Controle Ótimo - Aula 1 (Prova do Algoritmo da PD e Exemplo 1)

Adriano Almeida Gonçalves Siqueira e Marco H. Terra

Algoritmo da Programação Dinâmica:

Para cada condição inicial x_0, o custo ótimo $J^*(x_0)$ do problema básico é igual a $J_0(x_0)$, o último passo do seguinte algoritmo que evolui de modo reverso no tempo de $N-1$ até 0 :

- Passo N:

$$J_N(x_N) := g_N(x_N)$$

  para todo $x_N \in \mathcal{S}_k$
- Passo k:

$$J_k(x_k) = min_{u_k \in U_k(x_k)}\{E_{w_k}\{g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k))\}\}, \quad k = N-1, ..., 0$$

- Passo 0:

$$J^*(x_0) = J_0(x_0)$$

- Se $u_k^* = \mu_k^*(x_k)$ minimiza o lado direito da equação do passo k, a política $\pi^* = \{\mu_0^*, \mu_1^*, ..., \mu_{N-1}^*\}$ é ótima

Prova:
(I) Notação:
Para qualquer política admissível $\pi = \{\mu_0,\ \mu_1,\ ...,\ \mu_{N-1}\}$ e para cada $k = 0, 1, ...,\ N-1$ denote:

Políticas de controle admissíveis do instante k ao instante final N:

$$\pi^k := \{\mu_k, \mu_{k+1}, ..., \mu_{N-1}\}$$

Distúrbios do instante k ao instante final N:

$$d^k := \{w_k, w_{k+1}, ..., w_{N-1}\}$$

Custo ótimo para o problema que começa no estado $x_k =$x$_k$ no instante k e termina no instante N:

$$J_k^*(x_k) := min_{\pi^k} E_{d^k}\left\{g_N(x_N) + \sum_{i=k}^{N-1} g_i(x_i, \mu_i(x_i), w_i)\right\}$$

Custo ótimo para o problema que começa no estado x_N no instante final N:

$$J_N^*(x_N) := g_N(x_N)$$

(II) As funções de custo ótimas J_k^* são exatamente as funções J_k computadas pelo Algoritmo da Programação Dinâmica para todo k.

Prova por indução: A hipótese de indução vale para $\boldsymbol{N}$:

Segue pelas definições que $J_N^* = g_N = J_N$.

Suponha que a hipótese de indução valha para $k+1$, ou seja, suponha que a função custo ótima J_{k+1}^* seja dada exatamente pela função J_{k+1} computada pelo Algoritmo da Programação Dinâmica, ou seja, temos

$J_{k+1}^*\left(x_{k+1}\right) = J_{k+1}\left(x_{k+1}\right)$, para todo $x_{k+1} \in \mathcal{S}_{k+1}$

Vamos mostrar que a hipótese é válida para k:

Fixe $x_k \in \mathcal{S}_k$ e note que $\pi^k = \{\mu_k, \pi^{k+1}\}$ e $d^k = \{w_k, d^{k+1}\}$. Temos por definição:

$$
\begin{aligned}
J_k^*\left(x_k\right) :=& min_{\pi^k} E_{d_k} \left\{ g_N\left(x_N\right) + \sum_{i=k}^{N-1} g_i\left(x_i, \mu_i\left(x_i\right), w_i\right) \right\} \\
=& min_{\{\mu_k, \pi^{k+1}\}} E_{\{w_k, d^{k+1}\}} \left\{ g_k\left(x_k, \mu_k\left(x_k\right), w_k\right) + g_N\left(x_N\right) + \sum_{i=k+1}^{N-1} g_i\left(x_i, \mu_i\left(x_i\right), w_i\right) \right\}
\end{aligned}
$$

Sendo a distribuição de w_k independente da distribuição de $d^{k+1} := \{w_{k+1}, ..., w_{N-1}\}$,

$$
J_k^*\left(x_k\right) = min_{\{\mu_k, \pi^{k+1}\}} E_{w_k} \left\{ g_k\left(x_k, \mu_k\left(x_k\right), w_k\right) + E_{d^{k+1}} \left[g_N\left(x_N\right) + \sum_{i=k+1}^{N-1} g_i\left(x_i, \mu_i\left(x_i\right), w_i\right) \right] \right\}
$$

