Programação Dinâmica Determinística e Estocástica

Ruy Eduardo Campello

FURNAS-Centrais Elétricas S.A.
e Instituto Metodista Bennett

Prefácio

Estas notas foram preparadas como material didático do mini-curso, de 6 horas de duração, *Programação Dinâmica Determinística e Estocástica*, apresentado durante o *XXXIV Simpósio Brasileiro de Pesquisa Operacional da SOBRAPO* (8 a 11 de outubro, 2002), realizado no Instituto Militar de Engenharia – IME, no Rio de Janeiro. Foram resumidas e adaptadas do curso completo, com 54 horas, de *Modelagem Matemática e Programação Dinâmica* ministrado pelo autor no *Programa de Ciência da Computação do Instituto Metodista Bennett - IMB*.

Rio de Janeiro, 8 de outubro de 2002

Ruy Eduardo Campello

e-mail: campello@iis.com.br
campello@furnas.com.br

Índice

1. Programação Dinâmica

1.1. Introdução

A *Programação Dinâmica*, conhecida também como *otimização recursiva*, é um procedimento de otimização para resolver problemas de *decisão seqüencial* ou de *múltipos-estágios* relacionados. Entretanto, esta abordagem pode ser utilizada *induzindo* a propriedade seqüencial por conveniência computacional.

Sua essência é o *Princípio da Otimalidade de Richard Bellman*. Ao contrário de outros ramos da Programação Matemática, não pode ser definido um único algoritmo capaz de resolver diretamente todos os problemas de programação dinâmica. A multiplicidade de situações modeláveis pela técnica requer teoria e arte utilizando diferentes funções na formulação da *equação de otimalidade,* embora o princípio utilizado seja sempre o de Bellman.

A técnica da programação dinâmica permite transformar um problema de decisão seqüencial (em múltiplos estágios) contendo diversas variáveis interdependentes em uma série de subproblemas contendo poucas variáveis. A transformação é *invariante* preservando o número de soluções viáveis o valor da função objetivo associado a cada uma delas e, portanto, a própria solução ótima. De uma forma geral um problema de otimização com *n* variáveis de decisão é transformado em *n* subproblemas cada um deles com apenas uma varíavel de decisão (no caso unidimensional). O esforço computacional cresce exponencialmente com o número de variáveis, porém, apenas linearmente com o número de subproblemas. Assim, podem ser

obtidas reduções significativas no esforço computacional quando comparado a outras técnicas de otimização.

Em resumo, a Programação Dinâmica é uma técnica que se aplica à situações que exijam decisões sequenciais. Resolve problemas pela sua decomposição em sub-problemas resolvidos estágio por estágio oferecendo algumas vantagens em relação a outras técnicas de otimização. Pode tratar funções descontínuas, não diferenciáveis, não convexas, determinísticas ou estocásticas. A função objetivo deve, entretanto, ser *separável* e *monotônica*.

Como exemplo de um processo com múltiplos estágios por natureza considere uma situação simplificada de planejamento da produção de um único item durante t períodos, tal que:

y_j ... estoque no final do período j
u_j ... decisão do nível de produção no período j
w_j ... demanda conhecida pelo item no período j

A posição do estoque no início do primeiro período y_0 é conhecida. Portanto, em qualquer período (estágio) j a posição inicial do estoque y_{j-1} mais o nível de produção u_j menos a demanda w_j , considerada determinística neste caso, definem a posição do estoque (estado) no período seguinte. Ou seja, a evolução do processo pode ser representada por uma *transformação* (*função de transição de estado*) da forma:

$$y_j = r(y_{j-1}, u_j, w_j) = y_{j-1} + u_j - w_j \,, \quad j = 1, 2, ..., t$$

$$y_0 = \bar{y} \text{ (condição de contorno)}$$

Considera ainda os seguintes custos:

c_j

Em relação a modelagem da programação dinâmica cada estado deste processo é descrito completamente pelo nível do estoque no início de cada período, ou seja, para que a decisão seja tomada em cada estágio, e todos os subsequentes, é necessário conhecer apenas o nível do estoque no início do período. Esta característica torna o processo Markoviano sendo indispensável para que o Princípio da Otimalidade de Bellman possa ser aplicado.

"Um processo é Markoviano quando o futuro depender apenas da situação presente, ou seja, o passado não tem nenhuma influência nas decisões futuras."

Observe que sendo a demanda w_j , $j = 1, 2,..., t$ probabilística, teríamos um *problema de decisões sequenciais estocástico* em que o objetivo seria *minimizar o valor esperado do custo total do processo*, ou seja:

$$(P): \text{minimizar } E\{\sum_{j=1}^{t} g_j(y_{j-1} + u_j - w_j)\}$$

sujeito a:

$$y_j = y_{j-1} + u_j - w_j \,, \quad j = 1, 2, ..., t$$

$$y_0 = \bar{y}$$

$$u_j \geq 0 \,, \quad j = 1, 2, ..., t$$

O processo simplificado descrito anteriormente é claramente Markoviano, por outro lado, se estivéssemos lidando, por exemplo, com itens perecíveis para a tomada de decisões seria necessário conhecer quando cada item em estoque foi produzido. Logo, neste caso, o processo não seria mais

Markoviano. Entretanto, modificando a definição da variável de estado o processo pode ser transformado em Markoviano. Para tanto, basta definir o estado não como uma variável única representado o nível atual do estoque, mas uma matriz com duas colunas. A primeira coluna em cada linha representaria o período de produção de cada item em estoque e a segunda o número de itens produzidos neste período. Esta seria a maneira de transportar toda a informação do passado para o estado atual, permitindo então uma decisão segundo um processo Markoviano.

Como ilustração de um *processo sequencial por indução* considerar o seguinte problema de programação inteira:

$$
\begin{aligned}
&\textit{(P)}\text{: maximizar } x_0 = 8x_1 + 7x_2 \\
&\quad \text{sujeito a:} \\
&\qquad\qquad 2x_1 + \ x_2 \le 8 \\
&\qquad\qquad 5x_1 + 2x_2 \le 15 \\
&\qquad\qquad x_1, \quad x_2 \ge 0 \ \text{ e inteiros}
\end{aligned}
$$

Neste caso, o processo sequencial não resulta evidente em razão da natureza do problema de programação inteira, este porém, pode ser induzido. A decomposição em problemas menores pode ser caracterizada interpretando cada uma das variáveis x_1 e x_2 como sendo uma atividade e, a cada estágio, o nível de cada uma deve ser decidido. O termo independente de cada uma das restrições pode ser entendido como recurso disponível para realizar as atividades. A cada estágio, portanto, uma decisão deve ser tomada quanto ao nível de uma das atividades bem como os recursos a serem utilizados. As decisões são limitadas pelo nível dos recursos disponíveis no início de cada

estágio e, portanto, recursos remanescentes (não utilizados em cada estágio) podem ser alocados ao estágio seguinte. Assim, está caracterizado um processo sequencial Markoviano em que o conhecimento do nível de recursos disponíveis no início do estágio é suficiente para decidir de forma ótima neste estágio e, em conseqüência, em todos os estados subsequentes (neste caso apenas um).

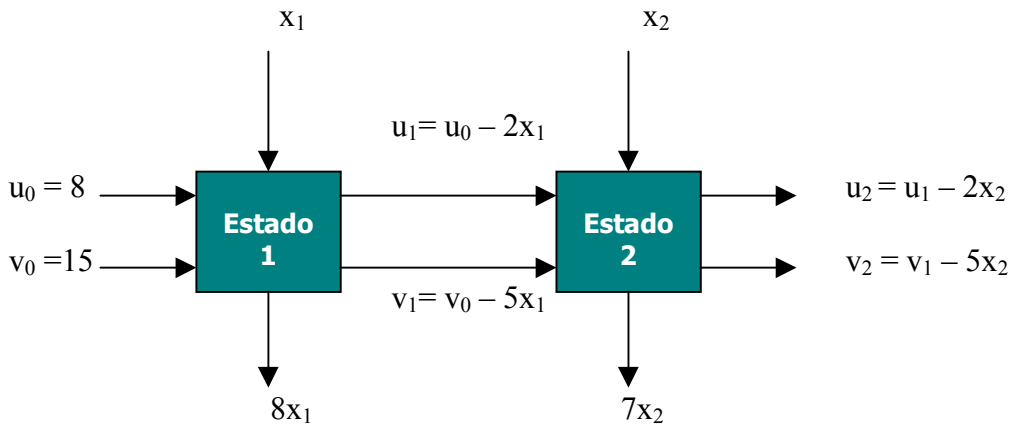
Como existem duas restrições o estado em cada estágio (t = 1, 2) fica definido por duas variáveis:

u_t .... primeiro recurso disponível (*restrição 1*) no estágio $t = 1, 2$

v_t .... segundo recurso disponível (*restrição 2*) no estágio $t = 1, 2$

$u_0 = 8$ e $v_0 = 15$ (condições de contorno)

Esquemático do processo sequencial induzido:

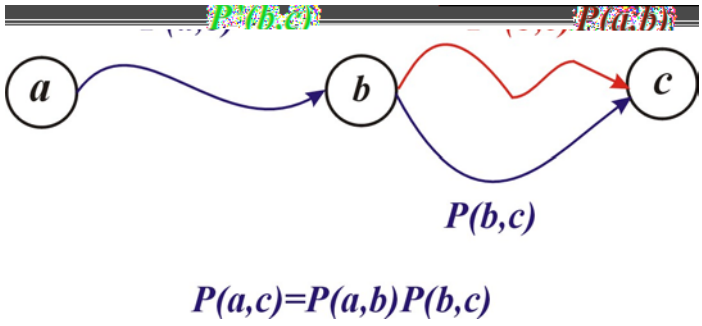


1.2. Princípio da Otimalidade de Bellman

O princípio de otimização devido a *Richard Bellman* (1957) é bastante intuitivo e será apresentado de maneira mais formal em 2.9.

> *"Uma trajetória ótima tem a seguinte propriedade: quaisquer que tenham sido os passos anteriores, a trajetória remanescente deverá ser uma trajetória ótima com respeito ao estado resultante dos passos anteriores, ou seja, uma política ótima é formada de subpolíticas ótimas."*

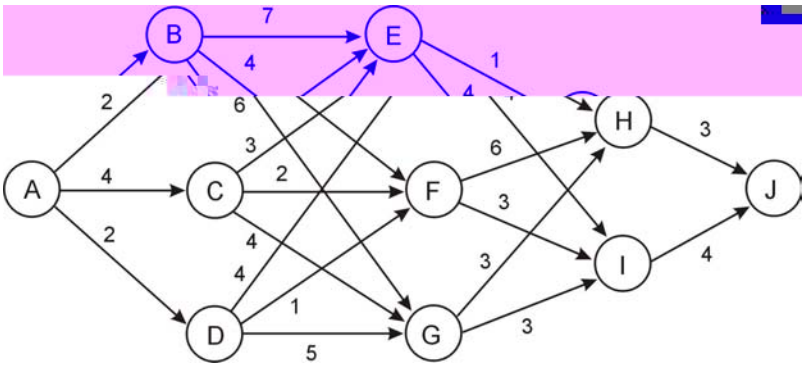
Informalmente, pode-se intuir o resultado pela argumentação a seguir. Digamos que *P(a,c)* seja uma trajetória ótima dos pontos *a* até *c* passando por um ponto intermediário qualquer *b*, como no esquema a seguir. Então, *P(a,c) = P(a,b)P(b,c)*.



Entretanto, por absurdo, assuma que outra trajetória, digamos *P'(b,c)*, seja a trajetória ótima de *b* até *c* e não *P(b,c)*. Se isto ocorre, a trajetória *P(a,b)P'(b,c)* deve ser melhor que *P(a,c)*. Entretanto, isto contraria a hipótese original de que *P(a,c)* seria a trajetória ótima de *a* até *c*. Portanto, P'(b,c) não pode ser melhor do que P(b,c) que, consequentemente, é a trajetória ótima de *b* até *c*.

1.3. Caminho mais Curto Determinístico

Determinar a(s) trajetória(s) ótima(s) de A até J no grafo ponderado a seguir.



Sistema

Grafo $G = (N,\Gamma^{+})$ em camadas, ponderado e orientado.

Estágios

$k = 0, 1, 2, 3, 4$ correspondendo a cada uma das camadas do grafo.

Estados

X_k .... conjunto de vértices no estágio $k = 0, 1, 2, 3, 4$

$X_0 = \{A\}$ e $X_4 = \{J\}$, ou seja, no estágio inicial $k = 0$ só há o estado A enquanto no final apenas J.

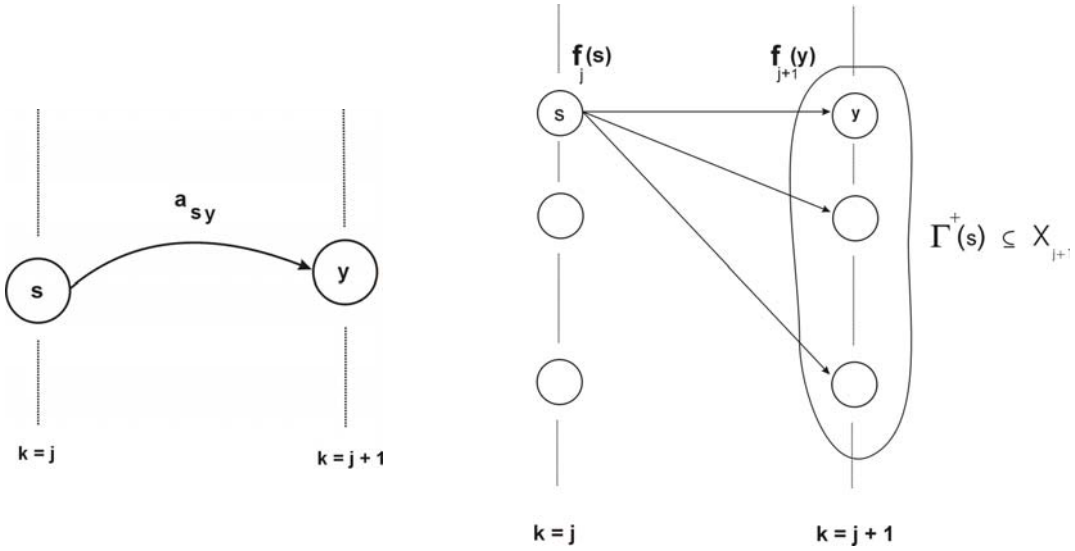
$N = \bigcup_{k=0}^{4} X_k$ .... conjunto de vértices do grafo G

Transição de Estado

$\Gamma^+(s)$ .... conjunto de *decisões admissíveis* no estado $s \in X_j$ ou, neste caso, de arcos incidentes para o exterior do estado\vértice $s \in X_j$

a_{sy} .... custo da transição do estado $s \in X_j$ para o estado $y \in X_{j+1}$

$f_j(s)$ .... custo/comprimento de um caminho mínimo de $s \in X_j$ até o alvo J



Função Critério (Equação Recursiva de Otimalidade)

$$f_j(s) = \underset{y \in \Gamma^+(s)}{\text{mínimo}} \{ a_{sy} + f_{j+1}(y) \}, \quad s \in X_j \ , \quad j = 0, 1, 2, 3$$

Condições de Contorno

$$f_4(s) = 0 \ , \quad \forall \ s \in X_4$$

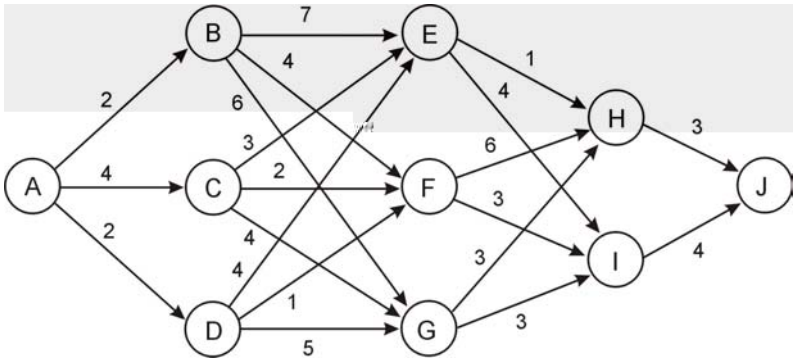
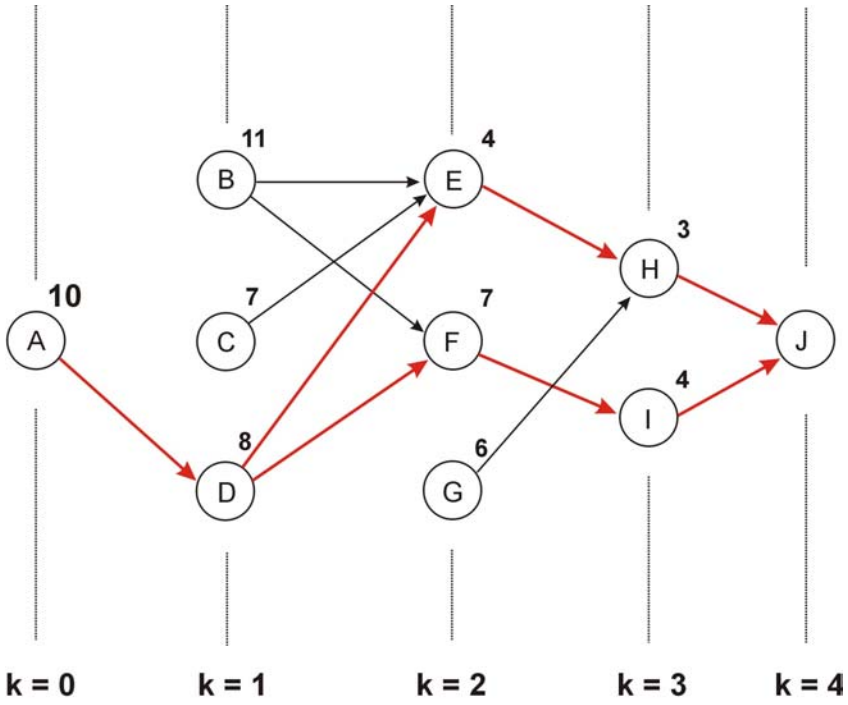


Diagrama Estado x Estágio

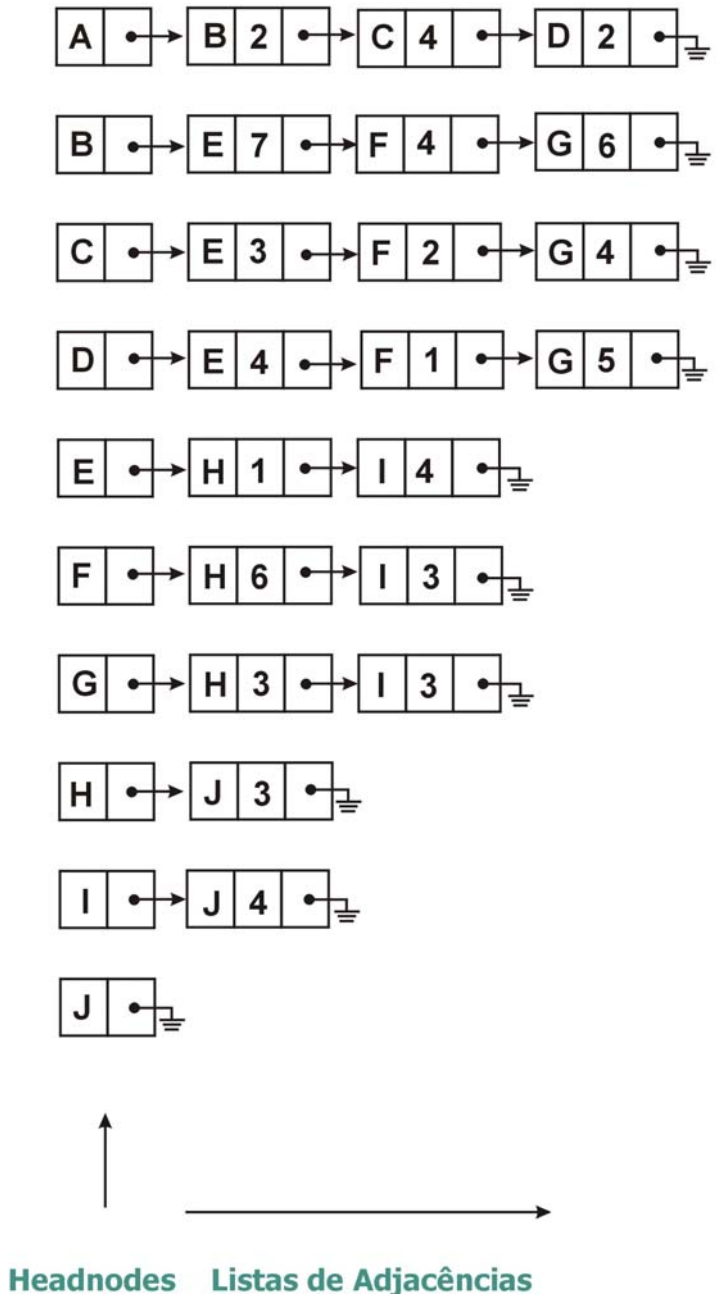


Trajetórias Ótimas: A D E H J

A D F I J

Custo Ótimo: 10

Estrutura de Adjacências do Grafo



Otimização Recursiva

$$f_3(s) = \underset{y \in \Gamma^+(s)}{\text{mínimo}}\{a_{sy} + f_4(y)\},\ s \in X_3 = \{H, I\}\ \text{e}\ f_4(s) = 0,\ \Gamma^+(s) = \{J\}, \forall s \in X_3$$

$$f_2(s) = \underset{y \in \Gamma^+(s)}{\text{mínimo}}\{a_{sy} + f_3(y)\},\ s \in X_2 = \{E, F, G\},\ \Gamma^+(s) = \{H, I\},\ \forall s \in X_2$$

$$f_1(s) = \underset{y \in \Gamma^+(s)}{\text{mínimo}}\{a_{sy} + f_2(y)\},\ s \in X_1 = \{B, C, D\},\ \Gamma^+(s) = \{E, F, G\},\ \forall s \in X_1$$

$$f_0(A) = \underset{y \in \Gamma^+(A)}{\text{mínimo}}\{a_{Ay} + f_1(y)\},\ \Gamma^+(A) = \{B, C, D\}$$

