UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

DEPARTAMENTO DE TELEMÁTICA

Programação Dinâmica Estocástica

IA-601 Sistemas Dinâmicos Estocásticos

João Bosco R. do Val

9 de outubro de 2001

arquivo Dinamic_Prog.tex

Custo de Operação de Cadeia de Markov

$$J = E[\sum_{k=0}^{N} c(x(k))]$$

1a. forma de expressar J

$$J = \sum_{k=0}^{N} \sum_{i \in X} \mathrm{Prob}(x(k) = i)\, c(i) =$$

$$\sum_{k=0}^{N} p(k).c = \sum_{k=0}^{N} p(0).P^k.c$$

podemos ainda considerar custo $c(k, i)$

Solução Recursiva:

$$V_i(k) = E[\sum_{l=k}^{N} c_{x(l)}(l) | x(k) = i]$$

(custo remanescente se partimos do estado i no instante k)

$$J = \sum_{i} V_i(0) p_i(0)$$

Fatos: $V_i(N) = c_i(N)$ e $V_i(k) = c_i(k) + \sum_{j \in X} p_{ij} V_j(k+1)$

Solução:

Via equação recusiva retrógrada

Imersão do problema numa classe maior de problemas

Problemas com horizonte infinito ($N = +\infty$)

$$E[\sum_{k=0}^{\infty} c(k, x(k))] \text{ (custo total não descontado)}$$

$$c < 0 \text{ (caso negativo)} \quad c > 0 \text{ (caso positivo)}$$

pode não existir!

$$E[\sum_{k=0}^{\infty} \beta^k c(k, x(k))],\ 0 < \beta < 1 \text{ (custo descontado)}$$

$$\lim \sup_{N\to\infty} \frac{1}{N} E[\sum_{k=0}^{N-1} c(x(k))] \text{ (critério de custo médio a longo prazo)}$$

Expressando equivalentemente,

$$E[\sum_{k=0}^{\infty} c(x(k))] = \sum_{k=0}^{\infty} p(0).P^k.c(k)$$

$$E[\sum_{k=0}^{\infty} \beta^k c(x(k))] = \sum_{k=0}^{\infty} \beta^k p(0).P^k.c(k)$$

$$\lim \sup_{N\to\infty} \frac{1}{N} E[\sum_{k=0}^{N-1} c(x(k))] = \lim \sup_{N\to\infty} \frac{1}{N} \sum_{k=0}^{\infty} p(0)P^k.c(k)$$

Caso especial $c(k) = c$

$$\lim \sup_{N\to\infty} \frac{1}{N} \sum_{k=0}^{\infty} p(0)P^k.c(k) = \pi.c \text{ (ergódico)}$$

Caso Custo Descontado

c não depende do estágio k, $0 < \beta < 1$. Definindo,

$$V_i(k) = E[\sum_{l=k}^{\infty} \beta^l c(x(l))|x(k) = i]$$

$$= E[\beta^k \sum_{l=0}^{\infty} \beta^l c(x(l))|x(0) = i] = \beta^k V_i(0)$$

Usando essa invariância ao deslocamento,

$$V_i(k) = \beta^k c_i + \sum_{j\in X} p_{ij} V_j(k+1)$$

$$\beta^k V_i(0) = \beta^k c_i + \sum_{j\in X} p_{ij} \beta^{k+1} V_j(0)$$

Em forma matricial: $V(0) = c + \sum_{j\in X} p_{ij} \beta V_j(0)$ e rank de $(I - \beta P)$ é completo:

$$V(0) = (I - \beta P)^{-1} c$$

Como $(I - \beta P)(I + \beta P + \beta^2 P^2 + \cdots) = I$

$$V(0) = (I - \beta P)^{-1} c = \sum_{k=0}^{\infty} \beta^k P^k.c$$

Cadeias de Markov Controlada

$P = [p_{ij}(u)]$ Custo Horizonte Finito ($= N$):

$$E[\sum_{k=0}^{N} c(x(k), u(k))]$$

como $u(k),\ k = 0, \ldots, N$ é escolhido?