Como o termo $g_k\left(x_k, \mu_k\left(x_k\right), w_k\right)$ não depende da política de controle futura π^{k+1}

$$
J_k^*\left(x_k\right) = min_{\mu_k} E_{w_k} \left\{ g_k\left(x_k, \mu_k\left(x_k\right), w_k\right) + min_{\pi^{k+1}} E_{d^{k+1}} \left[g_N\left(x_N\right) + \sum_{i=k+1}^{N-1} g_i\left(x_i, \mu_i\left(x_i\right), w_i\right) \right] \right\}
$$

Segue da definição do custo ótimo a partir do instante $k+1$ que

$$
J_k^*\left(x_k\right) = min_{\mu_k} E_{w_k} \left\{ g_k\left(x_k, \mu_k\left(x_k\right), w_k\right) + J_{k+1}^*\left(f_k\left(x_k, \mu_k\left(x_k\right), w_k\right)\right) \right\}
$$

Sendo válida a hipótese para $k+1$

$$
J_k^*\left(x_k\right) = min_{\mu_k} E_{w_k} \left\{ g_k\left(x_k, \mu_k\left(x_k\right), w_k\right) + J_{k+1}\left(f_k\left(x_k, \mu_k\left(x_k\right), w_k\right)\right) \right\}
$$

Sendo x_k fixo, variar a função: $\mu_k : S_k \to \mathcal{U}_k\left(x_k\right)$, $x_k \mapsto u_k$ é o mesmo que variar u_k, assim,

$$
\begin{aligned}
J_k^*\left(x_k\right) =& min_{u_k \in \mathcal{U}_k(x_k)} E_{w_k} \left\{ g_k\left(x_k, u_k, w_k\right) + J_{k+1}\left(f_k\left(x_k, u_k, w_k\right)\right) \right\} \\
=& J_k\left(x_k\right)
\end{aligned}
\qquad \square
$$

Exemplo1: Problema de Rota de Avião com gasto mínimo de combustível [Lewis p.294, ex 4.1-1]
Um avião pode voar da esquerda para a direita ao longo dos caminhos mostrados na figura 1.
As intercecções a,b,c,... representam cidades. Os números representam o combustível requerido para completar cada caminho (custo para cada instante $g_k\left(x_k, u_k\right)$). A cada instante k pode-se decidir movimentar-se para o próximo estágio $k+1$ indo para cima ($u_k = +1$), ou indo para baixo ($u_k = -1$).

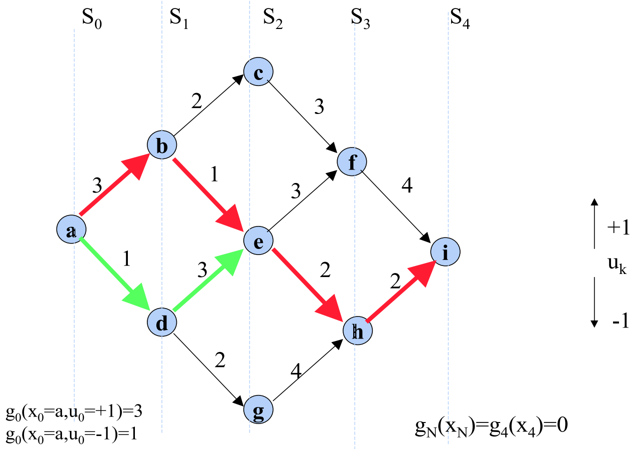

Fig. 1. Problema de Rota Ótima.

O problema é determinar o caminho que o avião deve percorrer partindo de a e chegando a i com gasto mínimo de combustível.

Vamos colocar o problema na forma padrão ([B95]p.10)

1) O estado corrente é o nó onde estamos fazendo a decisão corrente. Assim, o estado inicial é $x_0 = a$. No instante $k = 1$, o estado pode ser $x_1 = b$ ou $x_1 = d$. Pela figura, o espaço de estados para cada instante k é dado por

$$S_0 = \{a\},\ S_1 = \{b, d\},\ S_2 = \{c, e, g\},\ S_3 = \{f, h\},\ S_4 = \{i\}$$

2) A variável de decisão (controle) u_k a ser selecionada no instante k pertence ao espaço $C_k = \{+1, -1\}$.

Note que o controle u_k é retringido a tomar valores em um sub-conjunto $U_k\left(x_k\right) \subset C_k$ que depende do instante k e do valor corrente do estado x_k. Por exemplo, para $k = 2$, $x_2 = c$ e assim, temos $U_2\left(x_2 = c\right) = \{-1\} \subset C_1 = \{+1, -1\}$.