Estágio k = 3

S / y	a_{sJ} + $f_4(J)$ **J**	$f_3(s)$	Ir Para
H	3	3	**J**
I	4	4	**J**

Estágio k = 2

S / y	a_{sy} + $f_3(y)$ **H**	a_{sy} + $f_3(y)$ **I**	$f_2(s)$	Ir Para
E	4	8	4	**H**
F	9	7	7	**I**
G	6	7	6	**H**

Estágio k = 1

S / y	$a_{sy} + f_2(y)$			$f_1(s)$	Ir Para
	E	**F**	**G**		
B	11	11	12	11	**E ou F**
C	7	9	10	7	**E**
D	8	8	11	8	**E ou F**

Estágio k = 0

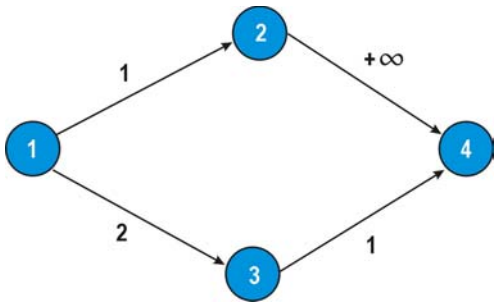
S / y	$a_{sy} + f_1(y)$			$f_0(s)$	Ir Para
	B	**C**	**D**		
A	13	11	10	**10**	**D**

1.4. Comentário sobre Algoritmos Míopes

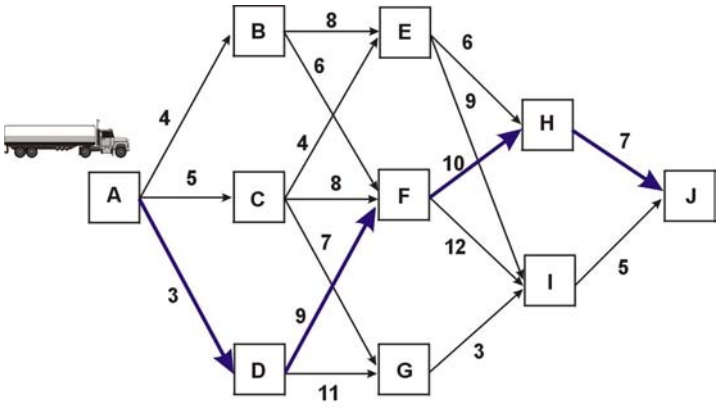
A técnica de construção de algoritmos heurísticos baseados na obtenção de uma boa solução, que eventualmente seja ótima, considerando a cada iteração a melhor decisão um passo à frente, ou seja, utilizando um critério de otimização meramente local, é bastante popular. Estas heurísticas são conhecidas genéricamente como *Míopes* ou *Gulosas* (Myopic/Greedy). Uma questão importante é discutir em que casos, ou para que classe de problemas, uma heurística do tipo míope garante a obtenção da solução ótima para qualquer instância.

Um algoritmo míope pode ser aplicado a qualquer problema cuja estrutura possa de alguma forma ser caracterizada como um *sistema de independência*, não havendo garantias de que a solução ótima será obtida a menos que este seja um *matróide* (Edmonds, 1971).

O problema de caminhos examinado infelizmente não se enquadra na classe de problemas para os quais a heurística míope oferece garantia de solução ótima. O exemplo trivial a seguir (caminho mais curto de 1 até 4) estabelece o contra-exemplo.



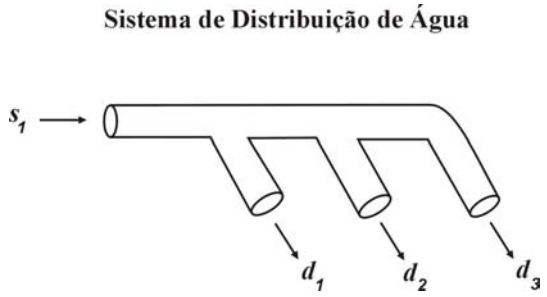
Entretanto, para a instância examinada anteriormente isto ocorre o que pode causar certo desconforto. Considere então a instância a seguir. Neste caso a aplicação da heurística míope leva a uma trajetória de valor igual a 29 quando a trajetória ótima tem valor 20, ou seja, um erro de 45% o que é muito significativo.



Trajetória Míope de Custo 29

1.5. Sistema de Distribuição de Água (Edgard and Himmelblau, 1989)

No sistema de distribuição de água, ilustrado a seguir, o valor máximo de s_1 é de 3.000 m^3 (por unidade de tempo). O fluxo s_1 deve ser distribuido nos três pontos mostrados na figura nas quantidades d_1, d_2 e d_3.



Assuma que o fluxo de água na saída de cada tubo depende apenas da quantidade de fluxo de água que chega a cada tubo. O retorno obtido pela distribuição da água em quantidades inteiras nos três tubos é informada a seguir:

Retorno pela Entrega de Água			
d_i (m^3x10^{-3})	$f_1(s_1,d_1)$ ($\$x10^{-3}$)	$f_2(s_2,d_2)$ ($\$x10^{-3}$)	$f_3(s_3,d_3)$ ($\$x10^{-3}$)
1	4	1	2
2	5	4	5
3	6	7	6

Qual a alocação de água nos tubos que maximiza o retorno do sistema.

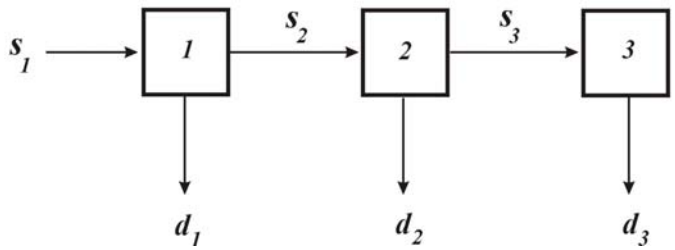
Modelagem Matemática

- Variáveis de Decisão

  s_i … fluxo de água no ponto de distribuição i = 1, 2, 3

  d_i ... fluxo de água no ponto de distribuição i = 1, 2, 3

Diagrama Funcional do Sistema de Distribuição de Água



- Restrições

Equações de Balanço de Material (Água)

$$\begin{cases} s_{i+1} = s_i - d_i \ , \ i = 1, 2 \\ s_3 = d_3 \end{cases}$$

Observe que estas equações implicam em que a restrição $s_1 = d_1 + d_2 + d_3$ é *redundante*, pois,

$$\begin{array}{rl} & \left\{\begin{array}{l} s_2 = s_1 - d_1 \\ s_3 = s_2 - d_2 \\ -s_3 = \quad - d_3 \end{array}\right. \\ (+) & \\ & s_2 = (s_1 + s_2) - (d_1 + d_2 + d_3) \ \therefore \\ \therefore & \ s_1 = d_1 + d_2 + d_3 \end{array}$$

Fluxo Máximo de Material (Água)

$$0 \leq d_1 + d_2 + d_3 \leq 3.000 \quad \text{e} \quad d_i \ \text{inteiro}, \ i = 1, 2, 3$$

- Função Objetivo

  $f_i(s_i,d_i)$ ... retorno obtido com a distribuição de d_i m^3 (por unidade de tempo) de água pelo ponto de distribuição $i = 1, 2, 3$

  x_0 ... retorno total obtido pela distribuição d_1, d_2 e d_3

$$x_0 = \sum_{i=1}^{3} f_i(s_i,d_i)$$

- Critério

$$\text{maximizar} \quad x_0 = \sum_{i=1}^{3} f_i(s_i,d_i)$$

- Modelo de Programação Matemática

$$\text{maximizar} \quad x_0 = \sum_{i=1}^{3} f_i(s_i,d_i)$$

sujeito a:

$$s_{i+1} = s_i - d_i \ , \quad i = 1, 2$$

$$s_3 = d_3$$

$$0 \leq d_1 + d_2 + d_3 \leq 3.000$$

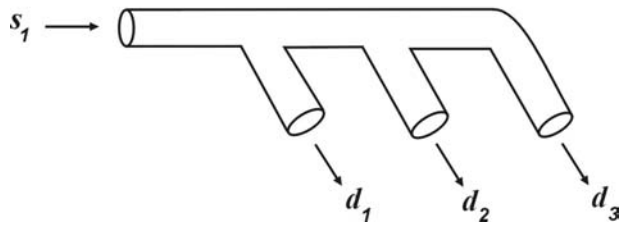
$$d_i \ , \text{ inteiro } \ i = 1, 2, 3$$

Modelagem da Programação Dinâmica

- Sistema

  Composto pelo sistema de distribuição de água com três pontos de distribuição, o fluxo na entrada de água e o esquema de retorno obtido pela repartição do fluxo de água entre os pontos.

Sistema de Distribuição de Água



- Estágios

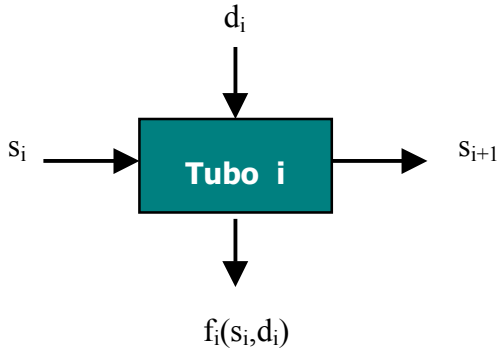
$k = i$ ... antes de alocar d_i m^3 (por unidade de tempo) ao ponto de distribuição $i = 1, 2, 3$;

$k = 4$ ... após distribuir o fluxo de água pelos três tubos.

- Estados

$\mathbf{s}_i$ ... fluxo de água que chega ao ponto de distribuição $i = 1, 2, 3$

- Equação de Transição de Estado



$s_{i+1} = s_i - d_i$, $i = 1, 2$

$s_3 = d_3$ (Condição de Contorno)

Função Critério

Sejam:

$f_i(s_i,d_i)$ ... retorno obtido com a distribuição de d_i m^3 (por unidade de tempo) de água pelo ponto de distribuição $i = 1, 2, 3$

d_i (m^3x10^{-3})	$f_1(s_1,d_1)$ ($\$x10^{-3}$)	$f_2(s_2,d_2)$ ($\$x10^{-3}$)	$f_3(s_3,d_3)$ ($\$x10^{-3}$)
1	2	1	4
2	5	4	5
3	6	7	6

$g^*_{i+1}(s_{i+1}) = g^*_{i+1}(s_i - d_i)$ ... retorno ótimo obtido por uma trajetória ótima que passa pelo estado s_i no estágio $i = 1, 2, 3$.

Condição de Contorno: $g^*_4(s_4) = g^*_4(s_3 - d_3) = g^*_4(0) = 0$

$$g^*_i(s_i) = \underset{d_i}{\text{máximo}} \{ f_i(s_i,d_i) + g^*_{i+1}(s_i - d_i) \}, \quad i = 1, 2, 3$$

com $g^*_4(s_4) = g^*_4(s_3 - d_3) = g^*_4(0) = 0$

Estágio $i = 3$

$$g^*_3(s_3) = \underset{\substack{0 \le s_3 \le 3 \\ 0 \le d_3 \le s_3}}{\text{máximo}} \{ f_3(s_3,d_3) + g^*_4(0) \}$$

Estágio $i = 2$

$$g_2^*(s_2) = \text{máximo } \{ f_2(s_2,d_2) + g_3^*(s_2 - d_2) \}$$
$$0 \le s_2 \le 3$$
$$0 \le d_2 \le s_2$$

Estágio $i = 1$

$$g_1^*(s_1) = \text{máximo } \{ f_1(s_1,d_1) + g_2^*(s_1 - d_1) \}$$
$$0 \le s_1 \le 3$$
$$0 \le d_1 \le s_1$$

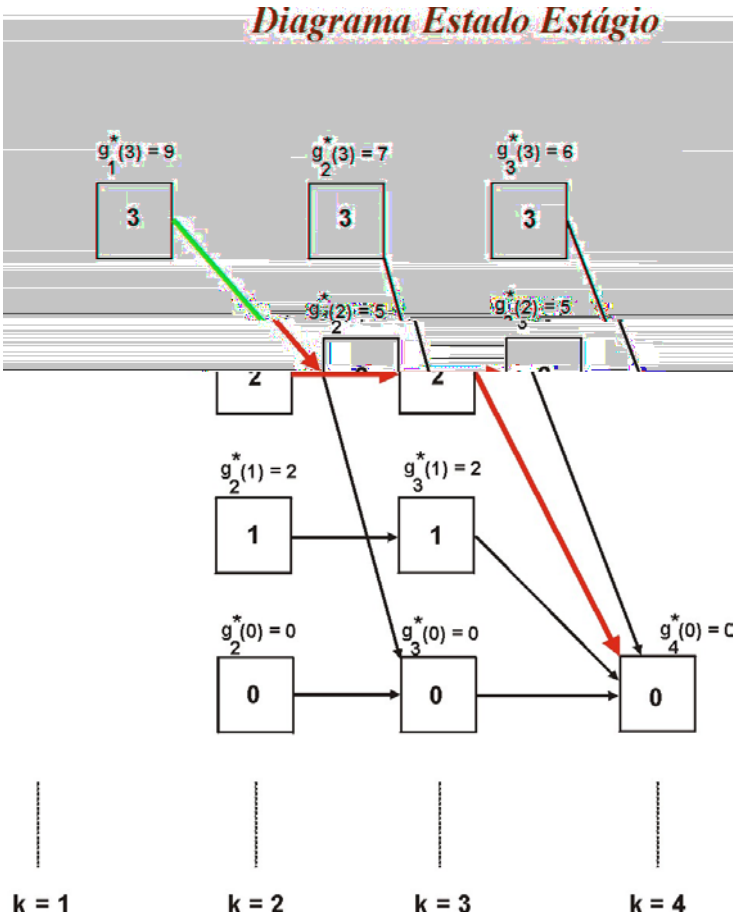
- Aplicação do Algoritmo de Programação Dinâmica

Estágio $i = 4$ $\quad s_4 = 0, \; g_4^*(0) = 0$

Estágio i = 3	d_3 / s_3	$f_3(s_3,d_3) + g_4^*(0)$				$g_3^*(s_3)$
		3	***2***	1	0	
	3	6	-	-	-	6
	2	-	5	-	-	***5***
	1	-	-	2	-	2
	0	-	-	-	0	0

Estágio i = 2	d_2 / s_2	$f_2(s_2,d_2) + g_3^*(s_2 - d_2)$				$g_2^*(s_2)$
		3	2	1	***0****	
	3	7	6	6	6	7
	2*	-	4	3	5	***5***
	1	-	-	1	2	2
	0	-	-	-	0	0

Estágio i = 1	$f_1(s_1,d_1) + g_2^*(s_1 - d_1)$				$g_1^*(s_1)$
s_1 \ d_1	3	2	***1****	***0***	
3*	6	7	9	6	***9***



Solução Ótima: $d^*_1 = 1,\ d^*_2 = 0,\ d^*_3 = 2$ e $x_0^* = 9$

1.6. Carregamento do Caminhão (Problema da Mochila)

Um caminhão tem 10 toneladas de capacidade de carga. Três produtos A, B e C estão disponíveis para transporte. Seus pesos e respectivos retornos pelo transporte estão na tabela a seguir.

Produtos	Valor ($)	Peso (t)/Unidade
A	**20**	**1**
B	**50**	**2**
C	**60**	**2**

Assumindo que, pelo menos um produto de cada tipo deve ser transportado, qual a composição de carga de maior retorno.

Passaremos, inicialmente, a formulação do modelo matemático definindo seus elementos básicos:

Modelo de Programação Matemática

- Variáveis de Decisão
- Restrições
- Função Objetivo
- Critério
- Modelo Matemático

(a) Variáveis de Decisão

x_j ..... quantidade do produto $j = 1, 2, 3$ alocado ao caminhão

(b) Restrições

b.1. Pelo menos um produto de cada tipo deve ser transportado no caminhão

$$x_j \geq 1, \quad j = 1, 2, 3$$

b.2. Capacidade máxima de carga do caminhão

$$x_1 + 2x_2 + 2x_3 \leq 10$$

b.3. Integralidade

$$x_j \text{ inteiro}, \quad j = 1, 2, 3$$

(c) Função Objetivo

$$x_0 = \left\{ \begin{array}{c} \text{Retorno pela} \\ \text{Composição} \\ \text{da Carga} \end{array} \right\} = 20x_1 + 50x_2 + 60x_3$$

(d) Critério

Maior retorno possível com o transporte da carga, ou seja,

$$\text{Maximizar} \quad x_0$$

(e) Modelo Matemático

$$(P): \text{ maximizar } x_0 = 20x_1 + 50x_2 + 60x_3$$

sujeito a:

$$x_1 + 2x_2 + 2x_3 \leq 10$$

$$x_j \geq 1 \text{ e inteiro}, \quad j = 1, 2, 3$$

A mudança de variável a seguir, permite transformar *(P)* em outro problema equivalente *(P')* de programação 0-1 o que facilitará a abordagem posterior.

Mudança de Variável:

$$x_j \geq 1 \quad \therefore \quad x_j - 1 \geq 0 \quad \therefore \quad y_j = x_j - 1 \geq 0\,, \quad j = 1, 2, 3$$

Faremos a seguinte transformação no problema *(P)*: $y_j + 1 = x_j\,, j = 1, 2, 3$

Restrição

$$x_1 + 2x_2 + 2x_3 \leq 10$$

$$(y_1 + 1) + 2(y_2 + 1) + 2(y_3 + 1) \leq 10 \quad \therefore$$

$$y_1 + 2y_2 + 2y_3 \leq 5$$

Função Objetivo

$$x_0 = 20(y_1 + 1) + 50(y_2 + 1) + 60(y_3 + 1)$$

$$= 20y_1 + 50y_2 + 60y_3 + 130$$

Fazendo $x'_0 = x_0 - 130$ temos o seguinte problema *(P')* transformado de *(P)*:

(P'): maximizar $x'_0 = 20y_1 + 50y_2 + 60y_3$

sujeito a:

$$y_1 + 2y_2 + 2y_3 \leq 5$$

$$y_j \geq 0 \text{ e inteiro } \quad j = 1, 2, 3$$

Os problemas *(P)* e *(P')* são equivalentes e podem ser resolvidos pela técnica da *Programação Dinâmica*.

Modelagem de (P') por Programação Dinâmica

Sistema

Caminhão, produtos, seus pesos e retorno no transporte

Estágios

$$\begin{cases} k = 0, & \text{antes de qualquer decisão} \\ k = j, & \text{após decidir o valor da variável } y_j \text{ , } j = 1, 2, 3 \end{cases}$$

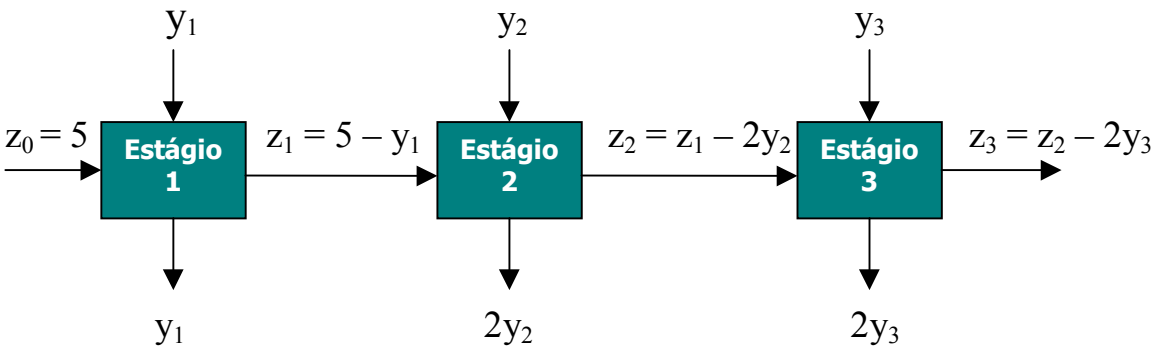
Estados

Definidos a cada estágio pela *capacidade de carga remanescente* z_j no caminhão.