Malha aberta: $u(k) = u_k,\ \forall k$

Malha fechada: $u(k)$ depende de $x(k), x(k-1), \ldots,$ e de $u(k-1), u(k-2), \ldots$

Em malha fechada o controle deve depender da história $u(t) = g_t(x^t, u^{t-1})$ e uma política ou estratégia de controle é a sequência de funções

$$g = (g_0, g_1, g_2, \ldots)$$

Política markoviana: $u(t) = g_t(x(t))$ onde $g_t : X \to U$

Fato: Sob uma política markoviana $g = (g_1, g_2, \ldots)$, $\{x(t)\}$ é uma cadeia de Markov com probabilidade de transição não-homogêna (variante no tempo):

$$P(x(t+1) = j | x(t) = i) = p_{ij}(g_t(i), t)$$

Política markoviana estacionária: $u(t) = g_0(x(t))$, isto é todas as g_t são iguais. Neste caso a CM se mantém homogênea.

Seja

$G = (g_0, g_1, \ldots)$ a classe de políticas gerais

G^M a classe de políticas markoviana

G^s a classe de políticas estacionárias

$G^s \subset G^M \subset G$

Considere política markoviana.

$$J^g = E^g[\sum_{k=0}^{N} c(x(k), u(k))]$$

$$V_i^g(t) = E^g[\sum_{k=t}^{N} c(x(k), u(k))|x(k) = i]$$

Recursão válida para políticas markovianas, $g \in G^M \subset G$:

$$V_i^g(N) = c(i, g_N(i), N)$$
$$V_i^g(t) = c(i, g_t(i), t) + \sum_{j \in X} p_{ij}(g_t(i), t) V_j^g(t+1)$$

Para políticas gerais, $g \in G$:

$$J^g(t) = E^g[\sum_{k=t}^{N} c(x(k), u(k))|x^t, u^{t-1}] \text{ (é uma v.a.)}$$

PRINCÍPIO DE COMPARAÇÃO: Seja $V_i(t)$, $i \in X$, $0 \leq t \leq N$ satisfazendo:

$$\begin{aligned} &V_i(N) \leq c(i, N) \\ &V_i(t) \leq c(i, u, t) + \sum_{j \in X} p_{ij}(u, t) V_j(t+1) \text{ para todo } u \in U \end{aligned}$$

Então $J^g(t) \geq V_{x(t)}(t)$ (q.c.) para todo g e t

Exemplo

Observações:

1) Imersão de um problema mais simples numa classe maior de problemas. Obtem-se a solução para a classe maior de problemas.

2) Programação Dinâmica pode ser computacionalmente intratável se os número de estágios e as decisões possíveis forem muito grandes.

3) PD determina a decisão a decisão ótima como uma função do estado e do número de estágios remanescentes. Com isso a política ótima em malha fechada é obtida.

4) DP é um procedimento recursivo retrógrado no tempo. Após determinar a solução no instante $t+1$, ele determina a solução ótima no instante t.

5) Custo Ótimo do estado x até o final =

$$= \text{Min}\begin{bmatrix} \text{Distância percorrida no} & & \text{Distância ótima de } x' \text{ até o fim} \\ \text{estágio presente quando} & + & \text{onde } x' \text{ é o estado para o qual} \\ \text{a decisão } u_i \text{ é tomada} & & \text{a decisão } u_i \text{ o conduz} \end{bmatrix}$$

$$\begin{array}{cc} |------------ & |--------------| \\ \text{custo imediato} & \text{custo ótimo remanescente} \end{array}$$

Prova do Princípio de Comparação:

Lema Suponha que $g_t^*(i)$ atinja o mínimo em

$$\min_{u \in U}\{c(i,u,t) + \sum_{j \in X} p_{ij}(u,t)V_j(t+1)\}$$

Seja $g^* = (g_0^*, g_1^*, \ldots) \in G^M$, então $V_i^{g^*}(t) = V_i(t)$.