3) O sistema dinâmico $x_{k+1} = f_k\left(x_k, u_k\right)$ pode ser dado de forma tabular pelos dados contidos na figura (por exemplo, é evidente que $x_2 = c$ quando $x_1 = b$ e $u_1 = +1$).

4) O custo a cada instante $g_k\left(x_k, u_k\right)$ pode ser dado de forma tabular pelos dados contidos na figura (por exemplo, é evidente que $g_2\left(x_2 = b, u_2 = 1\right) = 2$).

Podemos determinar a rota ótima de $x_0 = a$ para $x_N = i$ fazendo-se comparações com todos os possíveis caminhos $u = \left(u_0, ..., u_{N-1}\right)$ de $x_0 = a$ para $x_N = i$ conforme a expressão:

$$J^*\left(x_0 = a\right) = min_{u=\left(u_0,\ldots,u_{N-1}\right)}\{g_N\left(x_N\right) + \sum_{k=0}^{N-1} g_k\left(x_k, u_k\right)\}$$

A solução por este procedimento requer um grande número de cálculos.

Baseados no princípio de otimalidade de Bellman, vamos aplicar o algorítmo da Programação Dinâmica para reduzir o número de cálculos necessários (por restringir o número de decisões a serem tomadas).

Comece com $k = N = 4$.

Nenhuma decisão é necessária aqui. Assim, $J_4\left(x_4 = i\right) = g_4\left(x_4 = i\right) = 0$.

Agora decremente k.

$k = N - 1 = 3$

O espaço de estados é $S_3 = \{f, h\}$.

Se $x_3 = f$ o custo ótimo é

$$\begin{aligned} J_3\left(x_3 = f\right) =& min_{u_3 \in U_3(x_3 = f)} \{g_3\left(x_3 = f, u_3\right) + J_4\left(f_3\left(x_3 = f, u_3\right)\right)\} \\ =& g_3\left(x_3 = f, u_3 = -1\right) = 4 \end{aligned}$$

e a lei de controle ótimo é

$$\mu_3^*\left(x_3 = f\right) = arg\{min_{u_3 \in U_3(x_3 = f)} \{g_3\left(x_3 = f, u_3\right) + J_4\left(f_3\left(x_3 = f, u_3\right)\right)\}\} = -1$$

Se $x_3 = h$ o custo ótimo é

$$\begin{aligned} J_3\left(x_3 = h\right) =& min_{u_3 \in U_3(x_3 = h)} \{g_3\left(x_3 = h, u_3\right) + J_4\left(f_3\left(x_3 = h, u_3\right)\right)\} \\ =& g_3\left(x_3 = h, u_3 = -1\right) = 4 \end{aligned}$$

e a lei de controle ótimo é $\mu_3^*\left(x_3 = h\right) = 1$.
Assim, a função $J_3 : S_3 \to \mathbb{R}_+$ e a lei de controle admissível $\mu_3^* : S_3 \to C_3$ são dados pela tabela

x_3	$J_3\left(x_3\right)$	$\mu_3^*\left(x_3\right)$
f	4	-1
h	2	$+1$

Agora decremente k.

$k = 2$

O espaço de estados é $S_2 = \{c, e, g\}$.

Se $x_2 = c$ o custo ótimo é

$$\begin{aligned} J_2\left(x_2 = c\right) =& min_{u_2 \in U_2(x_2 = c)} \{g_2\left(x_2 = c, u_2\right) + J_3\left(f_2\left(x_2 = c, u_2\right)\right)\} \\ =& g_2\left(x_2 = c, u_2 = -1\right) + J_3\left(f\right) = 3 + 4 = 7 \end{aligned}$$

e a lei de controle ótimo é $\mu_2^*\left(x_2 = c\right) = -1$.

Se $x_2 = e$ o custo ótimo é

$$
\begin{aligned}
J_2(x_2=e) =& min_{u_2\in U_2(x_2=e)}\{g_2(x_2=e,u_2)+J_3(f_2(x_2=e,u_2))\}\\
=& min\{g_2(x_2=e,u_2=1)+J_3(f_2(x_2=e,u_2=1)),\\
& \quad g_2(x_2=e,u_2=-1)+J_3(f_2(x_2=e,u_2=-1))\}\\
=& min\{g_2(x_2=e,u_2=1)+J_3(f),g_2(x_2=e,u_2=-1)+J_3(h)\}\\
=& min\{3+4,2+2\}\ =4
\end{aligned}
$$

e a lei de controle ótimo é $\mu_2^*(x_2=e)=-1$.