Equação de Transição de Estado

$$\begin{cases} z_j = z_{j-1} - a_j.y_j \ , \quad j = 1, 2, 3 \\ z_0 = 5 \quad (\textit{Condição de Contorno}) \end{cases}$$

Esquemático do processo sequencial induzido:



Z_j ... conjunto dos estados viáveis no estágio $j = 0, 1, 2, 3$

$$\begin{cases} Z_0 = \{ 5 \} \\ Z_j = \{ 0, 1, 2, 3, 4, 5 \} \quad j = 1, 2, 3 \end{cases}$$

Função Critério

$g_j(z_j)$ ... retorno máximo de um caminho do estado z_j no estágio j até o alvo (estágio final)

$f_j(y_j)$ ... retorno obtido com y_j unidades do produto j (lucro elementar)

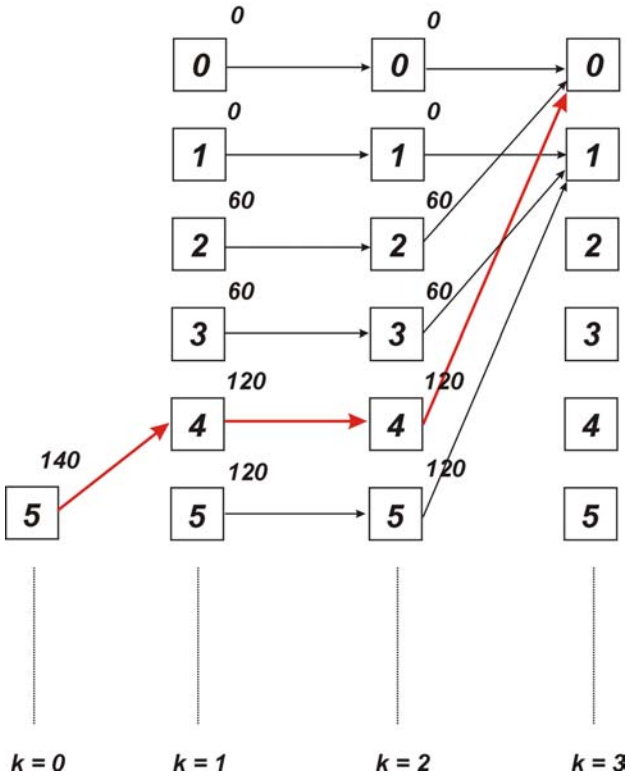
j	1	2	3
$f_j(y_j)$	$20y_1$	$50y_2$	$60y_3$

Equação Recursiva de Otimalidade

$$g^*_{j-1}(z_{j-1}) = \underset{z_j \in Z_j}{\text{máximo}} \{ f_j(y_j) + g^*_j(z_{j-1} - a_j.y_j) \}, \quad j = 1, 2, 3$$

$$g^*_3(z_3) = 0, \text{ com } z_3 = z_2 - a_3.x_3$$

Diagrama Estado x Estágio



Solução Ótima de *(P')*

$x_0^{*'} = 140$

$y_1^* = 1$

$y_2^* = 0$

$y_3^* = 2$

Solução Ótima de *(P)*

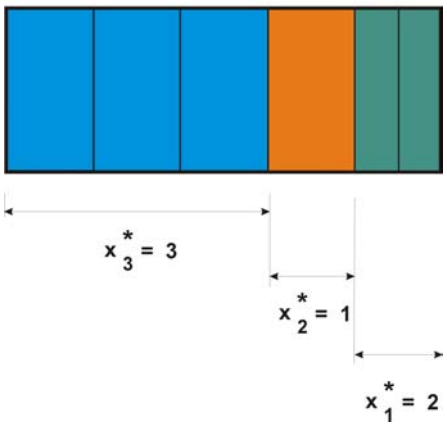
$x_0^* = 270$

$x_1^* = 2$

$x_2^* = 1$

$x_3^* = 3$

Solução Ótima



2. Programação Dinâmica Determinística com Horizonte Limitado

2.1. Conceitos e Definições

Serão apresentadas as definições e os conceitos a seguir, necessários a formalização da programação dinâmica determinística com horizonte limitado, embora muitos dos conceitos sejam os mesmos para os modelos probabilísticos e com horizonte ilimitado.

- ***Sistema***
- ***Estágios***
- ***Estados Viáveis, Estado Inicial e Alvo***
- ***Decisões Admissíveis***
- ***Equação de Transição de Estado***
- ***Custo (Lucro) Elementar***
- ***Política Admissível***
- ***Critério***
- ***Trajetórias***
- ***Problema de Programação Dinâmica***
- ***Princípio da Otimalidade de Bellman***
- ***Equação Recursiva de Otimalidade***

2.2. Sistema, Estágios, Estados e Alvo

Sistema

Pode ser *completamente descrito*, a cada *estágio*, pela especificação do seu *estado*.

Estágio

Variável discreta $\boldsymbol{k}$ que *determina a ordem* em que ocorrem modificações no sistema.

$$\boldsymbol{k} = 0, 1, 2, \ldots, t$$

Estágio Inicial *Estágio Final*

Estado

Variável $\boldsymbol{y(k)} = (\boldsymbol{y_1^k}, \boldsymbol{y_2^k}, ..., \boldsymbol{y_n^k}) \in \Re^{\boldsymbol{n}}$ que descreve completamente as características observáveis do sistema em cada estágio. O *mesmo estado* pode ocorrer em *diferentes estágios*.

Conjunto de Estados Viáveis no Estágio k

$\boldsymbol{Y(k)} \subset \Re^{\boldsymbol{n}}$, *estados* que a variável $\boldsymbol{y(k)}$ pode assumir no estágio $\boldsymbol{k}$.

O número de elementos de $\boldsymbol{Y(k)}$ é sempre *finito*.

Estado Inicial

Estado *único* em que se encontra o sistema no estágio inicial $\boldsymbol{k} = 0$, ou seja,

$$\boldsymbol{Y}(0) = \{\boldsymbol{y}(0)\} \quad \text{com} \quad \boldsymbol{y}(0) = (\boldsymbol{y}_1^0, \boldsymbol{y}_1^0, ..., \boldsymbol{y}_{\boldsymbol{n}}^0) \in \Re^{\boldsymbol{n}}$$

Alvo

Conjunto constituído dos *estados viáveis* $\boldsymbol{y(t)}$ no *estágio final* $\boldsymbol{t}$.

2.3. Decisões Admissíveis

Decisão

Variável $\boldsymbol{u(k)} = (\boldsymbol{u_1^k}, \boldsymbol{u_2^k}, ..., \boldsymbol{u_m^k}) \in \Re^{\boldsymbol{m}}$ que *aplicada ao sistema* quando este se encontra no *estado* $\boldsymbol{y(k)}$ influencia, de alguma forma, o estado em que o sistema se encontrará no *estágio* seguinte $(\boldsymbol{k}+1)$.

Conjunto de Decisões Admissíveis no Estágio k

$\boldsymbol{U(k)} \subset \Re^{\boldsymbol{m}}$, *decisões* que podem atuar sobre o sistema quando este se encontra no *estágio* $\boldsymbol{k}$ e no *estado* $\boldsymbol{y(k)}$. O número de elementos de $\boldsymbol{U(k)}$ é sempre *finito*.

2.4. Equação de Transição de Estado

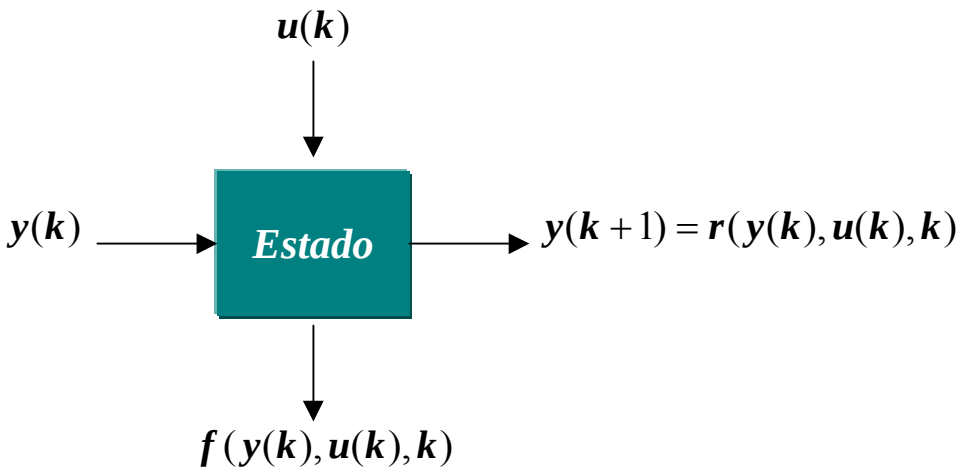
Relação entre o estado $y(k)$ em um dado estágio k, a decisão aplicada $u(k)$, e o estado resultante $y(k+1)$.

$$r : \Re^n \text{ x } \Re^m \text{ x } N \rightarrow \Re^n$$
$$(y(k), u(k), k) \mapsto r(y(k), u(k), k) = y(k+1)$$

2.5. Custo Elementar

Custo (lucro) devido à atuação da decisão $u(k)$ aplicada ao sistema no estado $y(k)$ e no estágio k.

$$f : \Re^n \text{ x } \Re^m \text{ x } N \rightarrow \Re^n$$
$$(y(k), u(k), k) \mapsto f(y(k), u(k), k)$$



2.6. Política Admissível

Política admissível aplicada a $\bar{y} = y(k_0) \in Y(k_0)$, $k_0 \in \{0, 1, 2, \ldots, t-1\}$ é uma seqüência de decisões $[u(k)]_{k=k_0}^{k=t-1} = (u(k_0), u(k_0+1), \ldots, u(t-1))$ tal que se definirmos $y(k+1) = r(y(k), u(k), k)$, $k = k_0, k_0+1, \ldots, t-1$ então:

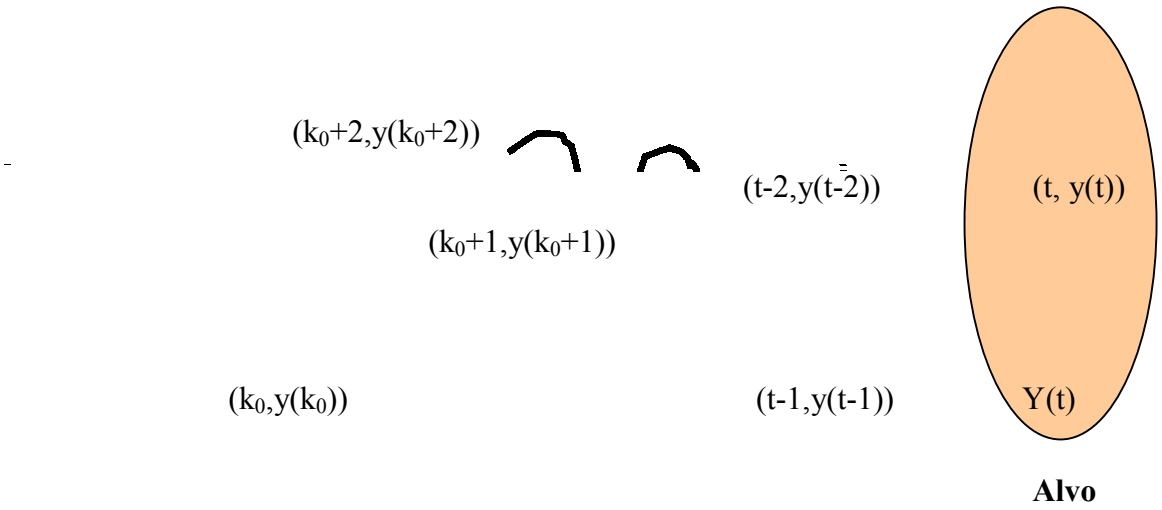
- $u(k) \in U(k)$, $k = k_0, k_0+1, \ldots, t-1$;
- $y(k) \in Y(k)$, $k = k_0+1, k_0+2, \ldots, t$.

O *conjunto das políticas admissíveis* aplicadas a $\bar{y}$ receberá a notação $\Omega(\bar{y}, k_0)$.

2.7. Trajetória

Trajetória *gerada por uma política admissível* $[u(k)]_{k=k_0}^{k=t-1}$ em $\bar{y} = y(k_0)$ é o *conjunto dos pontos* $(k, y(k))$, $k = k_0, k_0+1, \ldots, t$ e onde:

$$y(k+1) = r(y(k), u(k), k),\ k = k_0, k_0+1, \ldots, t-1$$



2.8. Função Critério

Para que o *Princípio da Otimalidade* possa ser utilizado é necessário que a função critério pertença a classe de funções *decomponíveis* (Nemhauser, 1966 e Mitten, 1964) e, para tanto, deverá ser *separável* e *monotônica* (*monótona não decrescente*).

Será utilizada a seguinte função:

$$g : (y(k_0), [u(k)]_{k=k_0}^{k=t-1}, k_0) \mapsto \sum_{k=k_0}^{t-1} f(y(k), u(k), k)$$

onde $[u(k)]_{k=k_0}^{k=t-1} \in \Omega(y(k_0), k_0)$ e

$$y(k+1) = r(y(k), u(k), k), \;\; k = k_0, k_0 + 1, ..., t-1.$$

A função $g(.)$ é *separável* quando, para $h : \Re^2 \to \Re$ e $g : \Re^{t-1-k_0} \to \Re$ tem-se:

$$g(y(k_0), [u(k)]_{k=k_0}^{k=t-1}, k_0) = h(f(y(k_0), u(k_0), k_0), g(y(k_0+1), [u(k)]_{k=k_0+1}^{k=t-1}, k_0+1))$$

Como foi definida, a função critério $g(.)$ é claramente separável, pois,

$$g(y(k_0), [u(k)]_{k=k_0}^{k=t-1}, k_0) = \sum_{k=k_0}^{t-1} f(y(k), u(k), k) =$$

$$= f(y(k_0), u(k_0), k_0) + \sum_{k=k_0+1}^{t-1} f(y(k), u(k), k)$$

Por outro lado, será *monótona não decrescente* quando um crescimento na função de retorno $f(.)$ implica em crescimento em $g(.)$, ou seja, para

$$f(y(k_0), \bar{u}(k_0), k_0) \geq f(y(k_0), u(k_0), k_0)$$

então $g(y(k_0), [\bar{u}(k)]_{k=k_0}^{k=t-1}, k_0) \geq g(y(k_0), [u(k)]_{k=k_0}^{k=t-1}, k_0)$

2.9. PPD e Princípio da Otimalidade de Bellman

Problema de Programação Dinâmica

Encontrar, *se existir*, uma *política admissível* $\left[u^*(k)\right]_{k=0}^{k=t-1}$ que, aplicada a $\boldsymbol{y}(0)$, leva o sistema a um estado $\boldsymbol{y}(\boldsymbol{t}) \in \boldsymbol{Y}(\boldsymbol{t})$ do estágio $\boldsymbol{t}$ e *minimiza* (*maximiza*) *o valor da função critério*, isto é:

$$\boldsymbol{g}^*(\boldsymbol{y}(0), \left[\boldsymbol{u}^*(\boldsymbol{k})\right]_{\boldsymbol{k}=0}^{\boldsymbol{k}=\boldsymbol{t}-1}, 0) \quad = \quad \underset{[\boldsymbol{u}(\boldsymbol{k})]_{\boldsymbol{k}=0}^{\boldsymbol{k}=\boldsymbol{t}-1} \in \Omega(\boldsymbol{y}(0),0)}{\text{mínimo}} \quad \boldsymbol{g}(\boldsymbol{y}(0), \left[\boldsymbol{u}(\boldsymbol{k})\right]_{\boldsymbol{k}=0}^{\boldsymbol{k}=\boldsymbol{t}-1}, 0)$$

Se existir uma política admissível, então existirá uma política ótima $\left[u^*(k)\right]_{k=0}^{k=t-1}$, pois, *o número de políticas admissíveis é finito*.

Princípio de Otimalidade de Bellman

Se $\left[\bar{\boldsymbol{u}}(\boldsymbol{k})\right]_{\boldsymbol{k}=\boldsymbol{k}_0}^{\boldsymbol{k}=\boldsymbol{t}-1}$, $\boldsymbol{k}_0 = 0, 1, 2, \ldots, t\text{-}1$ é uma *política ótima* considerando $\boldsymbol{y}(\boldsymbol{k}_0)$ como *estado inicial* então $\left[\bar{\boldsymbol{u}}(\boldsymbol{k})\right]_{\boldsymbol{k}=\boldsymbol{k}_0+1}^{\boldsymbol{k}=\boldsymbol{t}-1}$ será uma *política ótima* considerando $y(k_0+1) = r(y(k_0), \bar{u}(k_0), k_0)$ como *estado inicial*.

Demonstração

Seja $k_0 \in \{0, 1, 2, \ldots, t-1\}$ e uma *política ótima* $[\bar{u}(k)]_{k_0}^{t-1}$ considerando

$\bar{y} =$ •($_0$) (co) 1415(m)1587(o)-56(a)]TJ/TT18 1 Tf0 0.502 0.502 scn63.0943 0

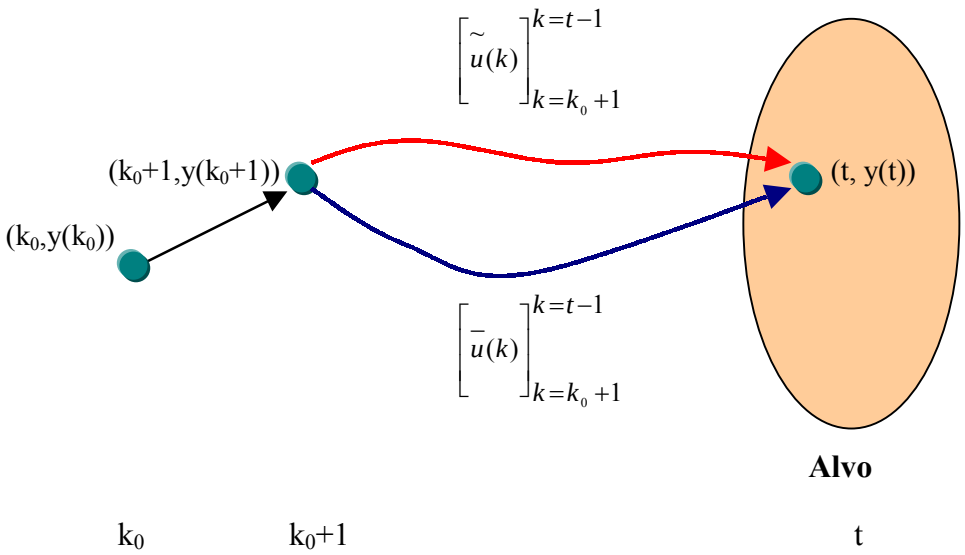
é a a

0

Admitamos, ***por absurdo***, que exista uma seqüência de decisões $\left[\tilde{\boldsymbol{u}}(\boldsymbol{k})\right]_{\boldsymbol{k}=\boldsymbol{k}_0+1}^{\boldsymbol{k}=\boldsymbol{t}-1} \in \Omega(\tilde{\boldsymbol{y}}, \boldsymbol{k}_0+1)$ aplicada a $\tilde{\boldsymbol{y}} = \boldsymbol{y}(\boldsymbol{k}_0+1) = \boldsymbol{r}(\bar{\boldsymbol{y}}, \tilde{\boldsymbol{u}}(\boldsymbol{k}_0), \boldsymbol{k}_0)$ e tal que:

$$\boldsymbol{g}^*(\tilde{\boldsymbol{y}}, \left[\tilde{\boldsymbol{u}}(\boldsymbol{k})\right]_{\boldsymbol{k}=\boldsymbol{k}_0+1}^{\boldsymbol{k}=\boldsymbol{t}-1}, \boldsymbol{k}_0+1) \;<\; \boldsymbol{g}(\tilde{\boldsymbol{y}}, \left[\bar{\boldsymbol{u}}(\boldsymbol{k})\right]_{\boldsymbol{k}=\boldsymbol{k}_0+1}^{\boldsymbol{k}=\boldsymbol{t}-1}, \boldsymbol{k}_0+1) \quad \text{------} \quad (1)$$

Estaremos admitindo, portanto, a existência de uma trajetória gerada por uma política admissível $\left[\tilde{\boldsymbol{u}}(\boldsymbol{k})\right]_{\boldsymbol{k}=\boldsymbol{k}_0+1}^{\boldsymbol{k}=\boldsymbol{t}-1}$ cujo valor da função critério correspondente é menor que o considerado ótimo.



Definindo uma seqüência formada pela primeira decisão de $\left[\bar{\boldsymbol{u}}(\boldsymbol{k})\right]_{\boldsymbol{k}_0}^{\boldsymbol{t}-1}$, ou seja, $\bar{\boldsymbol{u}}(\boldsymbol{k}_0)$ e toda a seqüência de decisões $\left[\tilde{\boldsymbol{u}}(\boldsymbol{k})\right]_{\boldsymbol{k}=\boldsymbol{k}_0+1}^{\boldsymbol{k}=\boldsymbol{t}-1}$. Obtem-se:

$$\left[\hat{\boldsymbol{u}}(\boldsymbol{k})\right]_{\boldsymbol{k}=\boldsymbol{k}_0}^{\boldsymbol{k}=\boldsymbol{t}-1} = (\bar{\boldsymbol{u}}(\boldsymbol{k}_0), \tilde{\boldsymbol{u}}(\boldsymbol{k}_0+1), \tilde{\boldsymbol{u}}(\boldsymbol{k}_0+2), \ldots, \tilde{\boldsymbol{u}}(\boldsymbol{t}-1))$$

Esta seqüência é, por definição, *admissível*: $\left[\hat{\boldsymbol{u}}(\boldsymbol{k})\right]_{\boldsymbol{k}=\boldsymbol{k}_0}^{\boldsymbol{k}=\boldsymbol{t}-1} \in \Omega(\bar{\boldsymbol{y}}, \boldsymbol{k}_0)$.

Além disto, de (1), $\boldsymbol{g}(\bar{\boldsymbol{y}},[\tilde{\boldsymbol{u}}(\boldsymbol{k})]_{\boldsymbol{k}=\boldsymbol{k}_0+1}^{\boldsymbol{k}=\boldsymbol{t}-1},\mathrm{k}_0+1) < \bar{\boldsymbol{g}}(\bar{\boldsymbol{y}},\boldsymbol{k}_0)$ o que é *absurdo* uma vez que contraria a definição de $\bar{\boldsymbol{g}}(\bar{\boldsymbol{y}},\boldsymbol{k}_0)$, isto é, o fato de que $[\bar{\boldsymbol{u}}(\boldsymbol{k})]_{\boldsymbol{k}_0}^{\boldsymbol{t}-1}$ ser uma *política ótima* aplicada a $\bar{\boldsymbol{y}} = \boldsymbol{y}(\boldsymbol{k}_0)$.