Teorema $J_0^{g^*} \leq J_0^g$ para todo $g \in G$. Então a política markoviana g^* é ótima.

Princípio de Otimalidade: Segmentos da política ótima são em si ótimos.

Problemas com Horizonte Infinito

$P = [p_{ij}(u)]$ invariante no tempo, $c(i,u)$, $0 < \beta < 1$

$$\min_{g \in G} E^g[\sum_{t=0}^{\infty} \beta^t c(x(t), u(t)|x(0) = i]$$

Como resolver? Suponha que se tenha

$$V_i^{N-1}(0) = \min_{g \in G} E^g[\sum_{t=0}^{N-1} \beta^t c(x(t), u(t)|x(0) = i]$$

Seja $W_i(N) = V_i^{N-1}(0) =$ custo ótimo de operação do sistema por N dias, iniciando no estado i. Então

$$W_i(N) = \min_{u \in U}\{c(i,u) + \sum_{j \in X} p_{ij}(u)\beta W_j(N-1)\}$$

e estamos interessados em $W_i(\infty)$. Espera-se que

$$W_i(\infty) = \min_{u \in U}\{c(i,u) + \sum_{j \in X} p_{ij}(u)\beta W_j(\infty)\}$$

e também que o controle ótimo seja na forma $u(t) = g(x(t))$ (ótimo estacionário).

Questões:

1. Existe função W que satisfaz a equação acima?

2. A equação tem solução única?

3. Como podemos determinar a solução?

4. A solução $W_i(\infty)$ é igual ao $\min_{g\in G} E^g[\sum_{t=0}^{\infty}\beta^t c(x(t),u(t)|x(0) = i]$?

5. A solução ótima será estacionária? E será ótimo a estratégia

$$g(i) = \arg\min\{c(i,u) + \sum_{j\in X} p_{ij}(u)\beta W_j(\infty)\}?$$

Teoria do Mapeamento Contrativo

Considere um operador $T : S \to S$ tal que

$$\|Tx - Ty\| \leq \beta\|x-y\|,\ 0<\beta<1 \text{ e todo } x \text{ e } y \in S$$

Se S for fechado, espera-se que exista x^* tal que $Tx^* = x^*$.

TEOREMA Seja S um espaço métrico completo. Então

1) Existe um $x^* \in S$ com $Tx^* = x^*$, isto é, x^* é um ponto fixo,

2) Existe apenas um ponto fixo,

3) Para todo x, $T^n x \to x^*$

Prova: Tome $z \in S$ fixo

$$\|T^{n+1}z - T^n z\| = \|T(T^n z) - T(T^{n-1}z)\| \leq$$
$$\beta\|T^n z - T^{n-1}z\| \leq \cdots \leq \beta^n\|Tz - z\|$$

Portanto $\{z, Tz, T^2z, \ldots\}$ é uma sequência de Cauchy, isto é existe um N tal que

$$\|T^{n+N}z - T^{m+N}z\| \leq \epsilon$$

para todo $n \geq 0,\ m \geq 0$. Assim, $T^n z \to x^*$, e como T é contínuo:

$$Tx^* = T(\lim_{n\to\infty} T^n z) =$$
$$\lim_{n\to\infty} T(T^n z) = \lim_{n\to\infty} T^{n+1}z = x^*$$

Então x^* é um ponto fixo, e como existe no máximo um único ponto fixo, $T^n z \to x^*$ para todo z. $\square$

De volta ao problema, queremos resolver

$$W_i(\infty) = \min_{u\in U}\{c(i,u) + \sum_{j\in X} p_{ij}(u)\beta W_j(\infty)\}$$

para cada i. Na forma vetorial

$$W = \min_{u\in U}\{c_u + \beta P(u)W\} := TW$$

com u tomado independentemente para cada elemento do vetor. Define-se

$$T : \mathbb{R}^\ell \to \mathbb{R}^\ell \text{ com } (TW)_i = \min_{u\in U}\{c(i,u) + \beta \sum_{j\in X} p_{ij}(u)W_j\}.$$

Queremos determinar W^* tal que $W^* = TW^*$.