Se $x_2=g$ o custo ótimo é

$$
\begin{aligned}
J_2(x_2=g) =& min_{u_2\in U_2(x_2=g)}\{g_2(x_2=g,u_2)+J_3(f_2(x_2=g,u_2))\}\\
=& g_2(x_2=g,u_2=1)+J_3(h)=4+2=6
\end{aligned}
$$

e a lei de controle ótimo é $\mu_2^*(x_2=g)=+1$.

Assim, a função $J_2:S_2\rightarrow\mathbb{R}_+$ e a lei de controle admissível $\mu_2^*:S_2\rightarrow C_2$ são dados pela tabela

x_2	$J_2(x_2)$	$\mu_2^*(x_2)$
c	7	-1
e	4	-1
g	6	$+1$

Agora decremente k.

$k=1$

O espaço de estados é $S_1=\{b,d\}$.

Se $x_1=b$ o custo ótimo é

$$
\begin{aligned}
J_1(x_1=b) =& min_{u_1\in U_1(x_1=b)}\{g_1(x_1=b,u_1)+J_2(f_1(x_1=b,u_1))\}\\
=& min\{g_1(x_1=b,u_1=1)+J_2(f_1(x_1=b,u_1=1)),\\
& \quad g_1(x_1=b,u_1=-1)+J_2(f_1(x_1=b,u_1=-1))\}\\
=& min\{g_1(x_1=b,u_1=1)+J_2(c),g_1(x_1=b,u_1=-1)+J_2(e)\}\\
=& min\{2+7,1+4\}\ =5
\end{aligned}
$$

e a lei de controle ótimo é $\mu_1^*(x_1=b)=-1$.

Se $x_1=d$ o custo ótimo é

$$
\begin{aligned}
J_1(x_1=d) =& min_{u_1\in U_1(x_1=d)}\{g_1(x_1=d,u_1)+J_2(f_1(x_1=d,u_1))\}\\
=& min\{g_1(x_1=d,u_1=1)+J_2(e),g_1(x_1=d,u_1=-1)+J_2(g)\}\\
=& min\{3+4,2+6\}\ =7
\end{aligned}
$$

e a lei de controle ótimo é $\mu_1^*(x_1=d)=+1$.

Assim, a função $J_1:S_1\rightarrow\mathbb{R}_+$ e a lei de controle admissível $\mu_1^*:S_1\rightarrow C_1$ são dados pela tabela

x_1	$J_1(x_1)$	$\mu_1^*(x_1)$
b	5	-1
d	7	$+1$

Agora decremente k.

$k = 0$

O espaço de estados é $S_0 = \{a\}$.

O custo ótimo é

$$\begin{aligned} J_0(x_0 = a) =& min_{u_0 \in U_0(x_0=a)} \{g_0(x_0 = a, u_0) + J_1(f_0(x_0 = a, u_0))\} \\ =& min\{g_0(x_0 = a, u_0 = 1) + J_1(b), g_0(x_0 = a, u_0 = -1) + J_1(d)\} \\ =& min\{3 + 5, 1 + 7\} = 8 \end{aligned}$$

e a lei de controle ótimo não é única e podemos fazer $\mu_0^{*1}(x_0 = a) = +1$ ou $\mu_0^{*2}(x_0 = a) = -1$.

Assim, pelo algoritmo da programação dinâmica chegamos à solução ótima do problema original

$$J^*(x_0 = a) = J_0(x_0 = a) = 8$$

e temos duas políticas (leis de realimentação) ótimas

$$\pi^{*i} = \{\mu_0^{*i}, \mu_1^*, \mu_2^*, \mu_3^*\}$$

Em particular, percorrendo a partir de $x_0 = a$ pela política ótima obtemos um sinal de controle ótimo em malha aberta e a correspondente trajetória ótima

$$u^* = \left(u_0^*, ..., u_{N-1}^*\right) = (u_0^*, u_1^*, u_2^*, u_3^*) = (+1, -1, -1, +1)$$

$$x^* = (a, b, e, h, i)$$

Para o outro sinal de controle ótimo em malha aberta e a correspondente trajetória ótima

$$u^* = (u_0^*, u_1^*, u_2^*, u_3^*) = (-1, +1, -1, +1)$$

$$x^* = (a, d, e, h, i)$$