Concluímos que:

$$\boldsymbol{g}(\bar{\boldsymbol{y}},[\tilde{\boldsymbol{u}}(\boldsymbol{k})]_{\boldsymbol{k}=\boldsymbol{k}_0+1}^{\boldsymbol{k}=\boldsymbol{t}-1},\boldsymbol{k}_0+1) = \text{mínimo}\ \boldsymbol{g}(\tilde{\boldsymbol{y}},[\boldsymbol{u}(\boldsymbol{k})]_{\boldsymbol{k}=\boldsymbol{k}_0+1}^{\boldsymbol{k}=\boldsymbol{t}-1},\boldsymbol{k}_0+1)$$

$$[\boldsymbol{u}(\boldsymbol{k})]_{\boldsymbol{k}_0+1}^{\boldsymbol{t}-1} \in \Omega(\tilde{\boldsymbol{y}},\boldsymbol{k}_0+1)$$

$$= \bar{\boldsymbol{g}}(\tilde{\boldsymbol{y}},\boldsymbol{k}_0+1)$$

Finalmente, obtemos como resultado a *Equação Recursiva de Otimalidade*:

$$\bar{\boldsymbol{g}}(\bar{\boldsymbol{y}},\boldsymbol{k}_0) = \text{mínimo}\ \{\boldsymbol{f}(\bar{\mathrm{y}},\boldsymbol{u}(\boldsymbol{k}_0),\boldsymbol{k}_0) + \bar{\mathrm{g}}(\boldsymbol{f}(\bar{\mathrm{y}},\boldsymbol{u}(\boldsymbol{k}_0),\boldsymbol{k}_0),\boldsymbol{k}_0+1)\}$$

$$\boldsymbol{u}(\boldsymbol{k}_0) \in \boldsymbol{U}(\bar{\boldsymbol{y}},\boldsymbol{k}_0)$$

$$\boldsymbol{f}(\bar{\mathrm{y}},\boldsymbol{u}(\boldsymbol{k}_0),\boldsymbol{k}_0) \in \boldsymbol{Y}(\boldsymbol{k}_0+1)$$

3. Programação Dinâmica Determinística com Horizonte Ilimitado

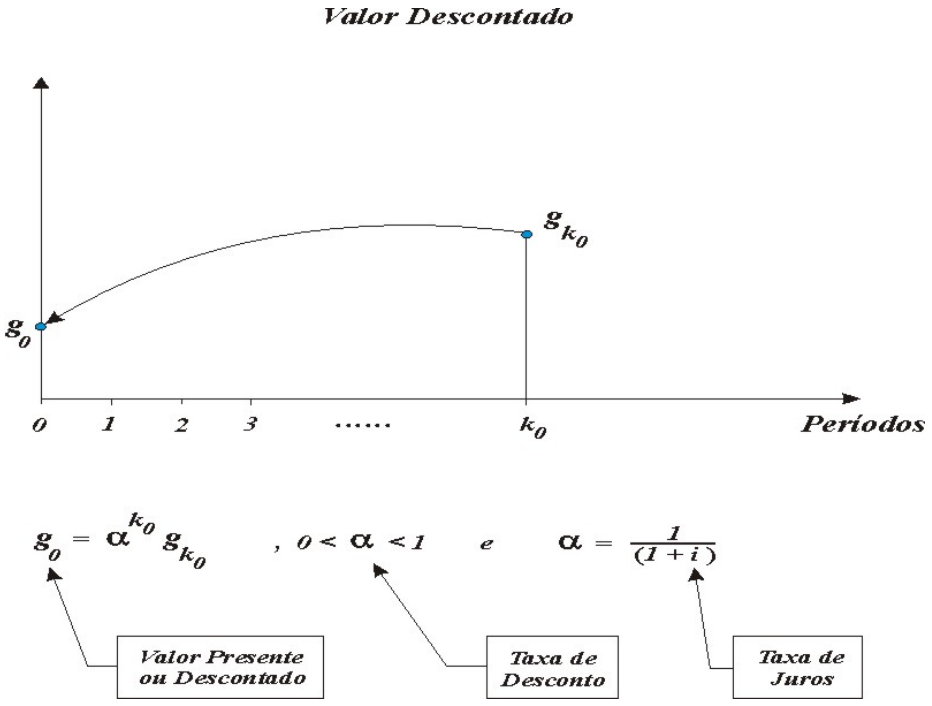
3.1. Condição de Utilização e Critério

Quando o horizonte é ilimitado a equação de otimalidade não pode mais ser resolvida por recursão "*backwards*" a partir do estágio final, portanto, condições devem ser estabelecidas para garantir a validade da equação.

Os valores da função critério devem ser finitos, ou seja, a imagem da função critério g deve ser *limitada inferiormente* (superiormente) no caso de minimização (maximização).

Critério do Valor Presente

Considerar uma taxa de desconto $0 < \alpha < 1$ que em muitas aplicações será representada por $\alpha = (1 + i/100)^{-1}$ onde $i\,\%$ é uma taxa de juros na unidade de tempo.



A função critério a ser utilizada é o *valor atual* (presente), no estágio k_0, dos custos (lucros) elementares, ou seja:

$$g(y(k_0),[u(k)]_{k_0}^{+\infty},k_0) = \sum_{k=k_0}^{+\infty} \alpha^k f(y(k),u(k),k)$$

Se todos os $f(y(k),u(k),k)$ são uniformemente limitados por, digamos B, e $\alpha < 1$ então $g(y(k_0),[u(k)]_{k_0}^{+\infty},k_0) < \dfrac{B\alpha^{k_0}}{(1-\alpha)}$, pois,

$$g(y(k_0),[u(k)]_{k_0}^{+\infty},k_0) = \sum_{k=k_0}^{+\infty} \alpha^k f(y(k),u(k),k) < \sum_{k=k_0}^{+\infty} \alpha^k B = \frac{B\alpha^{k_0}}{(1-\alpha)}$$

Com esta função é possível avaliar o *mérito relativo* de uma alternativa pela conversão de uma seqüência infinita de "retornos" em um número único.

3.2. Conceito de Estacionaridade

Um *Problema de Programação Dinâmica Determinístico* é *estacionário* quando:

- Sua *equação de transição de estado* não depende do estágio k, ou seja:

$$r(y(k),u(k),k) = r(y(k),u(k)), \forall\, k;$$

- A função de *custos (lucros) elementares* não depende do estágio k, ou seja:

$$f(y(k),u(k),k) = f(y(k),u(k)),\ \forall\, k;$$

- O conjunto de decisões que podem atuar sobre o sistema quando este se encontra no estágio k é função apenas do estado $y(k)$, ou seja:

$$U(y(k),k) = U(y(k));$$

- Se $y(k) \in Y(k),\ u(k) \in U(y(k))$ e $r(y(k),u(k)) \in Y(k+1)$ então $(\forall\, k',\ y(k') \in Y(k'),\ y(k') = y(k),\ u(k') = u(k)$ $\Rightarrow\ r(y(k'),u(k')) \in Y(k'+1))$;

- O número total de estados viáveis é *finito*;

- O número total de decisões admissíveis é *finito*.

Estaremos assumindo, portanto, que todas as funções de retorno, decisões e fenômenos externos (como requisitos de demanda) são idênticos para todos os períodos (estágios).

Em um *Problema de Programação Dinâmica Determinístico com Horizonte Ilimitado e Estacionário*, uma *política* $[\boldsymbol{u}(\boldsymbol{k})]_{\boldsymbol{k}=0}^{+\infty}$ é *estacionária* quando a trajetória $\{y(0), y(1), y(2),....\}$ associada a esta política é tal que $y(k) = y(k') \Rightarrow u(k) = u(k')$.

Concluímos então, que a aplicação de uma política estacionária requer apenas o conhecimento do estado atual do sistema e não a seqüência

histórica de eventos que conduzem à aquele estado, isto é, *a cada vez que o sistema retorna à aquele estado a mesma decisão será tomada.*

Além disto, concluímos que o número de políticas estacionárias admissíveis é finito e, portanto, *se existir uma política estacionária existirá um política estacionária ótima*.

Como não existem métodos gerais para resolução de problemas de *programação dinâmica determinísticos com horizonte ilimitado* não estacionários nos ocuparemos apenas dos que *satisfazem a hipótese de estacionaridade.*

3.3. Critério do Valor Presente em Problemas Estacionários

O objetivo será determinar uma política estacionária que minimiza $\sum_{k=0}^{n} \alpha^k f(y(k),u(k))$.

Aplicando o *Princípio da Otimalidade de Bellman* com $n \to +\infty$ tem-se:

$$g^*(y(k_0),k_0) = \underset{\substack{\boldsymbol{u}(\boldsymbol{k}_0) \in \boldsymbol{U}(\boldsymbol{y}(\boldsymbol{k}_0)) \\ \boldsymbol{r}(\boldsymbol{y}(\boldsymbol{k}_0),\boldsymbol{u}(\boldsymbol{k}_0)) \in \boldsymbol{Y}(\boldsymbol{k}_0+1)}}{mínimo} \{ \alpha^{k_0} f(y(k_0),u(k_0)) + g^*(r(y(k_0),u(k_0)),k_0+1) \}$$

Observar que $g^*(y(k_0),k_0)$ é o valor atual mínimo, no estágio $k=0$, da série $f(y(k_0),u(k_0))$, $f(y(k_0+1),u(k_0+1))$,..... obtida pela aplicação de política estacionária ótima ao estado $y(k_0)$ no estágio k_0 (*Princípio da Otimalidade*).

O valor presente da mesma série referenciada ao estágio $k = k_0$ será, portanto, dada por $\boldsymbol{g^*(y(k_0),k_0)/\alpha^{k_0}}$. Logo, no caso geral, $g^*(y(k_0),0)$ corresponderá ao valor atual de uma série que se inicia no estágio $k = 0$ e pode ser representada por:

$$g^*(y(k_0),k_0) = \alpha^{k_0} g^*(y(k_0),0)$$

e $$g^*(y(k_0+1),k_0+1) = \alpha^{k_0+1} g^*(r(y(k_0),u(k_0)),0)$$

Substituindo na *Equação Recursiva de Otimalidade com Horizonte Ilimitado* temos:

$$g^*(y(k_0),0) = \underset{\substack{u(k_0)\in U(y(k_0))\\ r(y(k_0),u(k_0))\in Y(k_0+1)}}{mínimo}\{\frac{\alpha^{k_0}}{\alpha^{k_0}} f(y(k_0),u(k_0)) + \frac{\alpha^{k_0+1}}{\alpha^{k_0}} g^*(r(y(k_0),u(k_0)),0)\}$$

$$= \underset{\substack{u(k_0)\in U(y(k_0))\\ r(y(k_0),u(k_0))\in Y(k_0+1)}}{mínimo}\{ f(y(k_0),u(k_0)) + \alpha\, g^*(r(y(k_0),u(k_0)),0)\}$$

É interessante observar que, na dedução da *Equação Recursiva de Otimalidade com Horizonte Ilimitado* foi utilizada implícitamente a hipótese de que *qualquer estado viável pode ser considerado estado inicial*. Embora, necessáriamente, isto não se verifique em todos os casos, é sempre possível adotar, levando em conta que os estados se repetem a cada estágio, que a série é infinita e desejamos determinar a decisão ótima para cada estado (*a política estacionária ótima não se preocupa com o estágio*), uma mudança

de escala dos estágios fazendo com que qualquer estado possa ser considerado como inicial.

Como a hipótese de *estacionaridade* implica em um número finito de estados que se repetirão a cada estágio, podemos simplificar a notação adotando i e j para representar os estados com $i = 1, 2,..., n$.

Sem perda de generalidade, assumiremos também que toda a decisão viável aplicada a um estado conduz a outro estado viável, ou seja:

$$u \in U(i) \Rightarrow j = r(i,u).$$

A *Equação Recursiva de Otimalidade com Horizonte Ilimitado* pode ser representada então por:

$$g^*(i) = \underset{\substack{u \in U(i) \\ j = r(i,u)}}{mínimo} \{ f(i,u) + \alpha\, g^*(j) \}$$

Portanto, *a política ótima é aquela que a cada estado viável i, em cada estágio k, determina a melhor decisão para passar ao próximo estágio.*

O sistema formado pelas equações de otimalidade *não pode ser resolvido diretamente*, pois, a decisão ótima para cada estado i só pode ser determinada se os valores ótimos $g^*(j),\ j = 1,2,...,n$ forem conhecidos, ou seja, seria necessário conhecer a própria estratégia ótima para cada estado.

3.4. Métodos de Solução da Equação Recursiva de Otimalidade com Horizonte Ilimitado

São dois os processos iterativos utilizados para resolver o problema:

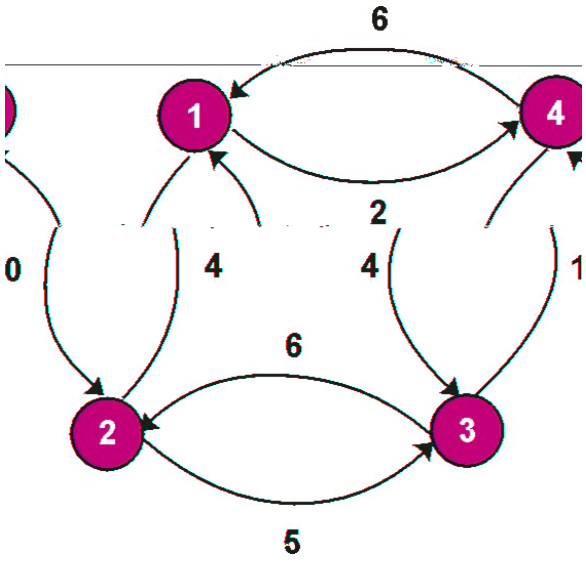
- Método das Aproximações no Espaço dos Critérios;
- Método das Aproximações no Espaço das Políticas.

A *Aproximação no Espaço dos Critérios* consiste em escolher *arbitrariamente* um conjunto inicial de critérios $g^*(i),\ i = 1,2,...,n$ e calcular novos valores utilizando a *Equação Recursiva de Otimalidade com Horizonte Ilimitado* $g^*(i) = \underset{\substack{u \in U(i) \\ j = r(i,u)}}{mínimo}\{ f(i,u) + \alpha\, g^*(j)\},\ i = 1,2,...,n.$

O procedimento deve ser repetido até que não haja *diferenças significativas* entre os valores de $g^*(i)$ em duas iterações consecutivas.

Exemplo

Determinar a política ótima para o problema representado pelo grafo orientado a seguir utilizando uma taxa de desconto $\alpha = 0{,}8$.



Solução:

Temos então o seguinte sistema de equações funcionais:

$$g^*(1) = mínimo\{(0 + 0{,}8.g^*(2)), (2 + 0{,}8.g^*(4))\}$$
$$g^*(2) = mínimo\{(4 + 0{,}8.g^*(1)), (5 + 0{,}8.g^*(3))\}$$
$$g^*(3) = mínimo\{(6 + 0{,}8.g^*(2)), (1 + 0{,}8.g^*(4))\}$$
$$g^*(4) = mínimo\{(6 + 0{,}8.g^*(1)), (4 + 0{,}8.g^*(3))\}$$

Vamos adotar, arbitrariamente, $g^*(i) = 10,\ i = 1, 2, 3, 4$.

As duas primeiras iterações são desenvolvidas a seguir:

1ª Iteração

$$g^*(1) = mínimo\{(0 + 0{,}8x10)\,,\,(2 + 0{,}8x10)\} = \ 8 \quad ..... \quad u(1) = 2$$
$$g^*(2) = mínimo\{(4 + 0{,}8x10)\,,\,(5 + 0{,}8x10)\} = 12 \quad ..... \quad u(2) = 1$$
$$g^*(3) = mínimo\{(6 + 0{,}8x10)\,,\,(1 + 0{,}8x10)\} = \ 9 \quad ..... \quad u(3) = 4$$
$$g^*(4) = mínimo\{(6 + 0{,}8x10)\,,\,(4 + 0{,}8x10)\} = 12 \quad ..... \quad u(4) = 3$$

2ª Iteração

$$g^*(1) = mínimo\{(0 + 0{,}8x12)\,,\ (2 + 0{,}8x12)\} = \ 8 \quad ..... \quad u(1) = 2$$
$$g^*(2) = mínimo\{(4 + 0{,}8x8)\ ,\ (5 + 0{,}8x9)\} \ = 12 \quad ..... \quad u(2) = 1$$
$$g^*(3) = mínimo\{(6 + 0{,}8x12)\,,\ (1 + 0{,}8x12)\} \ = \ 9 \quad ..... \quad u(3) = 4$$
$$g^*(4) = mínimo\{(6 + 0{,}8x8)\ ,\ (4 + 0{,}8x9)\} \ = 12 \quad ..... \quad u(4) = 3$$

Os resultados correspondentes a dezeseis (16) iterações são apresentados nos quadros da página a seguir onde a coluna Δ representa a diferença em módulo entre os valores obtidos para os critérios em duas iterações sucessivas. Na décima sexta iteração a diferença para todos os critérios é de 0,07, portanto, para uma tolerância de 0,10 poderíamos encerrar o procedimento com uma política ótima:

$$u^*(1) = 2,\ u^*(2) = 1,\ u^*(3) = 4 \ \text{ e } \ u^*(4) = 1.$$

Estado	*1ª*			*2ª*			*3ª*			*4ª*		
y	$g^*(y)$	*u*	Δ	$g^*(y)$	*u*	Δ	$g^*(y)$	*u*	Δ	$g^*(y)$	*u*	Δ
1	8	2	2,00	9,6	2	1,6	8,32	2	1,28	9,34	2	1,02
2	12	1	2,00	10,4	1	1,6	11,68	1	1,28	10,66	1	1,02
3	9	4	2,00	10,6	4	1,6	9,96	4	0,64	10,98	4	1,02
4	12	3	2,00	11,2	3	0,8	12,48	3	1,28	11,97	3	0,51

Estado	*5ª*			*6ª*			*7ª*			*8ª*		
y	$g^*(y)$	*u*	Δ	$g^*(y)$	*u*	Δ	$g^*(y)$	*u*	Δ	$g^*(y)$	*u*	Δ
1	8,52	2	0,82	9,18	2	0,66	8,66	2	0,52	9,08	2	0,42
2	11,48	1	0,82	10,82	1	0,66	11,34	1	0,52	10,92	1	0,42
3	10,57	4	0,41	11,23	4	0,66	10,97	4	0,26	11,39	4	0,42
4	12,79	3	0,82	12,46	3	0,33	12,98	3	0,52	12,77	3	0,21

Estado	*9ª*			*10ª*			*11ª*			*12ª*		
y	$g^*(y)$	*u*	Δ	$g^*(y)$	*u*	Δ	$g^*(y)$	*u*	Δ	$g^*(y)$	*u*	Δ
1	8,74	2	0,34	9,01	2	0,27	8,79	2	0,21	8,97	2	0,17
2	11,26	1	0,34	10,99	1	0,27	11,21	1	0,21	11,03	1	0,17
3	11,22	4	0,17	11,49	4	0,27	11,38	4	0,11	11,55	4	0,17
4	13,11	3	0,34	12,98	3	0,13	13,19	3	0,21	13,03	1	0,16

Estado	*13ª*			*14ª*			*15ª*			*16ª*		
y	$g^*(y)$	*u*	Δ	$g^*(y)$	*u*	Δ	$g^*(y)$	*u*	Δ	$g^*(y)$	*u*	Δ
1	8,82	2	0,14	8,94	2	0,11	8,85	2	0,09	8,92	2	0,07
2	11,17	1	0,14	11,06	1	0,11	11,15	1	0,09	11,08	1	0,07
3	11,43	4	0,12	11,54	4	0,11	11,45	4	0,09	11,52	4	0,07
4	13,17	1	0,14	13,06	1	0,11	13,15	1	0,09	13,08	1	0,07

A *convergência do procedimento é assintótica*, ou seja, não ocorre necessariamente, em número finito de iterações (Wagner, 1969). Adicionalmente, o fato de uma política admissível permanecer a mesma por diversas iterações não implica, necessariamente, em que esta seja a política estacionária ótima. Este fato pode ser observado no exemplo onde até a 11ª iteração a política manteve-se como (2, 1, 4, 3). Na realidade a política estacionária ótima pode não ser única.

Para o *Procedimento de Aproximações no Espaço das Políticas* vamos supor que $\{u^*(1), u^*(2), ..., u^*(n)\}$ seja uma *política estacionária ótima*. Portanto:

$$g^*(i) = \underset{\substack{\boldsymbol{u(i)} \in \boldsymbol{U(i)} \\ \boldsymbol{j = r(i,u(i))}}}{mínimo} \{ f(i,u(i)) + \alpha . g^*(j) \} = f(i,u^*(i)) + \alpha . g^*(j), \quad j = 1,2,...,n$$

Observe que, neste caso, o argumento de u^* é o estado i, pois, trata-se de política estacionária.

Escolhendo, arbitráriamente, uma *política admissível* $\{\bar{u}(1), \bar{u}(2), ..., \bar{u}(n)\}$ e, resolvendo o sistema $\bar{g}(i) = f(i,\bar{u}(i)) + \alpha . \bar{g}(j), \; i, j = 1,2,...,n$ obtemos os valores de $\bar{g}(i), i = 1,2,...,n.$

Testamos, então, estes valores nas *equações de otimalidade* e, sempre que ocorrer:

$$\underset{\substack{\boldsymbol{u(i)} \in \boldsymbol{U(i)} \\ \boldsymbol{j = r(i,u(i))}}}{mínimo} \{ f(i,u(i)) + \alpha . \bar{g}(j) \} = f(i,u'(i)) + \alpha . \bar{g}(j') < \bar{g}(i)$$

substituimos na política admissível $\bar{u}(i)$ por $\bar{u}'(i)$. Com esta *nova política* resolvemos o sistema novamente até que não haja nenhuma mudança na *política admissível* que será então uma *política ótima*.

Consideremos o mesmo exemplo utilizado anteriormente escolhendo, arbitráriamente, a política $\bar{u}_0(1)=2, \bar{u}_0(2)=3, \bar{u}_0(3)=4, \bar{u}_0(4)=3$ onde $\bar{u}_s(i)$ e $\bar{g}_s(i)$ representam, respectivamente, a decisão admissível no estado i e o valor do critério na iteração s.