Defina a norma $\|x\| = \max_i |x_i|$.

FATO: T é um mapeamento contrativo.

Prova: Seja W e V. É preciso mostrar que $\|TW - TV\| \leq \beta\|W - V\|$, isto é,

$$\max_i |(TW)_i - (TV)_i| \leq \beta \max_i |W_i - V_i|$$

Mas,

$$\begin{aligned}
&(TW)_i - (TV)_i = \\
&\min_{u \in U}\{c(i,u) + \beta \sum_{j \in X} p_{ij}(u) W_j\} - \\
&\min_{u \in U}\{c(i,u) + \beta \sum_{j \in X} p_{ij}(u) V_j\} = \\
&\min_{u \in U}\{c(i,u) + \beta \sum_{j \in X} p_{ij}(u) W_j\} - \\
&- c(i,u^*) - \beta \sum_{j \in X} p_{ij}(u^*) V_j \leq \\
&c(i,u^*) + \beta \sum_{j \in X} p_{ij}(u^*) W_j - c(i,u^*) - \beta \sum_{j \in X} p_{ij}(u^*) V_j = \\
&\beta \sum_{j \in X} p_{ij}(u^*)(W_j - V_j) \leq \beta \max_j |W_j - V_j|
\end{aligned}$$

Portanto, $\|TW - TV\| \leq \beta\|W - V\|$ $\square$

TEOREMA:

1. Existe um único W satisfazendo

$$W_i = \min_{u \in U} \{c(i,u) + \sum_{j \in X} p_{ij}(u) \beta W_j\}$$

2. Escolha V qualquer. Então $\lim_{n \to \infty} T^n V = W =$ único ponto fixo.

TEOREMA:

Se $W = TW$ então

$$W_i = \min_{g \in G} E^g[\sum_{t=0}^{\infty} \beta^t c(x(t), u(t)) | x(0) = i]$$

Para cada i, suponha que $g(i)$ atinja o mínimo em o

$$W_i = \min_{u \in U} \{c(i,u) + \sum_{j \in X} p_{ij}(u) \beta W_j\}$$

Então g é ótimo e estacionário.

Prova: Assuma que $0 \leq c(i,u) \leq M$ sem perda de generalidade.

$$\min_{g \in G} E^g[\sum_{t=0}^{\infty} \beta^t c(x(t), u(t)) | x(0) = i] \geq$$

$$\min_{g \in G} E^g[\sum_{t=0}^{n-1} \beta^t c(x(t), u(t)) | x(0) = i] =$$

$$= (T^n 0)_i \to W_i$$

Para a desigualdade reversa,

$$
\begin{aligned}
&\min_{g\in G} E^g[\sum_{t=0}^{\infty}\beta^t c(x(t),u(t))|x(0)=i] \le \\
&\min_{g\in G} E^g[\sum_{t=0}^{n-1}\beta^t c(x(t),u(t)) + \sum_{t=n}^{\infty}\beta^t M|x(0)=i] = \\
&\min_{g\in G} E^g[\sum_{t=0}^{n-1}\beta^t c(x(t),u(t))|x(0)=i] + \frac{\beta^n M}{1-\beta} = \\
&= (T^n 0)_i + \frac{\beta^n M}{1-\beta} \to W_i
\end{aligned}
$$

O que prova a 1a. parte. Para mostrar que g que atinge o mínimo, lembre que para qualquer política $h \in G^s$ existe Z tal que

$$Z_i = c(i,h(i)) + \beta\sum_{j\in X} p_{ij}(h(i))Z_j$$

e

$$
\begin{aligned}
Z_i = E^h[\sum_{t=0}^{\infty}\beta^t c(x(t),u(t))|x(0)=i] \ge \\
E^g[\sum_{t=0}^{\infty}\beta^t c(x(t),u(t))|x(0)=i] = W_i
\end{aligned}
$$

