice Hamiltoniana

Answer. Supponiamo che H non abbia autovalori puramente immaginari, dunque n a parte reale positiva (instabili) e n a parte reale negativa (stabili).

Esiste semptre una trasformazione non-singolare U tale che:

$$U^{-1} H U = \begin{pmatrix} \Lambda_s & 0 \\ 0 & \Lambda_u \end{pmatrix}$$

in cui Λ_s / Λ_u raccolgono tutti i blocchi di Jordan associati ad autovalori **stabili/instabili**.

Partizioniamo, di conseguenza, anche U in maniera coerente come:

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$$

dove:

$$\begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} / \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix}$$

hanno come colonne gli autovettori generalizzati di H corrispondenti agli autovalori **stabili/instabili.**

Ora, consideriamo il **cambio di coordiante** per il sistema Hamiltoniano:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = U \begin{pmatrix} \hat{X} \\ \hat{Y} \end{pmatrix} \Rightarrow \begin{pmatrix} \hat{X} \\ \hat{Y} \end{pmatrix} = U^{-1} \begin{pmatrix} X \\ Y \end{pmatrix}$$

Il sistema nelle nuove coordinate diventa:

$$\begin{pmatrix} \hat{X} \\ \hat{Y} \end{pmatrix} = U^{-1} \begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} = U^{-1} H \begin{pmatrix} X \\ Y \end{pmatrix} = U^{-1} H U \begin{pmatrix} \hat{X} \\ \hat{Y} \end{pmatrix} = \begin{pmatrix} \Lambda_s & 0 \\ 0 & \Lambda_u \end{pmatrix} \begin{pmatrix} \hat{X} \\ \hat{Y} \end{pmatrix}$$

Questo è un sistema **lineare e disaccoppiato** per $\hat{X}$ e $\hat{Y}$.

La soluzione al tempo T può essere trovata nel seguente modo:

$$\begin{aligned} \hat{X}(T) = e^{\Lambda_s (T-t)} \hat{X}(t) \implies \hat{X}(t) = e^{-\Lambda_s (T-t)} \hat{X}(T) \\ \hat{Y}(T) = e^{\Lambda_u (T-t)} \hat{Y}(t) \implies \hat{Y}(t) = e^{-\Lambda_u (T-t)} \hat{Y}(T) \end{aligned}$$

Ora imponiamo le condizioni al contorno ottenendo:

$$\begin{pmatrix} X(T) \\ Y(T) \end{pmatrix} = \begin{pmatrix} I \\ M \end{pmatrix} = \begin{pmatrix} U_{11} \hat{X}(T) + U_{12} \hat{Y}(T) \\ U_{21} \hat{X}(T) + U_{22} \hat{Y}(T) \end{pmatrix}$$

che si utulizzano per ricavare $\hat{Y}(T)$ in funzione di $\hat{X}(T)$:

$$M(U_{11} \hat{X}(T) + U_{12} \hat{Y}(T)) = U_{21} \hat{X}(T) + U_{22} \hat{Y}(T)$$

$$\Rightarrow \hat{Y}(T) = -(U_{22} - M U_{12})^{-1} (U_{21} - M U_{11}) \hat{X}(T) = G \hat{X}(T)$$

Dalla relazione $[X(t)^T, Y(t)^T]^T = U [\hat{X}(t)^T, \hat{Y}(t)^T]^T$, si ottiene:

$$X(t) = U_{11} \hat{X}(t) + U_{12} \hat{Y}(t) = U_{11} e^{-\Lambda_s (T-t)} \hat{X}(t) + U_{12} e^{-\Lambda_u (T-t)} \hat{Y}(t)$$

$$= U_{11} e^{-\Lambda_s (T-t)} \hat{X}(t) + U_{12} e^{-\Lambda_u (T-t)} G \hat{X}(t) =$$

$$= [U_{11} + U_{12} e^{-\Lambda_u (T-t)} G e^{\Lambda_s (T-t)}] e^{-\Lambda_s (T-t)} \hat{X}(t)$$

Ripetendo lo stesso ragionamento per $\hat{Y}(t)$:

$$Y(t) = [U_{21} + U_{22} e^{-\Lambda_u (T-t)} G e^{\Lambda_s (T-t)}] e^{-\Lambda_s (T-t)} \hat{X}(t)$$

Avendo ora calcolato esplicitamente $X(t)$ e $Y(t)$, possiamo ottenere $P(t) = Y(t) X(t)^{-1}$:

$$P(t) = [U_{21} + U_{22} e^{-\Lambda_u (T-t)} G e^{\Lambda_s (T-t)}] [U_{11} + U_{12} e^{-\Lambda_u (T-t)} G e^{\Lambda_s (T-t)}]^{-1}$$

Allora la soluzione della DRE ottenuta clacolando solo autovalore e autovettori di H.