1ª Iteração

$$\begin{cases} \bar{g}_0(1) = 0 + \alpha.\bar{g}_0(2) \\ \bar{g}_0(2) = 5 + \alpha.\bar{g}_0(3) \\ \bar{g}_0(3) = 1 + \alpha.\bar{g}_0(4) \\ \bar{g}_0(4) = 4 + \alpha.\bar{g}_0(3) \end{cases} \quad \therefore$$

$$\begin{cases} \bar{g}_0(1) - 0{,}8\text{x}\,\bar{g}_0(2) = 0 \\ \bar{g}_0(2) - 0{,}8\text{x}\,\bar{g}_0(3) = 5 \\ \bar{g}_0(3) - 0{,}8\text{x}\,\bar{g}_0(4) = 1 \\ -0{,}8\mathbf{x}\,\bar{g}_0(3) + \bar{g}_0(4) = 4 \end{cases}$$

Resolvendo este sistema de quatro equações lineares e quatro incógnitas, obtemos:

$$\bar{g}_0(1) = 11{,}3$$

$$\bar{g}_0(2) = 14{,}3$$

$$\bar{g}_0(3) = 11{,}7$$

$$\bar{g}_0(4) = 13{,}3$$

Testando estes valores nas *equações de otimalidade*:

$$\begin{cases} \text{mínimo}\{(0+0{,}8\text{x}14{,}3)\ ,\ (2+0{,}8\text{x}13{,}3)\} = 11{,}3 = \bar{g}_0(1) \\ \text{mínimo}\{(4+0{,}8\text{x}11{,}3)\ ,\ (5+0{,}8\text{x}11{,}7)\} = 13{,}0 < \bar{g}_0(2) \ \Rightarrow \bar{u}_1(2) = 1 \\ \text{mínimo}\{(6+0{,}8\text{x}14{,}3)\ ,\ (1+0{,}8\text{x}13{,}3)\} = 11{,}7 = \bar{g}_0(3) \\ \text{mínimo}\{(6+0{,}8\text{x}11{,}3)\ ,\ (4+0{,}8\text{x}11{,}7)\} = 13{,}3 = \bar{g}_0(4) \end{cases}$$

A *nova política admissível* passa a ser então:

$$\bar{u}_1(1) = 2, \bar{u}_1(2) = 1, \bar{u}_1(3) = 4, \bar{u}_1(4) = 3$$

2ª Iteração

Com a *nova política admissível* o sistema de equações passa a ser o seguinte:

$$\begin{cases} \bar{g}_1(1) = 0 + \alpha.\bar{g}_1(2) \\ \bar{g}_1(2) = 4 + \alpha.\bar{g}_1(1) \\ \bar{g}_1(3) = 1 + \alpha.\bar{g}_1(4) \\ \bar{g}_1(4) = 4 + \alpha.\bar{g}_1(3) \end{cases} \quad \therefore$$

$$\begin{cases} \bar{g}_1(1) - 0{,}8\text{x}\,\bar{g}_1(2) = 0 \\ \bar{g}_1(2) - 0{,}8\text{x}\,\bar{g}_1(1) = 4 \\ \bar{g}_1(3) - 0{,}8\text{x}\,\bar{g}_1(4) = 1 \\ -0{,}8\mathbf{x}\,\bar{g}_1(3) + \bar{g}_1(4) = 4 \end{cases}$$

Sua solução é dada por:

$$\bar{g}_1(1) = 8{,}9$$

$$\bar{g}_1(2) = 11{,}1$$

$$\bar{g}_1(3) = 11{,}7$$

$$\bar{g}_1(4) = 13{,}3$$

Testando estes valores nas *equações de otimalidade*:

$$\begin{cases} \text{mínimo}\{(8,9)\ ,\ (2+0,8\text{x}13,3)\} \ = \ 8,9 \ = \bar{g}_1(1) \\ \text{mínimo}\{(11,1)\ ,\ (5+0,8\text{x}11,7)\} = 13,0 = \bar{g}_1(2) \\ \text{mínimo}\{(6+0,8\text{x}14,3)\ ,\ (11,7)\} = 11,7 = \bar{g}_1(3) \\ \text{mínimo}\{(6+0,8\text{x}11,3)\ ,\ (13,3)\} = 13,1 < \bar{g}_1(4) \ \Rightarrow \bar{u}_2(4) = 1 \end{cases}$$

A *nova política admissível* passa a ser então:

$$\bar{u}_2(1) = 2, \bar{u}_2(2) = 1, \bar{u}_2(3) = 4, \bar{u}_2(4) = 1$$

3ª Iteração

Com a *nova política admissível* o sistema de equações na terceira iteração passa a ser o seguinte:

$$\begin{cases} \bar{g}_2(1) = 0 + \alpha.\bar{g}_2(2) \\ \bar{g}_2(2) = 4 + \alpha.\bar{g}_2(1) \\ \bar{g}_2(3) = 1 + \alpha.\bar{g}_2(4) \\ \bar{g}_2(4) = 6 + \alpha.\bar{g}_2(3) \end{cases} \quad \therefore$$

$$\begin{cases} \bar{g}_2(1) - 0{,}8\text{x}\,\bar{g}_2(2) = 0 \\ \bar{g}_2(2) - 0{,}8\text{x}\,\bar{g}_2(1) = 4 \\ \bar{g}_2(3) - 0{,}8\text{x}\,\bar{g}_2(4) = 1 \\ -0{,}8\mathbf{x}\,\bar{g}_2(1) + \bar{g}_2(4) = 4 \end{cases}$$

Sua solução é dada por:

$$\bar{g}_2(1) = 8{,}9$$

$$\bar{g}_2(2) = 11{,}1$$

$$\bar{g}_2(3) = 11{,}5$$

$$\bar{g}_2(4) = 13{,}1$$

Testando estes valores nas *equações de otimalidade*:

$$\begin{cases} \text{mínimo}\{(8{,}9)\ ,\ (2+0{,}8\text{x}13{,}1)\} = 8{,}9 = \bar{g}_2(1) \\ \text{mínimo}\{(11{,}1)\ ,\ (5+0{,}8\text{x}11{,}5)\} = 11{,}1 = \bar{g}_2(2) \\ \text{mínimo}\{(6+0{,}8\text{x}11{,}1)\ ,\ (11{,}5)\} = 11{,}5 = \bar{g}_2(3) \\ \text{mínimo}\{(13{,}1)\ ,\ (4+0{,}8\text{x}11{,}5)\} = 13{,}1 = \bar{g}_2(4) \end{cases}$$

Como não há alterações na política admissível o procedimento se encerra com a seguinte solução ótima:

$$u^*(1) = 2, u^*(2) = 1, u^*(3) = 4, u^*(4) = 1$$
$$g^*(1) = 8{,}9\ ,\ g^*(2) = 11{,}1\ ,\ g^*(3) = 11{,}5\ ,\ g^*(4) = 13{,}1$$

Embora o método seja *convergente para uma política estacionária ótima em um número finito de iterações* requer maior esforço computacional a cada etapa, pois, deve resolver um sistema de equações lineares. Quanto melhor for a estimativa inicial da política admissível mais rápida será a convergência.

4. Programação Dinâmica Probabilística com Horizonte Limitado

4.1. Conceito

Nos modelos determinísticos quando uma decisão atua sobre o sistema o estado resultante é completamente previsível. Portanto, quando uma seqüência de decisões admissíveis atua, a partir de um estado inicial, todas as transições de estado e seus custos ou retornos correspondentes são conhecidos com precisão.

Os princípios da Programação Dinâmica podem ser estendidos para modelos estocásticos permitindo transições de estado que envolvem incertezas. No caso da *Programação Dinâmica Probabilística com Horizonte Limitado* atuam sobre o sistema *fatores aleatórios* de tal forma que a decisão que atua em um estado de determinado estágio não determina completamente o estado que o sistema assumirá no estágio seguinte. Não há, como no caso determinístico, uma trajetória ótima e uma política ótima. A solução será representada por um conjunto de decisões ótimas com cada uma delas associada a um estado do estágio correspondente. Ao conjunto de todas as decisões ótimas denominamos estratégia ótima.

4.2. Equação Recursiva de Otimalidade

Como um componente estocástico está presente, seja na forma de um distúrbio aleatório ou ruído dependendo do contexto, a função critério adotada será o *valor esperado* da soma das contribuições em cada estágio.
Logo:

$$g(y(k_0),k_0) = E\left[\sum_{k=k_0}^{t-1} f(y(k),u(k),k)\right]$$

que preserva as propriedades da aditividade e monotonicidade.

Portanto, a *Equação Recursiva de Otimalidade*, neste caso, é representada pelo valor esperado ótimo (*maximio* ou *minimo*) da soma das contribuições de cada estágio, ou seja:

$$g^*(\bar{y}, k_0) = \underset{u(k_0) \in U(\bar{y}, k_0)}{mínimo}\, E\left[\sum_{k=k_0}^{t-1} f(y(k), u(k), k)\right] =$$

$$= \underset{u(k_0) \in U(\bar{y}, k_0)}{mínimo}\, E\left[f(\bar{y}, u(k_0), k_0) + \sum_{k=k_0+1}^{t-1} f(y(k), u(k), k)\right] =$$

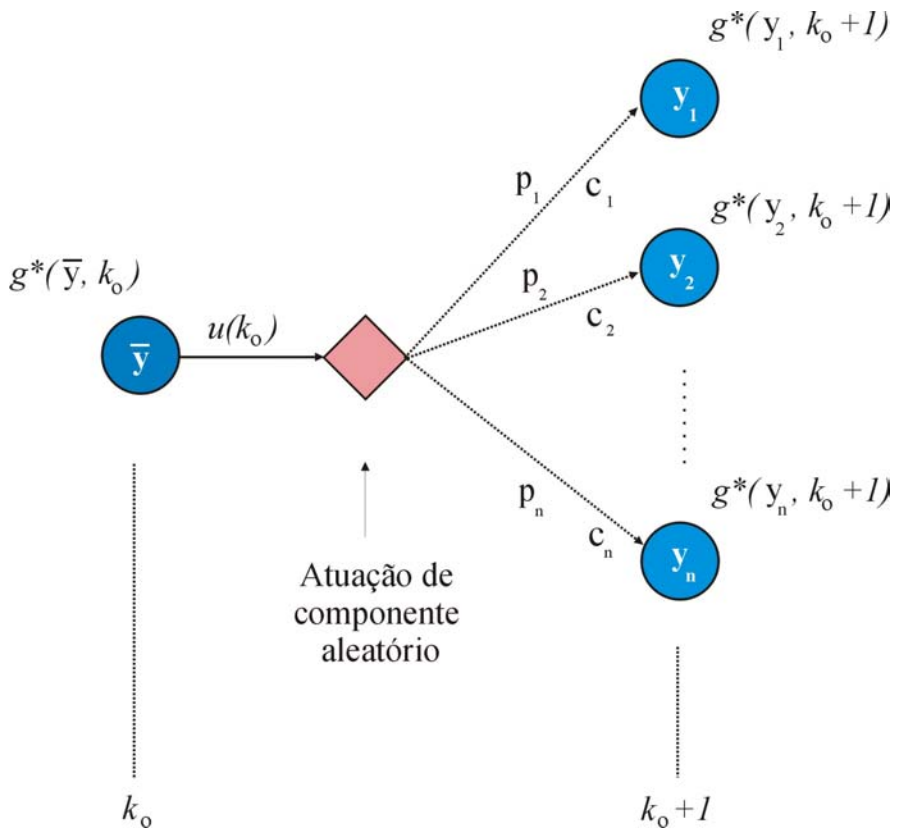
$$= \underset{u(k_0) \in U(\bar{y}, k_0)}{mínimo}\left\{ E\left[f(\bar{y}, u(k_0), k_0)\right] + \underset{u(k_0+1) \in U(y(k_0+1), k_0+1)}{mínimo}\left\{ E\left[\sum_{k=k_0+1}^{t-1} f(y(k), u(k), k)\right] \right\}\right\}$$

$$= \underset{u(k_0) \in U(\bar{y}, k_0)}{mínimo}\left\{ E\left[f(\bar{y}, u(k_0), k_0)\right] + g^*(r(\bar{y}, u(k_0), k_0), k_0 + 1) \right\}$$

onde $r(\bar{y}, u(k_0), k_0)$ e $f(\bar{y}, u(k_0), k_0)$ são variáveis aleatórias com distribuição de probabilidade conhecida. Como as distribuições de probabilidades dos novos estados são conhecidas temos um *problema de decisão sob risco*.

É oportuno observar que, nos casos em que os custos (lucros/ganhos) elementares $f(\bar{y},u(k_0),k_0)$ não dependem do estágio k_0 os cálculos ficam simplificados, pois, os valores esperados $E\left[f(\bar{y},u(k_0),k_0)\right]$ devem ser calculados apenas uma vez.

Sendo $y_1, y_2, ..., y_n$ os estados admissíveis no estágio k_0+1 o problema pode ser esquematizado como a seguir, onde cada mudança de estágio ocorre em dois momentos: no primeiro a decisão $u(k_0)$ é aplicada ao estado $\bar{y}$ no estágio k_0, em seguida atua a componente aleatória levando então ao estado resultante da decisão no estágio k_0+1.



Para uma *distribuição de probabilidades* dos estados $y_1, y_2, ..., y_n$ resultantes da aplicação da decisão $u(k_0)$ sobre $\bar{y}$ dada por $P\{Y_j = y_j, u(k_0)\} = p_j$, $j = 1,2,...,n$ e sendo c_j a contribuição à função objetivo (custo/lucro/ganho elementar) quando o estado resultante for y_j teremos:

$$E[f(\bar{y}, u(k_0), k_0) + g^*(r(\bar{y}, u(k_0), k_0), k_0 + 1)] =$$

$$= \sum_{j=1}^{n} E\left[(f(\bar{y}, u(k_0), k_0) + g^*(r(\bar{y}, u(k_0), k_0), k_0 + 1)) / Y_j = y_j, u(k_0)\right] P\{Y_j = y_j, u(k_0)\} =$$

$$= \sum_{j=1}^{n} p_j \left[c_j + g^*(y_j, k_0 + 1)\right]$$

Em alguns casos é possível resolver a *equação recursiva de otimalidade* explicitamente, com considerável ganho computacional, e explorar as propriedades estruturais do sistema relacionandas a estratégia ótima. O exemplo a seguir reforça esta idéia e os conceitos apresentados.

4.3. Resolução Explícita da Equação Recursiva de Otimalidade

Considere um sistema, que se encontra inicialmente no estado i, com $k = 1,2,...,t$ estágios e para cada estágio existem $j = 1,2,...,n$ estados admissíveis. Quando uma decisão $u \in U$ (conjunto finito) é aplicada ao estado i há um retorno $R(i,u)$, sendo que o próximo estado será j com probabilidades $p_{ij}(u)$ conhecidas, $j = 1,2,...,n$.

Seja $V_t(i)$ o retorno máximo esperado no último estágio para este sistema, considerando i como estado inicial. Quando $k=1$, isto é, com o sistema atuando em um único estágio então a decisão ótima é dada por:

$$V_1(i) = \underset{\boldsymbol{u \in U}}{\textit{máximo}}\{ R(i,u) \} \qquad (1)$$

Considerando agora o sistema inicialmente no estado i e atuando em $k>1$ estágios, se o próximo estado for j então repete-se o problema anterior, isto é, um sistema iniciando em j atuando em $t-1$ estágios. Portanto, o melhor que se poderá obter, em relação ao valor esperado do retorno quando a decisão $u \in U$ for tomada é:

$$R(i,u) + \sum_j p_{ij}(u) V_{t-1}(j)$$

Como $V_t(i)$ é o melhor que se pode obter sem restrições para a ação inicial $u \in U$ temos a seguinte *equação recursiva de otimalidade*:

$$V_t(i) = \underset{\boldsymbol{u \in U}}{\textit{máximo}}\left\{ R(i,u) + \sum_j p_{ij}(u) V_{t-1}(j) \right\} \qquad (2)$$

Observe que esta equação (2), como não poderia deixar de ser, é equivalente a obtida anteriormente, com critério de minimização, aplicadas simplificações de notação que o caso permite.

A equação (2) pode ser resolvida recursivamente para $V_t(i)$ obtendo inicialmente $V_1(i)$ e, em seguida, utilizando o resultado com $t = 2$ na *equação recursiva de otimalidade* (2) obtendo $V_2(i)$ e assim sucessivamente.

Considere agora a aplicação deste modelo a um jogo simples no qual o jogador, em cada uma das t rodadas, pode apostar qualquer quantia não-negativa, limitada pela sua disponibilidade no momento da aposta, podendo ganhar aquela quantia com probabilidade p ou perder a referida quantia com probabilidade $q = 1 - p$. O jogador deseja, naturalmente, encerradas as t rodadas, maximizar seu lucro esperado. Nestas condições qual deve ser a estratégia ótima que *maximiza o valor esperado do logaritmo de sua disponibilidade inicial* ?

Seja, como anteriormente, $V_t(x)$ o retorno esperado máximo quando o jogador tiver x para apostar e t rodadas para jogar. A decisão em cada rodada é o *valor da aposta* e ficará definido como uma fração $0 \le \alpha \le 1$ da disponibilidade x no momento. Então a equação recursiva de otimalidade é:

$$V_t(x) = \underset{0 \le \alpha \le 1}{máximo} \left\{ pV_{t-1}(x + \alpha x) + (1 - p)V_{t-1}(x - \alpha x) \right\}$$

com a condição de contorno $V_0(x) = \log x$.

Fazendo, na equação anterior, $t = 1$ e utilizando a condição de contorno:

$$V_1(x) = \underset{0 \le \alpha \le 1}{máximo}\{\ p\log(x+\alpha x) + q\log(x-\alpha x)\ \} =$$

$$= \underset{0 \le \alpha \le 1}{máximo}\{\ p\log(1+\alpha)x + q\log(1-\alpha)x\ \} =$$

$$= \underset{0 \le \alpha \le 1}{máximo}\{\ [p\log(1+\alpha) + q\log(1-\alpha)] + \log x\ \} =$$

$$= \underset{0 \le \alpha \le 1}{máximo}\{\ \mathrm{g}(\alpha) + \log x\ \} \text{ com } \quad g(\alpha) = p\log(1+\alpha) + q\log(1-\alpha)$$

O máximo da função $g(\alpha)$ é obtido para $\alpha = p - q$ da seguinte forma:

$$\begin{cases} \dfrac{dg(\alpha)}{d\alpha} = \dfrac{p}{1+\alpha} - \dfrac{q}{1-\alpha} = 0 \quad \therefore \ \alpha = \dfrac{p-q}{p+q} = p - q \\ \dfrac{d^2 g(\alpha)}{d\alpha^2} = -\dfrac{p}{(1+\alpha)^2} - \dfrac{q}{(1-\alpha)^2} < 0 \end{cases}$$

Portanto, $V_1(x) = C + \log x$, $x > 0$ e $C = \log 2 + p\log p + q\log q$

Utilizando a equação recursiva de otimalidade para $t = 2$ tem-se:

$$V_2(x) = \underset{0 \le \alpha \le 1}{máximo}\{\ [p\log(x+\alpha x) + q\log(x-\alpha x)] + C\}$$

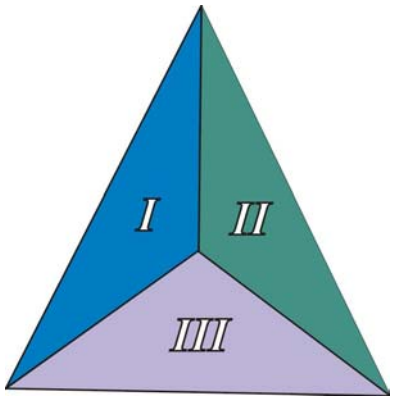
Resolvendo esta equação da mesma forma que a anterior:

$$V_2(x) = 2C + \log x$$

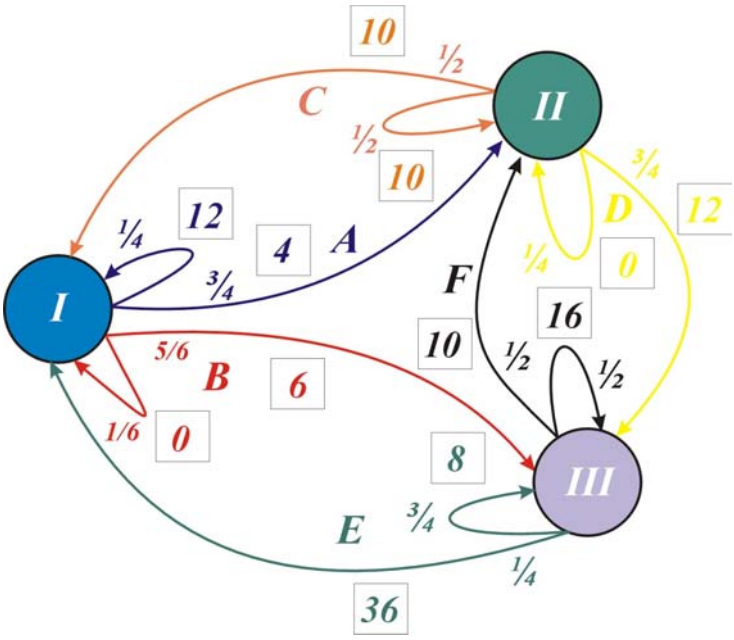
Portanto, é fácil deduzir que $V_t(x) = tC + \log x$ e a estratégia ótima do jogador será apostar uma fração $\alpha = p - q = 2p - 1$ da quantidade de recursos x em qualquer dos estágios/rodadas do jogo. É interessante notar que sendo $\alpha = p - q$ e $0 \le \alpha \le 1$ então $1/2 < p \le 1$. Caso $p \le 1/2$, caracterizando uma condição desfavorável de jogo, $\alpha = 0$ e a melhor estratégia é não jogar.