Princípios Básicos de Solução:

1) Aproximações sucessivas / iteração de ponto fixo / iteração de valor:

a) Escolha $f_0 = \begin{pmatrix} f_0(1) \\ \vdots \\ f_0(J) \end{pmatrix}$

b) Resolva iterativamente

$$f_{n+1}(i) = \min_{u \in U}\{c(i,u) + \beta \sum_{j \in X} p_{ij}(u) f_n(j)\}$$

então $\lim_{n\to\infty} f_n(i) = W_i$

FORMA ALGORÍTMICA DA ITERAÇÃO DE VALOR

1. Selecione $V^0 : E \to \mathbb{R}$, especifique $\epsilon > 0$; $n = 0$;
2. Para cada $i \in E$ calcule

$$W^{n+1}(i) = \max_{u \in U_i}\{c(i,u) + \sum_{j \in E} \beta p(j|i,u) W^n(j)\}$$

3. Se $\|W^{n+1} - W^n\| < \epsilon(1-\beta)/2\beta$ vá para o passo 4.; caso contrário, $n = n+1$ e retorne ao passo 2.;

4. Para cada $i \in E$ escolha

$$u_\epsilon(i) = \arg\max_{u \in U_i} \{c(i,u) + \sum_{j \in E} \beta p(j|i,u) W^{n+1}(j)\}$$

e pare.

Resultado: $\|W^{n+1} - W\| < \epsilon/2$ (Puterman, sec. 6.3.2)

ACELERAÇÃO: ANÁLOGO AO GAUSS-SIEDEL

Considere estados ordenados:

$$i_1, i_2, \ldots, i_J$$

2' Tome $j = 1$ e calcule

$$W^{n+1}(i_j) = \max_{u \in U_{i_j}} \{c(i_j,u) + \beta[\sum_{\ell < j} p(i_\ell|i_j,u) W^{n+1}(i_\ell) + \sum_{\ell \geq j} p(i_\ell|i_j,u) W^{n}(i_\ell)]\}$$

Se $j = J$, vá para 3.; caso contrário faça $j = j + 1$ e retorne ao cálculo de $W^{n+1}(i_j)$ acima;

2) Iteração de política

Política estacionária $g : X \mapsto U$ com $|X| = J$ e $|U|$ finito. Número de possíveis decisões $|U|^J$

Como melhorar uma dada política estacionária? $V^g = T_g V^g$
Ordenação:

$$W \leq V \Leftrightarrow W(j) \leq V(j), \quad j = 1, \ldots, J$$

$$W < V \Leftrightarrow \begin{cases} W(j_0) < V(j_0) \\ W(j) \leq V(j) \text{ para outros } j \end{cases}$$

Suponha que $W \leq V \Rightarrow T_g W \leq T_g V$ (T_g é operador monotônico).
Segue que $TW \leq TV$ (T também é monotônico)

Procedimento de iteração de política
a) Iniciar com g_0; calcular V^{g_0}. Duas possibilidades:

$$\begin{cases} V^{g_0} = TV^{g_0} \\ V^{g_0} > TV^{g_0} \end{cases}$$

b) $V^{g_0} > TV^{g_0}$. Obtenha g_1 com

$$(TV^{g_0})_i = \min_{u \in U} \{c(i, u) + \beta \sum_{j \in X} p_{ij}(u) V^g(j)\}$$
$$= c(i, g_1(i)) + \beta \sum_{j \in X} p_{ij}(g_1(i)) V^g(j)$$

Com isso $V^{g_0} > T_{g_1}V^{g_0} \Rightarrow T_{g_1}V^{g_0} \geq (T_{g_1})^2 V^{g_0}$ e

$$V^{g_0} > T_{g_1}V^{g_0} \geq (T_{g_1})^2 V^{g_0} \geq \cdots \geq (T_{g_1})^n V^{g_0} \rightarrow V^{g_1}$$