4.4. Resolução Recursiva da Equação de Otimalidade

Considere o jogo a seguir onde uma pessoa compra uma ficha que é lançada ao acaso sobre uma mesa triangular como a do esquema abaixo.



São iguais as probabilidades da ficha cair nas regiões I, II ou III. Se ao ser laçada a ficha cair na região I, o jogador escolhe e aciona uma das roletas A ou B. Se a roleta A for a escolhida a sua ficha terá probabilidade ¼ de permanecer na região I, ganhando 12, e uma probabilidade de ¾ de passar para a região II, ganhando 4. Se a roleta B for a escolhida algo semelhante ocorre com probabilidades e ganhos descritos na figura a seguir.



Se cair na região II, pode escolher entre as roletas C e D, e caindo na região III, pode escolher entre as roletas E e F. As probabilidades e os ganhos respectivos encontram-se indicados na figura anterior.

Após a primeira jogada, o jogador tem direito a mais duas partindo da região resultante do lance anterior e mantendo-se todas as outras condições.

O problema do jogador é determinar que roleta deve ser escolhida, a cada jogada, de modo a *maximizar seu ganho esperado*.

Os estados serão caracterizados, a cada estágio, pela região (I, II ou III) em que cada jogador estiver, em função das jogadas. Os estágios serão definidos por:

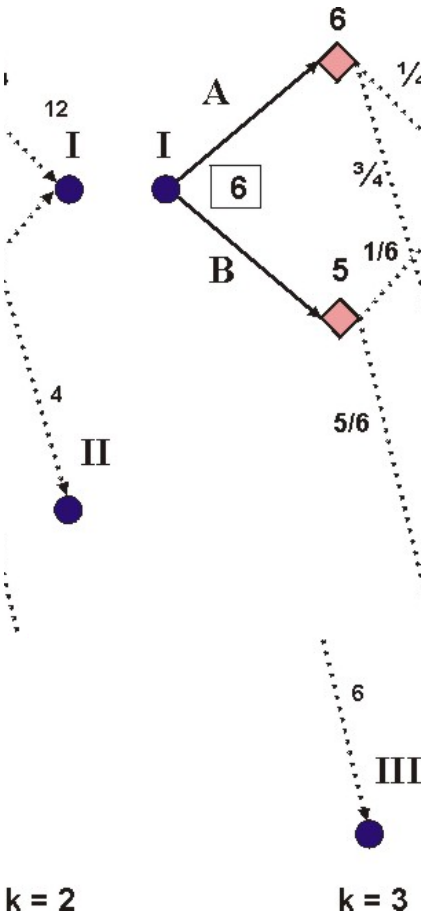
$$\begin{cases} k = 0, & \text{antes de realizar a primeira jogada} \\ k = j, & \text{após realizar a j-ésima jogada } (j = 1, 2, 3) \end{cases}$$

Aplicando o *Princípio da Otimalidade de Bellman*, no penúltimo estágio ($k = 2$), quando falta realizar a última jogada, não é possível conhecer com certeza em que região (estado) estará o jogador. Observe que o mesmo ocorre no problema determinístico com horizonte limitado, pois, *a política ótima não é conhecida a priori*. No caso *probabilístico*, porém, não sabemos exatamente qual será o resultado de cada decisão *em função da atuação do componente aleatório*. Entretanto, podemos afirmar que, estando o jogador na região I e escolhendo a *roleta A* o seu ganho esperado será:

$$\frac{1}{4} \text{x } 12 + \frac{3}{4} \text{x } 4 = 6$$

Se escolher a *roleta B*, seu ganho esperado será:

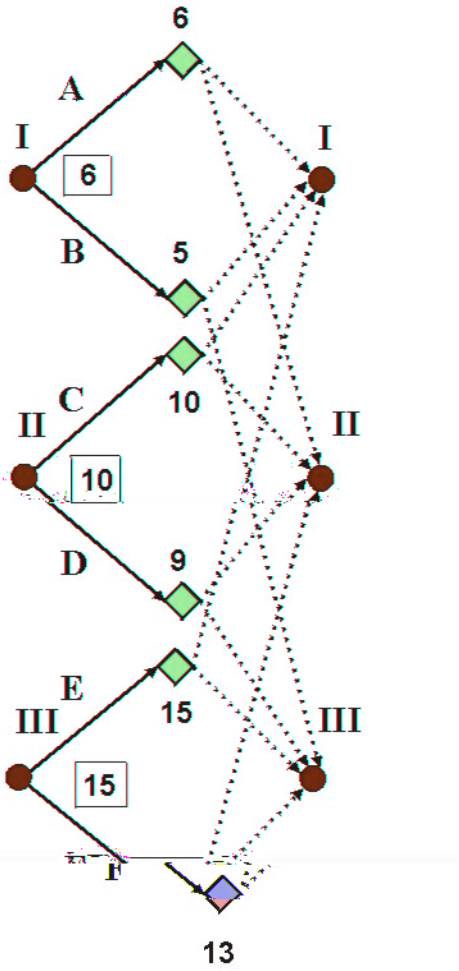
$$\frac{1}{6} \text{x } 0 + \frac{5}{6} \text{x } 6 = 5$$



Então o jogador, estando na região I, no estágio $k = 2$, deve optar pela roleta A, o que lhe proporcionará o *maior ganho esperado* $(g^*(I,2) = 6)$.

Procedendo de forma análoga para as regiões II e III obteremos o *diagrama estado x estágio* parcial correspondente aos estágios $k = 2$ e $k = 3$. Caso o jogador só tivesse direito a uma jogada essas seriam as suas decisões ótimas $(g^*(II,2) = 10$ e $g^*(III,2) = 15)$.

iagrama Estado x EstágioD



k = 2 **k = 3**

Suponhamos agora que faltam ainda duas jogadas para o encerramento da partida. O mesmo raciocínio se repete aqui, ou seja, o fator aleatório faz com que não se possa determinar com certeza em que região estará o jogador.

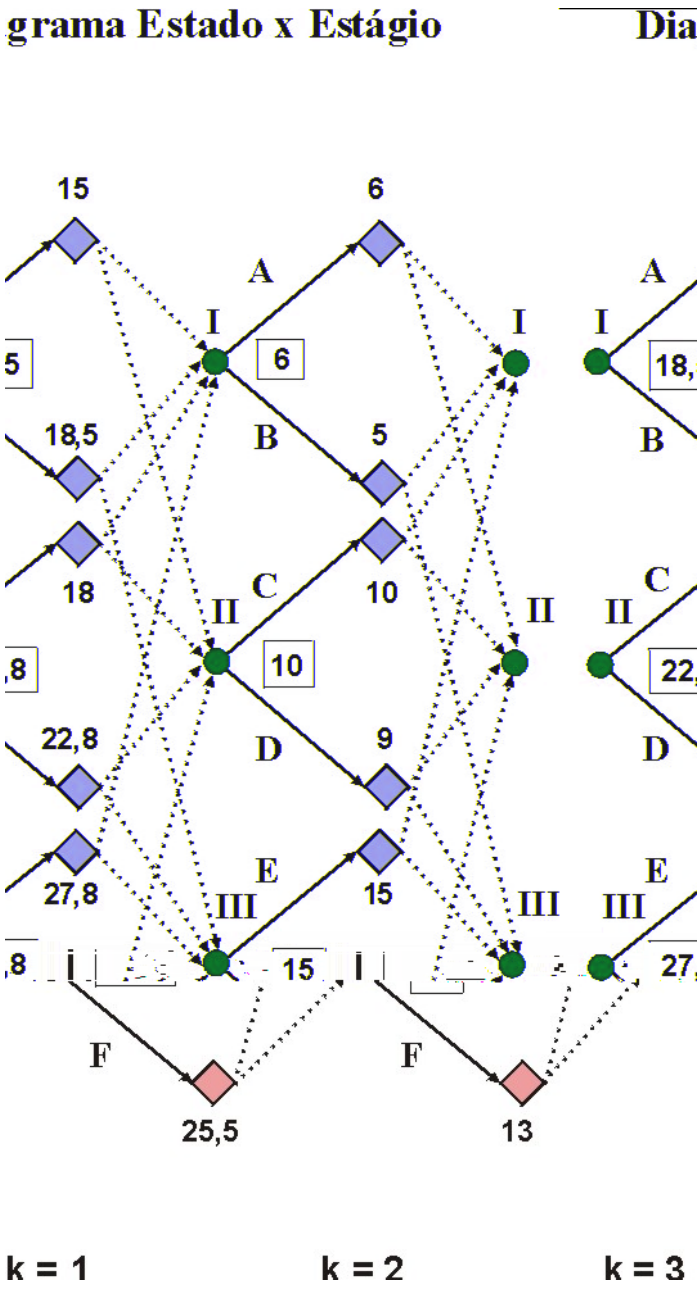
Se o jogador tomar a decisão de escolher a roleta A, tem probabilidade ¼ de continuar na região I, com ganho de 12, e uma chance de 3/4 de ir para a região II, com ganho de 4. Se ficar em I, sua próxima decisão será optar pela roleta A novamente, com ganho esperado de 6. Se for para a região II, sua próxima decisão será a roleta C, com ganho esperado de 10. Então, tomando a decisão A, o *ganho adicional esperado máximo* será:

$$\frac{1}{4}(12+6)+\frac{3}{4}(4+10)=4{,}5+10{,}5=15$$

Se tomar a decisão B, o seu *ganho adicional esperado máximo* será:

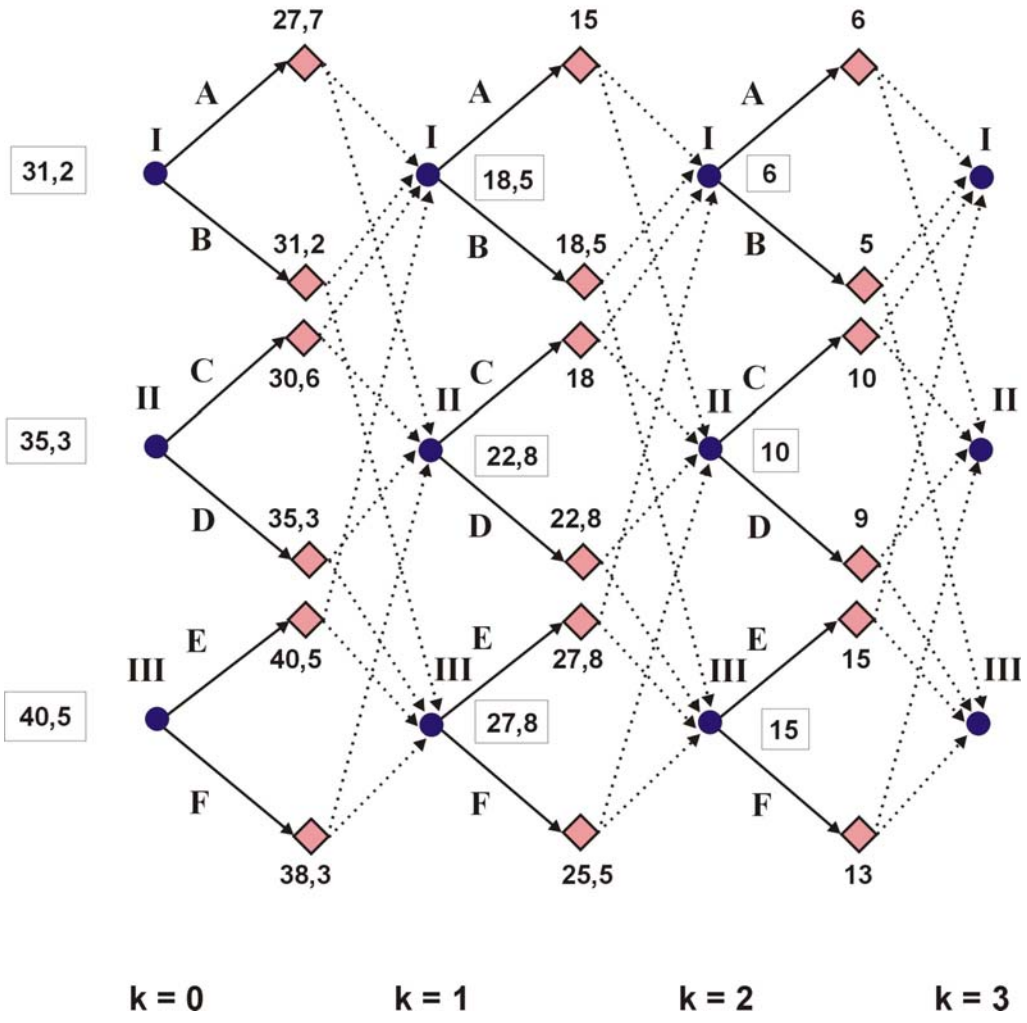
$$\frac{1}{6}(0+6)+\frac{5}{6}(6+15)=1{,}0+17{,}5=18{,}5$$

Portanto, se imediatamente antes da segunda jogada, a ficha estiver no *estado I*, a melhor decisão será a *roleta B* $(g^*(I,1)=18{,}5)$. Procedendo da mesma forma para as regiões II e III obteremos o *diagrama estado x estágio* parcial a seguir:



Aplicando o mesmo raciocínio para o estágio $k = 0$, ou seja, antes da primeira jogada, teremos o diagrama estado estágio correspondente ao problema.

Diagrama Estado x Estágio



Estratégia ótima:

$$g^*(I,0) = 31{,}2 \ , \ g^*(II,0) = 35{,}3 \text{ e } g^*(III,0) = 40{,}5$$

$$u^*(0) = (B,D,E) \ , \ u^*(1) = (B,D,E) \text{ e } u^*(2) = (A,C,E)$$

A estratégia ótima, ou conjunto de decisões ótimas para cada estágio, será então:

Estratégia Ótima		
Estágio	**Estado**	**Decisão**
$k = 0$	*I*	Roleta B
	II	Roleta D
	III	Roleta E
$k = 1$	*I*	Roleta B
	II	Roleta D
	III	Roleta E
$k = 2$	*I*	Roleta A
	II	Roleta C
	III	Roleta E

O ganho esperado ótimo $E[G]$, considerando que os estados no estágio inicial, $k_0 = k = 0$, são *igualmente prováveis* é dado por:

$$E[G] = \frac{g^*(I,0) + g^*(II,0) + g^*(III,0)}{3} = \frac{1}{3}(31{,}2 + 35{,}3 + 40{,}5) = 35{,}7$$

Estágio $k = 2$

$y(2)$	$u(2)$	$y(2+1)$ ... variável aleatória					$g^*(y(2),2)$	$\bar{u}(y(2),2)$
		y	p_j	c_j	$g^*(y,2+1)$	E		
I	A	I	¼	12	0	6	6	A
		II	¾	4	0			
	B	I	1/6	0	0	5		
		III	5/6	6	0			
II	C	I	½	10	0	10	10	C
		II	½	10	0			
	D	II	¼	0	0	9		
		III	¾	12	0			
III	E	I	¼	36	0	15	15	E
		III	¾	8	0			
	F	II	½	10	0	13		
		III	½	16	0			

Estágio $k = 1$

$y(1)$	$u(1)$	$y(1+1)$ ... variável aleatória					$g^*(y(1),1)$	$\bar{u}(y(1),1)$
		y	p_j	c_j	$g^*(y,1+1)$	E		
I	A	I	¼	12	6	15	18,5	B
		II	¾	4	10			
	B	I	1/6	0	6	18,5		
		III	5/6	6	15			
II	C	I	½	10	6	18	22,8	D
		II	½	10	10			
	D	II	¼	0	10	22,8		
		III	¾	12	15			
III	E	I	¼	36	6	27,8	27,8	E
		III	¾	8	15			
	F	II	½	10	10	25,5		
		III	½	16	15			

Estágio $k = 0$

$y(0)$	$u(0)$	$y(0+1)$ ... variável aleatória					$g^*(y(0),0)$	$\bar{u}(y(0),0)$
		y	p_j	c_j	$g^*(y,0+1)$	E		
I	A	I	¼	12	18,5	27,7	31,2	B
		II	¾	4	22,8			
	B	I	1/6	0	18,5	31,2		
		III	5/6	6	27,8			
II	C	I	½	10	18,5	30,6	35,3	D
		II	½	10	22,8			
	D	II	¼	0	22,8	35,3		
		III	¾	12	27,8			
III	E	I	¼	36	18,5	40,5	40,5	E
		III	¾	8	27,8			
	F	II	½	10	22,8	38,3		
		III	½	16	27,8			

4.5. Um Jogo de Cartas

Considere um jogo que consiste em sortear uma carta do baralho, observar seu valor, recolocar a carta de volta e embaralhar. Esta operação pode ser repetida mais *três vezes* se o jogador desejar. Em qualquer etapa o jogo pode ser interrompido e o jogador ganhará um valor proporcional ao da última carta mostrada.

O jogador terá um ganho de $\$10 \times (\text{Valor da Carta } k)$, onde $k = 1, 2, 3, 4$ é a etapa do jogo, sendo que o Ás = 1, Valete = 11, Dama = 12 e Rei = 13 e as demais cartas têm o valor de face normal, independente do naipe.

k=1 k=2 k=3

Jogo encerrado na terceira tentativa com ganho de $10 x 11 = $110

k=1 k=2 k=3 k=4

Jogo encerrado na quarta e última tentativa com ganho de $10 x 5 = $50

As regras básicas do jogo são, portanto, as seguintes:

- Os ganhos correspondem ao da última carta sorteada (carta de ordem k), não podendo o jogador em nenhuma hipótese, optar por uma carta anterior de maior valor;
- O jogador poderá optar pelo encerramento do jogo antes da última carta ($k = 4$) ser sorteada;
- A quarta carta sorteada ($k = 4$) define o valor do jogo e, necessariamente, o encerra.

Qual a *estratégia* que *maximiza o ganho esperado* do jogador ?

Solução

Os *estágios* $k = 0, 1, 2, 3, 4$ do jogo serão definidos por:

$$k = \begin{cases} 0 & \ldots\ldots \text{ antes do jogo iniciar} \\ 1, 2, 3, 4 & \ldots\ldots \text{ após concluir a } k-\textit{ésima} \text{ jogada} \end{cases}$$

Os *estados* $i = 1, 2, ..., 13$ representam os valores possíveis das cartas nos *estágios* $k = 1, 2, 3, 4$. No *estágio* $k = 0$ o jogo ainda não iniciou e, portanto, somente o *estado inicial* $i = 0$ se aplica. Nos demais *estágios* k o *estado* $i = 0$ se aplicará quando o jogo for encerrado no *estágio anterior* $k - 1$.

Só há duas *decisões admissíveis* em cada estado:

$$u(i) = \begin{cases} u_1 \text{ , continuar jogando} \\ u_2 \text{ , encerrar a partida} \end{cases}$$

Note que para o *estado* $i = 0$, nos estágios $k = 1, 2, 3, 4$, não atuará nenhuma decisão uma vez que o jogo já estará encerrado.

A tomada da decisão é influenciada pelo *estado* i e pelo *estágio* k. Por exemplo:

- Se $i = 1$ (Ás) e $k < 4$ a melhor decisão deverá ser prosseguir jogando ($u(1) = u_1$), pois, o valor da próxima carta será igual, na pior das hipóteses, não tendo o jogador nada a perder arriscando;

- Se $i \neq 1$ e $k < 4$ então a decisão não é evidente a não ser que $i = 13$ (Rei) e, neste caso, o jogador para com o maior ganho possível ($130), caso contrário, poderá ou não arriscar buscando um ganho maior.

Como, por hipótese, os eventos são *equiprováveis*, a probabilidade do sistema alcançar o *estado* j quando a *decisão* $u(i) = u_1$ atuar é dada por:

$$P\{j / u(i) = u_1\} = p_j = \frac{4}{52} = \frac{1}{13} \quad , \quad j = 1, 2, ..., 13$$

Esta distribuição de probabilidades independe do estágio considerado.

Se no *estado* i, do terceiro estágio (k = 3), a decisão $u(i) = u_1$ for aplicada, e o jogador receber a quarta carta, o *valor esperado do ganho* será:

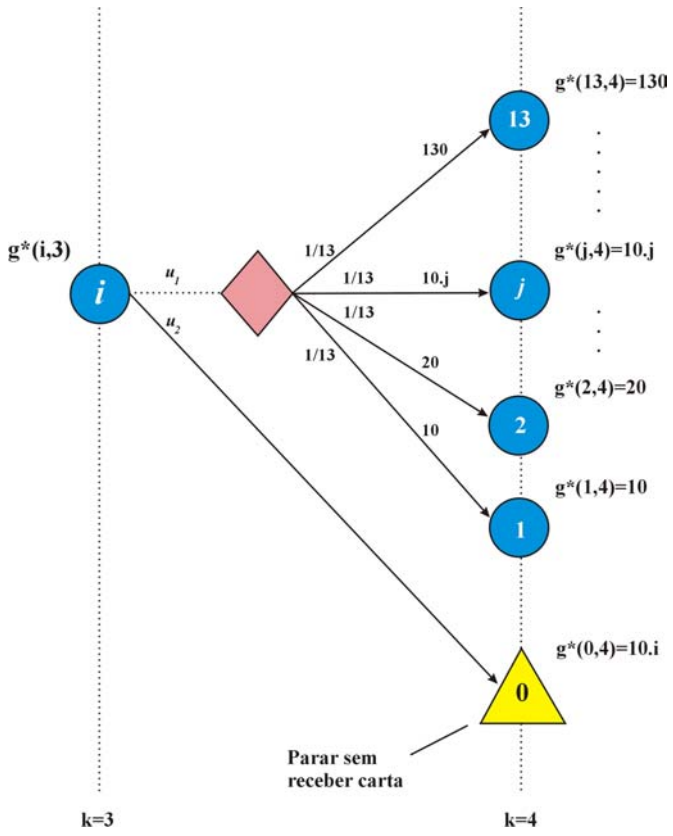
$$\sum_{j=1}^{13} E\left[j / u(i) = u_1\right] P\{\, j / u(i) = u_1\} = \sum_{j=1}^{13} 10\, j\; p_j = 10 \sum_{j=1}^{13} \frac{j}{13} = \frac{10}{13} \sum_{j=1}^{13} \mathrm{j} = 70\,,\ \mathrm{i} = 1,2,...,13$$

Entretanto, se a decisão for $u(i) = u_2$ (parar), o ganho será $\$10 \times$ i. Portanto, o *ganho esperado ótimo* $g^*(i,3)$, em cada estado *i* do *estágio* $k = 3$, é dado por:

$$g^*(i,3) = máximo\{\ 10 \times i\ ,\ 70\ \}\ ,\ i = 1, 2, ..., 13$$

$$\therefore \quad g^*(i,3) = \begin{cases} \$70 & ,\ \text{se}\ i \le 7 \\ \$10 \times i & ,\ \text{se}\ i \ge 8 \end{cases} \quad ,\ i = 1, 2, ..., 13$$

Segue-se que, concluido o *estágio* $k = 3$ e a terceira carta ter sido exibida, se $i \le 7$ a *decisão ótima* será $u^*(i) = u_1$, continuar jogando até a última etapa, pois, o maior ganho possível é \$70, enquanto para $i \ge 8$ a *decisão ótima* é $u^*(i) = u_2$, parar.



Esquema do Estágio $k = 3$

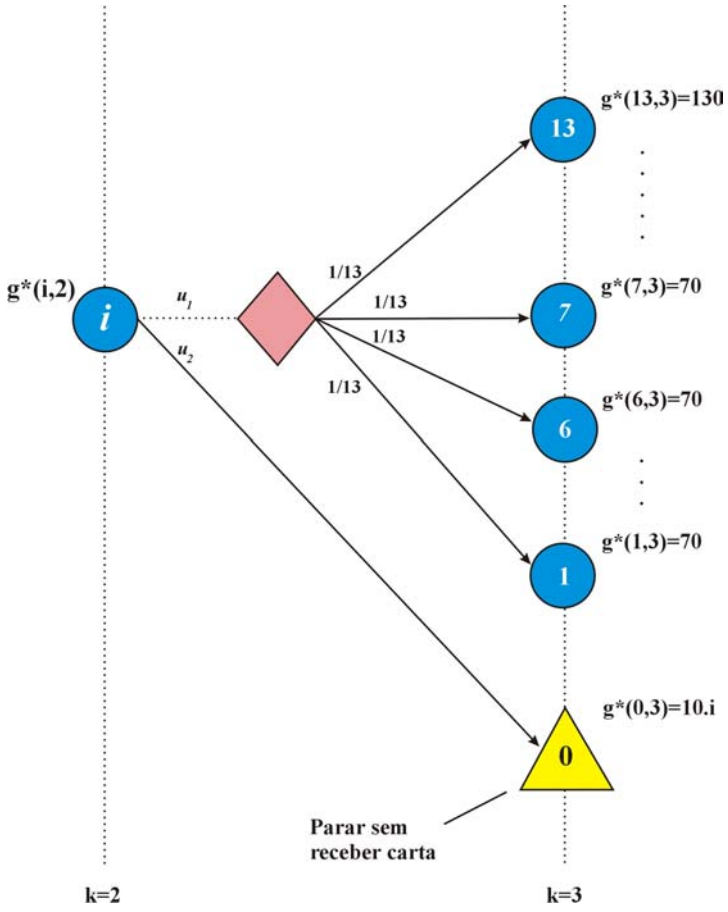
Para um *estado* i do *estágio* $k = 2$ se for aplicada a decisão $u(i) = u_1$ de continuar jogando, com probabilidade $p_j = 1/13$, a transição poderá levar aos estados $j = 8, 9, 10, 11, 12, 13$ (oito, nove, dez, Valete, Dama e Rei) com ganho de $\$10 \times j$. Porém, com probabilidade $\sum_{j=1}^{7} p_j = 7/13$, o estado resultante poderá ter *valor menor ou igual a 7*.

Considerando que, no estágio $k = 3$, a decisão ótima para todo estado $i \leq 7$ é $u^*(i) = u_1$ com $g^*(i,3) = \$\ 70$ enquanto para $i \geq 8$ será $u^*(i) = u_2$ com $g^*(i,3) = \$10 \times i$, o *ganho esperado ótimo* $g^*(i,2)$ para os estados $i = 1, 2, ..., 13$ do estágio $k = 2$ é dado por:

$$g^*(i,2) = máximo\left\{ \frac{7}{13}\sum_{j=1}^{7} g^*(j,3) + \frac{1}{13}\sum_{j=8}^{13} g^*(j,3)\ , 10 \times i \right\}, i = 1, 2, ..., 13$$

$$\therefore \quad g^*(i,2) = \begin{cases} \$86{,}154 & , \text{ se } i \leq 8 \\ \$10 \times i & , \text{ se } i \geq 9 \end{cases}$$

Portanto, se no *estágio* $k = 2$ tivermos uma carta $i \leq 8$ a *decisão ótima* é $u^*(i) = u_1$, ou seja, continuar jogando. Caso contrário, se $i \geq 9$ então $u^*(i) = u_2$, ou seja, encerrar o jogo.



Esquema do Estágio $k = 2$

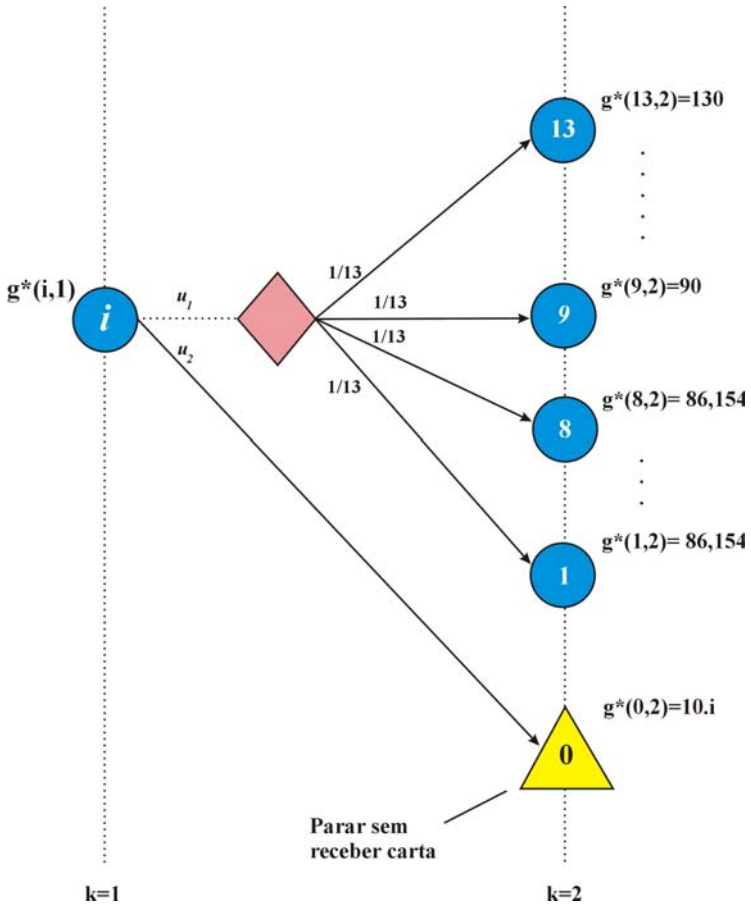
No *estágio* $k = 1$, estando o sistema no *estado* i, se for aplicada a decisão $u(i) = u_1$ de continuar jogando, com probabilidade $p_j = 1/13$ a transição poderá levar aos estados $j = 9, 10, 11, 12, 13$ (nove, dez, Valete, Dama e Rei) com ganho correspondente $\$10 \times j$. Entretanto, com probabilidade $\sum_{j=1}^{8} p_j = 8/13$ poderá ser uma *carta de menor valor do que 9*.

No *estágio* $k = 2$, para todo o *estado* $i \leq 8$ tem-se $g^*(i,2) = \$86{,}154$ enquanto para $i \geq 9$, $g^*(i,2) = \$10 \times i$, segue-se então que:

$$g^*(i,1) = máximo\left\{ \frac{8}{13}\sum_{j=1}^{8} g^*(j,2) + \frac{1}{13}\sum_{j=9}^{13} g^*(j,2) \ , \ 10 \times i \right\} =$$

$$= máximo\left\{ 95{,}325 \ , \ 10 \times i \right\} , \ i = 1, 2, ..., 13$$

$$\therefore \quad g^*(i,1) = \begin{cases} \$95{,}325 & , \text{ se } i \leq 9 \\ \$10 \times i & , \text{ se } i \geq 10 \end{cases}$$

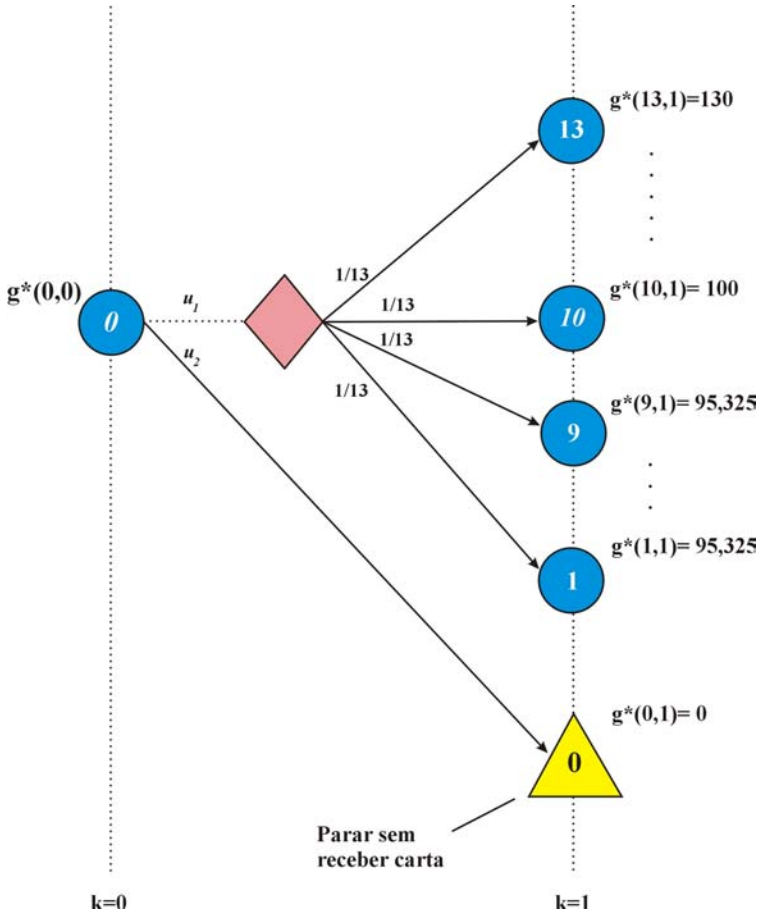


Esquema do Estágio $k = 1$

No estágio $k = 0$, antes de qualquer carta ter sido exibida, o único estado é $i = 0$, quando o jogo ainda não iniciou e nenhuma carta foi observada. A única decisão admissível é $u(i) = u_1$, ou seja, jogar recebendo carta que, com probabilidade $p_j = 1/13$ poderá ser $j = 10, 11, 12, 13$ (dez, Valete, Dama e Rei) e com probabilidade $\sum_{j=1}^{9} p_j = 9/13$ *carta de menor valor*. Então:

$$g^*(0,0) = \textit{máximo}\left\{ \frac{9}{13}\sum_{j=1}^{9} g^*(j,1) + \frac{1}{13}\sum_{j=10}^{13} g^*(j,1) \ , \ 10 \times 0 \right\} =$$

$$= \textit{máximo}\left\{ 101{,}379 \ , \ 0 \right\} = \$101{,}379$$

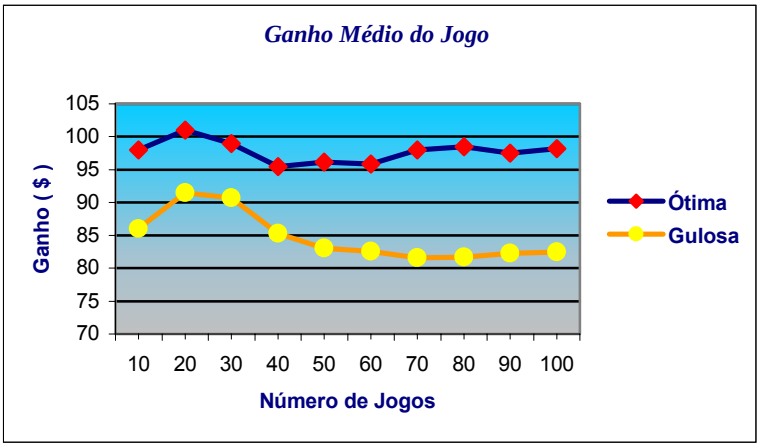


Esquema do Estágio $k = 0$

Ou seja, se as *decisões ótimas* atuarem sobre o sistema o *ganho esperado ótimo* do jogo será de ***$101,379***.

Estágio k	Estado i	Decisão Ótima $u^*(i)$
0	0	u_1
1	$1 \le i \le 9$ $i \ge 10$	u_1 u_2
2	$1 \le i \le 8$ $i \ge 9$	u_1 u_2
3	$1 \le i \le 7$ $i \ge 8$	u_1 u_2

A seguir apresenta-se o resultado do *ganho médio* obtido em uma simulação considerando 100 repetições do jogo utilizando a *estratégia ótima* ($98,20) ou uma *estratégia gulosa ou míope ($82,40)*, ou seja, continuar jogando até a última etapa na tentativa de maximizar o ganho.



4.6. Manufatura de Produto

Uma fábrica aceitou encomenda para fornecer *um único item* de determinado produto. O cliente estabeleceu requisitos severos de qualidade de tal forma que o fabricante pode ter que produzir mais de um item até obter um considerado aceitável pelo cliente.

O fabricante estima, em função das características do produto e do processo produtivo que pode ser executado em lotes, que a probabilidade de produzir um item compatível com os requisitos do cliente é de 50% não havendo possibilidade de reparar um item defeituoso que deve então ser imediatamente descartado.

Portanto, a variável aleatória $\boldsymbol{x}$ que representa o número de items *defeituosos* em um lote de tamanho $\boldsymbol{u}$ terá *distribuição binomial* com parâmetros $(\boldsymbol{u}, \boldsymbol{p})$, ou seja:

$$\boldsymbol{P}\{\boldsymbol{X} = \boldsymbol{x}\} = \binom{\boldsymbol{u}}{\boldsymbol{x}} \boldsymbol{p}^{\boldsymbol{x}} (1-\boldsymbol{p})^{\boldsymbol{u}-\boldsymbol{x}}, \ \boldsymbol{x} = 0, 1, 2, ..., \boldsymbol{u}$$

A probabilidade de produção de um lote com *todos os items defeituosos* considerando $\boldsymbol{p} = \frac{1}{2}$ é dada por $\boldsymbol{P}\{\boldsymbol{X} = \boldsymbol{u}\} = \left(\frac{1}{2}\right)^{\boldsymbol{u}}$ e com *pelo menos um item aceitável* $\boldsymbol{P}\{\boldsymbol{X} < \boldsymbol{u}\} = 1 - \left(\frac{1}{2}\right)^{\boldsymbol{u}}$.

Os custos unitários de produção são estimados em \$100 por item (mesmo quando defeituoso). Items produzidos em excesso não tem qualquer valor para a fábrica. Há ainda um custo fixo de \$300 para preparação do processo de produção em lotes independente do tamanho do lote. Se a inspeção revelar que *um lote inteiro foi recusado*, ou seja, *não houver um único item*

aceitável, um custo adicional de \$300 será imputado. O fabricante tem tempo disponível para produção de apenas três lotes na tentativa de obtenção de item aceitável. Se um item aceitável não for obtido ao final da terceira tentativa de produção, os custos do fabricante devido a perda de receita e penalidades contratuais serão de \$1,600.

O objetivo é determinar a política relativa ao tamanho do lote em cada uma das três tentativas de produção que *minimize o custo total esperado do fabricante*.

Na formulação do modelo de programação dinâmica os estágios serão as "corridas de produção". Portanto:

$$\begin{cases} k = 0 & \text{... antes da primeira corrida} \\ k = j & \text{... após a j-ésima corrida, j = 1, 2, 3} \end{cases}$$

Como só há necessidade de produzir um item aceitável, o estado do sistema pode ser representado por uma variável bivalente y_j tal que:

$$y_j = \begin{cases} 1, \text{ se nenhum item aceitável foi produzido} \\ 0, \text{ caso contrário} \end{cases}$$

A variável de decisão u_j será definida como o tamanho do lote na corrida de ordem $j = 1, 2, 3$. O número de items defeituosos r_j em uma "corrida" de tamanho u_j é, neste caso, uma *variável aleatória* com distribuição binomial com parâmetros (u_j, p).

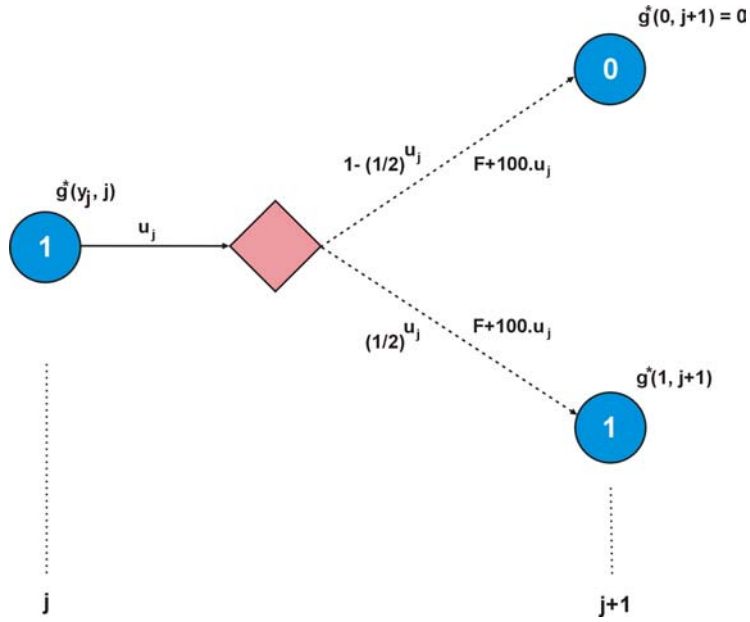
- Probabilidade de nenhum item ser aceitável:

$$\boldsymbol{P}\{\boldsymbol{r_j} = 0\} = \binom{\boldsymbol{u_j}}{0} \boldsymbol{p}(1-\boldsymbol{p})^{\boldsymbol{u_j}} = \left(\frac{1}{2}\right)^{\boldsymbol{u_j}}$$

- Probabilidade de pelo menos um item aceitável:

$$\boldsymbol{P}\{\boldsymbol{r_j} \geq 1\} = 1 - \boldsymbol{P}\{\boldsymbol{r_j} = 0\} = 1 - \left(\frac{1}{2}\right)^{\boldsymbol{u_j}}$$

O esquema da transição de estado com o componente aleatório pode ser representado como a seguir:



A *equação recursiva de otimalidade* é dada por:

$$\boldsymbol{g}^*(1,\boldsymbol{j}) = \underset{\boldsymbol{u_j} = 0,1,2,\ldots}{\boldsymbol{mínimo}} \{ (1-\left(\frac{1}{2}\right)^{\boldsymbol{u_j}}) \left[\boldsymbol{F} + 100\boldsymbol{u_j} + \boldsymbol{g}^*(0,\boldsymbol{j}+1)\right] + \left(\frac{1}{2}\right)^{\boldsymbol{u_j}} \left[\boldsymbol{F} + 100\boldsymbol{u_j} + \boldsymbol{g}^*(1,\boldsymbol{j}+1)\right] \}$$

com $g^*(0, j) = 0,\ g^*(1,3) = 1{,}600$ e

$$F = \begin{cases} 300\ , & u_j > 0 \\ \\ 0\ , & u_j = 0 \end{cases} \qquad \forall\ j$$

Assumindo que já foram realizadas duas "corridas de produção" sem que se obtivesse nenhum item aceitável, caso contrário o processo estaria encerrado, deve-se decidir o tamanho do lote de produção para a terceira corrida u_3.

Caso não se obtenha nenhum item aceitável haverá uma penalidade de \$1,600, isto é, $g^*(1,3) = 1{,}600$. A *equação recursiva de otimalidade* é dada por:

$$g^*(1,2) = \underset{u_3 = 0,1,2,\ldots}{\textbf{mínimo}} \{ F + 100u_3 + \left(\frac{1}{2}\right)^{u_3} g^*(1,3) \}$$

com $g^*(1,3) = 1{,}600$ e $F = \begin{cases} 0, & \text{se } u_3 = 0 \\ \\ 300, & \text{se } u_3 > 0 \end{cases}$

Os quadros e o diagrama estado estágio a seguir resumem o processo de cálculos.