FORMA ALGORÍTMICA DE ITERAÇÃO DE POLÍTICA
(Puterman sec. 6.4.1)

1. Selecione uma política $g_0 \in G^s$ e faça $n = 0$;
2. Avalie o custo da política g_n resolvendo
$$(I - \beta P_{g_n})W^n = c_{g_n}$$
3. Melhoria de política: escolha u_{n+1} satisfazendo
$$u_{n+1} \in \arg\min_{u \in U}(c_u + \beta P_u W^n)$$
tomando $u_{n+1}(i) = u_n(i)$ sempre que possível;
4. Se $u_{n+1}(i) = u_n(i)$ para todo $i \in E$ pare e $g^* = u^n$ é ótimo. Caso contrário faça $n = n + 1$ e retorne ao passo 2.

3) Solução via Programação Linear (Puterman sec. 6.9)

Suponha que o conjunto de ações admissíveis U_i para todo i seja finito. ($|U_i|$ finito).
Sabemos que se

$$V \geq c_u + \beta P_u V, \quad \forall u \in U$$

então $V \geq W$.

Idéia básica: minimizar V satisfazendo a equação acima:

$$\text{minimizar} \sum_{i \in E} \alpha(i) V(i) \qquad (\alpha(i) > 0)$$

sujeito à

$$V(i) - \beta \sum_{j \in E} p(j|i,u) V(j) \geq c(i,u) \qquad \text{(problema primal)}$$

$$u \in U_i, \, \forall i \in E$$

Obtendo o dual:

$$\sum_{i\in E}\{\alpha(i)V(i)+x(i,u)[V(i)-\beta\sum_{j\in E}p(j|i,u)V(j)-c(i,u)]\}$$
$$=\sum_{i\in E}\{-x(i,u)c(i,u)+V(i)[x(i,u)-\beta\sum_{j\in E}p(i|j,u)x(j,u)+\alpha(i)]\}$$

Problema PL dual:

$$\text{maximizar}\ \sum_{i\in E}\sum_{u\in U_i}x(i,u)c(i,u)$$

sujeito à

$$\sum_{u\in U_i}x(i,u)-\beta\sum_{j\in E}\sum_{u\in U_j}p(i|j,u)x(j,u)+\alpha(i)=0,\ \forall i\in E$$

Custo Médio de Longo Prazo

Custo médio de uma política estacionária. Suponha que P_g seja irredutível

$$V^g = \lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N-1} (P_g)^t \, c_g = \begin{bmatrix} \pi_g.c_g \\ \vdots \\ \pi_g.c_g \end{bmatrix}$$

Outra forma: Considere o custo associado a N dias W_N. Queremos estudar

$$\lim_{N\to\infty} \frac{W_N(\cdot)}{N}$$

e espera-se que $\lim_{N\to\infty} W_N(i)/N = J^*$ não dependa de i, isto é

$$W_N(i) = NJ^* + O(N)$$

Assumimos que $O(N) \to w(i)$, mais precisamente,

$$\lim_{N\to\infty} (W_N(i) - NJ^*) = w(i)$$

para cada i.

Equação da Programação Dinâmica

Como

$$W_N(i) = \min_{u \in U}\{c_{iu} + \sum_{j \in X} p_{ij}(u) W_{N-1}(j)\}$$

então,

$$W_N(i) - NJ^* = \min_{u \in U}\{c_{iu} + \sum_{j \in X} p_{ij}(u)\Big(W_{N-1}(j) - NJ^*\Big)\}$$

$$= \min_{u \in U}\{c_{iu} + \sum_{j \in X} p_{ij}(u)\Big(W_{N-1}(j) - (N-1)J^*\Big)\} - J^*$$

tomando o limite

$$w(i) + J^* = \min_{u \in U}\{c_{iu} + \sum_{j \in X} p_{ij}(u) w(j)\}$$

(Eq. da PD)

FATO: Se w e J^* satisfazem a equação de PD acima então J^* é o custo ótimo independente de i. Se $g^*(i)$ atinge o mínimo na equação de PD, então g^* é ótimo.

Prova.

QUESTÃO: Quando existe w e J^* satisfazendo a equação de DP?

TEOREMA: Suponha que para toda política estacionária g, P_g seja irredutível. Então existe solução para a equação de PD.

Prova: Ponto básico. Para cada g a solução da equação $v = vP_g$ é na forma $v = k\pi_g$.
Re-escrevendo:

$$(P_g' - I)\pi_g' = 0$$

similar a $Ax = 0 \Leftrightarrow x \in \mathcal{N}(A)$. Portanto,

$$\pi_g' \in \mathcal{N}(P_g' - I) = \{k\pi_g' \text{ para } k \text{ real}\}$$

Notação:

$$\mathcal{N}(A) = \{x : Ax = 0\}, \qquad \mathcal{R}(A) = \{Ax : \text{para todo } x\}$$

são subespaços. Se $\mathcal{S}$ é subespaço,

$$\mathcal{S}^\perp = \{x : x \perp s \text{ para todo } s \in \mathcal{S}\}.$$

FATO: $\mathcal{N}(A) = \mathcal{R}(A')^\perp$
Pois, $x \in \mathcal{N}(A) \Rightarrow Ax = 0 \Rightarrow y'Ax = 0 \ \forall y \Rightarrow x'A'y = 0 \ \forall y \Rightarrow$ $< x, A'y >= 0 \ \forall y \Rightarrow x \perp A'y \ \forall y \Rightarrow x \perp \mathcal{R}(A') \Rightarrow x \in \mathcal{R}(A')^\perp$
Considere $J^g = \pi_g \, c_g$ custo de g. Definindo $e = (1 \ \ 1 \cdots 1)$ podemos re-escrever

$$\pi_g(J^g e - c_g) = 0$$

portanto,

$$J^g e - c_g \perp \pi'_g \equiv \mathcal{N}(P'_g - I)$$

e então,

$$J^g e - c_g \in \mathcal{R}(P_g - I)$$

e existe um vetor w_g tal que

$$J^g e - c_g = (P_g - I) w_g,$$

ou

$$w_g(i) + J^g = c_{i,g} + \sum_{j \in X} p_{ij}(g(i)) w_g(j)$$

Considere agora g^* a melhor política entre $g \in G^s$, isto é

$$J^{g^*} \leq J^g \, \forall g \in G^s.$$

Então existe w_g^* e J^{g^*} tal que

$$\begin{aligned} w_{g^*}(i) + J^{g^*} &= c_{i,g*} + \sum_{j \in X} p_{ij}(g^*(i)) w_{g^*}(j) \\ &\geq \min_{u \in U} \{c_{i,u} + \sum_{j \in X} p_{ij}(u) w_{g^*}(j)\} \\ &= c_{i,\widehat{g}} + \sum_{j \in X} p_{ij}(\widehat{g}) w_{g^*}(j) \end{aligned}$$

em forma vetorial: $w_{g^*} + J^{g^*} e \geq c_{\widehat{g}} + P_{\widehat{g}} w_{g^*}$ o que leva a concluir que

$$\pi_{\widehat{g}} w_g + J^{g^*} \pi_{\widehat{g}} e \geq \pi_{\widehat{g}} c_{\widehat{g}} + \pi_{\widehat{g}} P_{\widehat{g}} w_{g^*}$$

$$\pi_{\widehat{g}} w_g + J^{g^*} \geq J^{\widehat{g}} + \pi_{\widehat{g}} w_{g^*}$$

a desigualdade estrita não pode ocorrer em nenhum estado i , pois $\pi_{\widehat{g}}(i) > 0$. Então

$$w_{g^*}(i) + J^{g^*} = \min_{u \in U}\{c_{iu} + \sum_{j \in X} p_{ij}(u) w_{g^*}(j)\}$$

e existe solução w e J^* para a equação de PD. Provamos que $g^* \in$G^s é ótima dentre todas as políticas $g \in G$.