Estágio k = 2

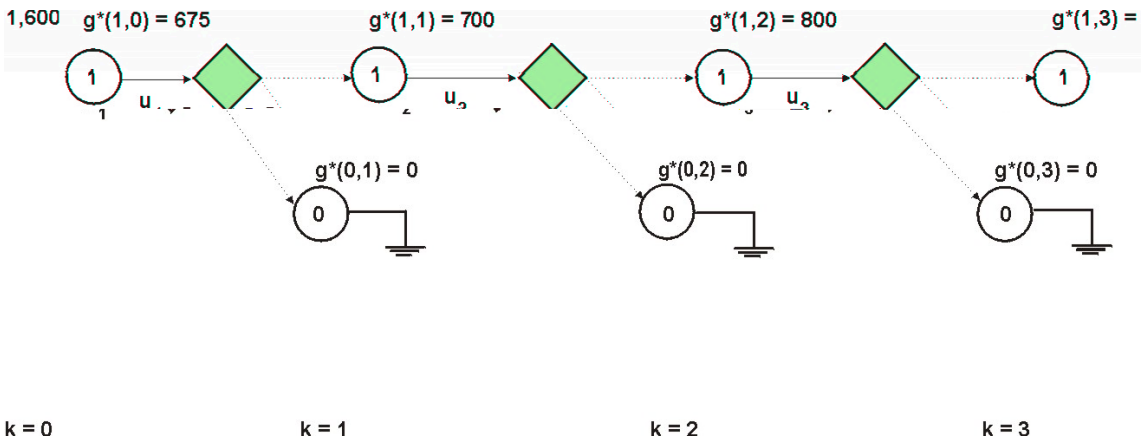
y \ u_3	$F+100u_3+1{,}600\left(\frac{1}{2}\right)^{u_3}$							$g^*(1,2)$	u_3^*
	0	1	2	3	4	5	6		
0	0							0	0
1	1,600	1,200	900	800	800	850	925	800	3 ou 4

Estágio k = 1

y \ u_2	$F+100u_2+\left(\frac{1}{2}\right)^{u_2} g^*(1,2)$							$g^*(1,1)$	u_2^*
	0	1	2	3	4	5	6		
0	0							0	0
1	800	800	700	700	750	825	912	700	2 ou 3

Estágio k = 0

y \ u_1	$F+100u_1+\left(\frac{1}{2}\right)^{u_1} g^*(1,1)$							$g^*(1,0)$	u_1^*
	0	1	2	3	4	5	6		
0	0							0	0
1	700	750	675	687	744	822	911	675	2



A estratégia ótima será, portanto, $\boldsymbol{u}_1^* = 2, \boldsymbol{u}_2^* = 2 \text{ ou } 3$ e $\boldsymbol{u}_3^* = 3 \text{ ou } 4$, ou seja, produzir dois items (2) na primeira corrida e, se nenhum aceitável for obtido, produzir dois (2) ou três (3) items na segunda corrida, caso nenhum aceitável for obtido então produzir três (3) ou quatro (4) items na terceira e última corrida de produção. O *custo esperado total* desta estratégia será $ 675.

5. Programação Dinâmica Probabilística com Horizonte Ilimitado

5.1. Conceito

Nos modelos determinísticos quando uma decisão atua sobre o sistema as *mudanças de estado ocorrem de maneira previsível* sem envolver nenhuma incerteza. Por esta razão, não estando presentes fatores aleatórios, adotada uma seqüência de decisões a partir de um estado inicial (política), as transições de estado bem como os custos (lucros) elementares correspondentes são conhecidos de forma exata.

Os princípios da *Programação Dinâmica*, como visto no caso da *Programação Dinâmica Probabilística com Horizonte Limitado*, podem ser estendidos para modelos estocásticos com horizonte ilimitado, permitindo que as transições de estado envolvam componentes estocásticas. A estratégia neste caso é adotar como critério a minimização (maximização) dos custos (lucros) esperados em presença das incertezas envolvidas.

No caso da *Programação Dinâmica Probabilística com Horizonte Ilimitado*, *as distribuições de probabilidade variam de estágio a estágio*. Estaremos, portanto, lidando com sistemas dinâmicos para os quais decisões seqüenciais estarão vinculadas a diferentes processos aleatórios. Na modelagem destes sistemas dinâmicos será adotada a característica fundamental da evolução segundo *Cadeias de Markov*, isto é, *políticas estacionárias* onde as transições de estado só dependem do estado em que se encontra o sistema e da decisão a ser aplicada, ou seja, uma transição de estado não dependerá do passado.

Serão tratados nesta abordagem apenas os problemas com as seguintes características:

- Número de *estados viáveis finito* $i = 1, 2, ..., m$;
- Número de *estratégias estacionárias finito* U_s, $s = 1, 2, ..., q$;
- Uma vez definida uma estratégia estacionária U_s, se o sistema se encontrar em um estado i, *a probabilidade de que ele passe ao estado* j *não dependerá dos estados anteriormente assumidos*, *e nem do estágio em que esta transição estiver ocorrendo*;
- O custo elementar associado a cada transição de estado *não dependerá do estágio em que ela ocorre*.

Estas hipóteses caracterizam a *evolução estocástica do sistema*, ou seja, os estados assumidos no processo, como uma *Cadeia de Markov Homogênea*. Portanto, a cada estratégia estacionária U_s haverá uma matriz de transição de estado $\pi\ (U_s)$ da forma:

$$\pi\ (U_s) = \begin{bmatrix} p_{11}^s & p_{12}^s & .... & p_{1m}^s \\ p_{21}^s & p_{22}^s & .... & p_{2m}^s \\ \vdots & \vdots & & \vdots \\ p_{m1}^s & p_{m2}^s & .... & p_{mm}^s \end{bmatrix}$$

Será adotado o critério da *minimização (maximização) do valor esperado da série de custos (lucros) elementares* para obtenção das estratégias estacionárias ótimas.

5.2. Critério do Valor Atual Esperado

O problema consiste em resolver o sistema de equações extremas:

$$g^*(i) = \underset{\substack{u \in U(i) \\ j = r(i,u)}}{mínimo} E\left[\ f(i,u) + \alpha . g^*(j)\ \right] \quad , \quad i = 1, 2, ..., m$$

onde $0 \leq \alpha < 1$ é a *taxa de descontos*, $f(i,u)$ e $j = r(i,u)$ representam respectivamente a *variável aleatória custo elementar esperado* e a *variável aleatória novo estado* quando o sistema se encontrar no estado i e a decisão $u \in U(i)$ for tomada.

Se $p_j(i,u)$ e $c_j(i,u)$ forem, respectivamente, a probabilidade do sistema evoluir do estado i para o estado j e o custo elementar quando a decisão $u \in U(i)$ for aplicada, as *equações de otimalidade* podem ser escritas como:

$$g^*(i) = \underset{\substack{u \in U(i) \\ j = r(i,u)}}{mínimo} \left\{ \sum_j c_j(i,u).p_j(i,u) + \alpha . \sum_j g^*(j).p_j(i,u) \right\}, \ i = 1, 2, ..., m$$

$$\sum_j p_j(i,u) = 1, \quad f(i,u) = \sum_j c_j(i,u).p_j(i,u), \ u \in U(i) \text{ e } i = 1, 2, ..., m.$$

O valor $g^*(i)$ representa o valor presente (atual) esperado mínimo de iniciar no estado i e utilizar uma estratégia ótima sobre um horizonte ilimitado. Note que a hipótese adotada de que o sistema evoluirá segundo uma *Cadeia de Markov Homogênea* implica em que as probabilidades $p_j(i,u)$ dependem apenas de i e u e não da história do sistema em relação à decisões e estados anteriores.

A justificativa para a validade desta equação funcional é intuitiva, sendo que a prova formal pode ser encontrada em Ross (1983). Se o sistema em exame se encontrar inicialmente no estado i e a decisão a ser aplicada é $u \in U(i)$ levando ao estado j, então, de imediato, há um custo esperado associado a esta decisão $f(i,u)$. Como na evolução dinâmica do sistema há ainda um número infinito de estágios a partir do estado j, segundo o *Princípio da Otimalidade de Bellman*, deve ser acrescentado um custo descontado ótimo $g^*(j)$. Além disto, como o estado j será alcançado com probabilidade $p_j(i,u)$ e devemos multiplicar $g^*(j)$ pela taxa de descontos α para obter o valor presente esperado no estado i. Segue-se que

$$f(i,u) + \alpha \sum_j p_j(i,u).g^*(j)$$

será o *valor atual (presente) esperado mínimo* quando o estado inicial for i e a decisão que atua sobre o sistema for $u \in U(i)$. Conseqüentemente, com probabilidade 1, a decisão ótima no estado i é aquela cujo valor atual esperado é mínimo, o que nos leva a equação de otimalidade.

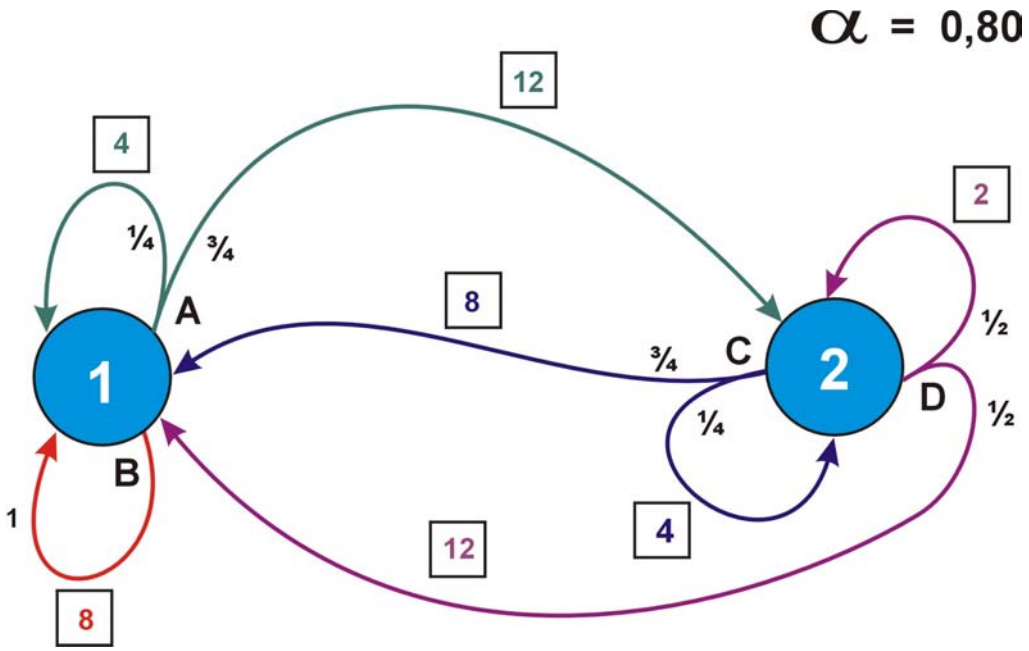
O sistema de equações de otimalidade, como no caso determinístico, pode ser resolvido por dois métodos:

- ***Aproximações no Espaço dos Critérios***
- ***Aproximações no Espaço de Políticas***

5.3. Método das Aproximações no Espaço dos Critérios

As características deste método foram abordadas no problema determinístico e permitem passar diretamente para a resolução de um exemplo.

Exemplo



O sistema, representado esquemáticamente no diagrama anterior, pode se encontrar a cada estágio, nos estados **1** e **2**. Em cada estado, podem ser tomadas uma dentre duas decisões: **A** e **B** para o **estado 1** e **C** e **D** para o **estado 2**. No diagrama estão representados os custos elementares bem como as probabilidades das respectivas transições de estado.

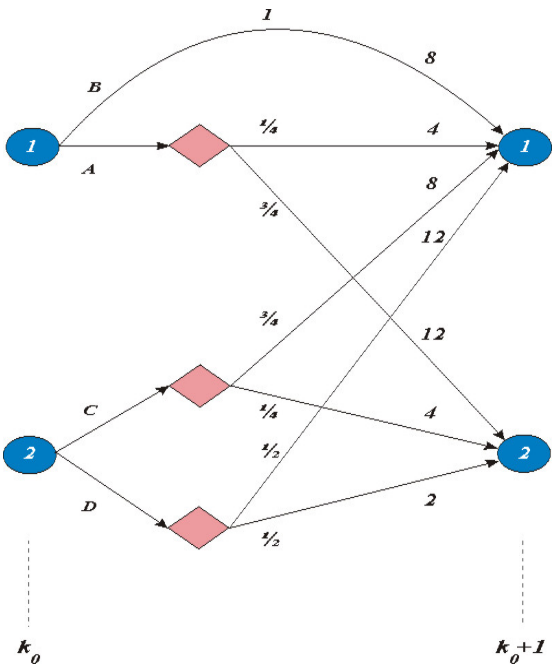
Só há quatro (04) estratégias estacionárias: **(A,C)**, **(A,D)**, **(B,C)** e **(B,D)**. A cada uma das políticas corresponderá uma matriz de transição de estado. Por exemplo, para a política **(A,C)** temos:

$$\pi\,(\mathrm{A,C}) = \begin{bmatrix} 1/4 & 3/4 \\ 3/4 & 1/4 \end{bmatrix}$$

A taxa de descontos considerada para efeito de cálculo do *valor atual esperado* é $\alpha = 0{,}80$.

Utilizaremos como estimativas, tanto para $g^*(1)$ quanto para $g^*(2)$, o valor **40** e para regra de parada $\Delta_t(i) = \left|g_t^*(i) - g_{t-1}^*(i)\right| < 10^{-2},\ i = 1, 2$ em qualquer iteração t.

Solução:



Observando o esquema do processo, representado acima, as equações recursivas de otimalidade correspondentes são:

$$g^*(1) = mínimo\{\underbrace{(10 + \alpha(\frac{1}{4}g^*(1) + \frac{3}{4}g^*(2))}_{A}\ ,\ \underbrace{(8 + \alpha\, g^*(1))}_{B}\}$$

$$g^*(2) = mínimo\{\underbrace{(7 + \alpha(\frac{3}{4}g^*(1) + \frac{1}{4}g^*(2))}_{C}, \underbrace{(7 + \alpha(\frac{1}{2}g^*(1) + \frac{1}{2}g^*(2))}_{D}\}$$

1ª Iteração

$$g_1^*(1) = mínimo\{(\frac{1}{4}\times 4 + \frac{3}{4}\times 12 \ + \ 0{,}8(\frac{1}{4}\times 40 + \frac{3}{4}\times 40)) \, , (8 + 0{,}8\times 40)\} =$$

$$= mínimo\{(10 + 0{,}8\times 40) \ , \ (8 + 0{,}8\times 40)\} = 40 \ \ldots \text{ Decisão B}$$

$$g_1^*(2) = mínimo\{(\frac{3}{4}\times 8 + \frac{1}{4}\times 4 \ + \ 0{,}8(\frac{3}{4}\times 40 + \frac{1}{4}\times 40)) \, , (\frac{1}{2}\times 2 + \frac{1}{2}\times 12 +$$

$$0{,}8(\frac{1}{2}\times 40 + \frac{1}{2}\times 40))\} =$$

$$= mínimo\{(7 + 0{,}8\times 40) \ , \ (7 + 0{,}8\times 40)\} = 39 \ \ldots \text{ Decisão C ou D}$$

2ª Iteração

$$g_2^*(1) = mínimo\{(10 + \ 0{,}8(\frac{1}{4}\times 40 + \frac{3}{4}\times 39)) \, , (8 + 0{,}8\times 40)\} =$$

$$= mínimo\{41{,}4 \, , 40\} = 40 \qquad \ldots \text{ Decisão B}$$

$$g_2^*(2) = mínimo\{(7 + \ 0{,}8(\frac{3}{4}\times 40 + \frac{1}{4}\times 39)) \, , (7 + 0{,}8(\frac{1}{2}\times 40 + \frac{1}{2}\times 39))\} =$$

$$= mínimo\{38{,}8 \, , 38{,}6\} = 38{,}6 \ \ldots \text{ Decisão D}$$

3ª Iteração

$$g_3^*(1) = mínimo\{(10 + \ 0{,}8(\frac{1}{4}\times 40 + \frac{3}{4}\times 38{,}6)) \, , (8 + 0{,}8\times 40)\} =$$

$$= mínimo\{41{,}16 \, , 40\} = 40 \qquad \ldots \text{ Decisão B}$$

$$g_3^*(2) = mínimo\{(7 + 0{,}8(\frac{3}{4}\times 40 + \frac{1}{4}\times 38{,}6)) \, , (7 + 0{,}8(\frac{1}{2}\times 40 + \frac{1}{2}\times 38{,}6))\} =$$

$$= mínimo\{38{,}72 \, , 38{,}44\} = 38{,}44 \ \ldots \text{ Decisão D}$$

4ª Iteração

$$g_4^*(1) = mínimo\{(10 + 0{,}8(\frac{1}{4}\times 40 + \frac{3}{4}\times 38{,}44)), (8 + 0{,}8\times 40)\} =$$
$$= mínimo\{41{,}064 , 40\} = 40 \qquad \text{... Decisão B}$$

$$g_4^*(2) = mínimo\{(7 + 0{,}8(\frac{3}{4}\times 40 + \frac{1}{4}\times 38{,}44) , (7 + 0{,}8(\frac{1}{2}\times 40 + \frac{1}{2}\times 38{,}44))\} =$$
$$= mínimo\{38{,}688 , 38{,}376\} = 38{,}376 \text{ ... Decisão D}$$

5ª Iteração

$$g_5^*(1) = mínimo\{(10 + 0{,}8(\frac{1}{4}\times 40 + \frac{3}{4}\times 38{,}376)), (8 + 0{,}8\times 40)\} =$$
$$= mínimo\{41{,}0256 , 40\} = 40 \qquad \text{... Decisão B}$$

$$g_5^*(2) = mínimo\{(7 + 0{,}8(\frac{3}{4}\times 40 + \frac{1}{4}\times 38{,}376) , (7 + 0{,}8(\frac{1}{2}\times 40 + \frac{1}{2}\times 38{,}376))\} =$$
$$= mínimo\{38{,}6752 , 38{,}3504\} = 38{,}3504 \text{ ... Decisão D}$$

6ª Iteração

$$g_6^*(1) = mínimo\{(10 + 0{,}8(\frac{1}{4}\times 40 + \frac{3}{4}\times 38{,}3504)), (8 + 0{,}8\times 40)\} =$$
$$= mínimo\{41{,}0102 , 40\} = 40 \qquad \text{... Decisão B}$$

$$g_6^*(2) = mínimo\{(7 + 0{,}8(\frac{3}{4}\times 40 + \frac{1}{4}\times 38{,}3504) , (7 + 0{,}8(\frac{1}{2}\times 40 + \frac{1}{2}\times 38{,}3504))\} =$$
$$= mínimo\{38{,}6700 , 38{,}3402\} = 38{,}3402 \text{ ... Decisão D}$$

7ª Iteração

$$g_7^*(1) = mínimo\{(10 + 0{,}8(\frac{1}{4}\times 40 + \frac{3}{4}\times 38{,}3402)), (8 + 0{,}8\times 40)\} =$$
$$= mínimo\{41{,}0041, 40\} = 40 \quad \text{... Decisão B}$$

$$g_7^*(2) = mínimo\{(7 + 0{,}8(\frac{3}{4}\times 40 + \frac{1}{4}\times 38{,}3402), (7 + 0{,}8(\frac{1}{2}\times 40 + \frac{1}{2}\times 38{,}3402))\} =$$
$$= mínimo\{38{,}6680, 38{,}3402\} = 38{,}3361 \text{ ... Decisão D}$$

Como a regra de parada atuou, ou seja,

$$\left|g_7^*(i) - g_6^*(i)\right| < 10^{-2},\ i = 1, 2$$

a *estratégia ótima* é **(B,D)** com valores ótimos

$$g^*(1) = 40 \text{ e } g^*(2) = 38{,}33.$$

A evolução das iterações é resumida no quadro a seguir:

Algoritmo da Aproximações no Espaço dos Critérios						
Iteração t	$g^*_t(1)$	$\Delta_t(1)$	$u^*(1)$	$g^*_t(2)$	$\Delta_t(2)$	$u^*(2)$
1	40	0	B	39	1	C ou D
2	40	0	B	38,6	0,4	D
3	40	0	B	38,44	0,16	D
4	40	0	B	38,376	0,064	D
5	40	0	B	38,3504	0,0256	D
6	40	0	B	38,3402	0,0102	D
7	**40**	**0**	**B**	**38,3361**	**0,0041**	**D**

5.4. Método das Aproximações no Espaço das Políticas

Como no método anterior vamos passar diretamente ao exemplo adotando, na solução dos sistemas de equações lineares, uma precisão de 10^{-1}.

1ª Iteração

Escolhemos, inicialmente, a *estratégia* **(A,C),** ou seja**,**

$$\bar{u}_0(1) = A \text{ e } \bar{u}_0(2) = C$$

obtendo o seguinte *sistema de equações lineares*:

$$\begin{cases} \bar{g}_1(1) = (\frac{1}{4}\times 4 + \frac{3}{4}\times 12) + 0{,}8(\frac{1}{4}\times \bar{g}_1(1) + \frac{3}{4}\bar{g}_1(2)) \\ \bar{g}_1(2) = (\frac{3}{4}\times 8 + \frac{1}{4}\times 4) + 0{,}8(\frac{3}{4}\times \bar{g}_1(1) + \frac{1}{4}\times \bar{g}_1(2)) \end{cases}$$

cuja solução é $\bar{g}_1(1) = 43{,}5$ e $\bar{g}_1(2) = 41{,}5$.

Testando estes valores nas *equações de otimalidade*:

$$\textit{mínimo}\{\, 43{,}5\,,\, (8 + 0{,}8\times 43{,}5)\,\} = 42{,}8 < \bar{g}_1(1) \Rightarrow \bar{u}_2(1) = B$$

$$\textit{mínimo}\{\, 41{,}5\,,\, ((\frac{1}{2}\times 2 + \frac{1}{2}\times 12) + 0{,}8(\frac{1}{2}\times 43{,}5 + \frac{1}{2}\times 41{,}5))\,\} = 41{,}5 = \bar{g}_1(2)$$

A nova *política admissível* passa a ser: $\bar{u}_2(1) = B$ e $\bar{u}_2(2) = C$.

2ª Iteração

Com a nova *estratégia (política admissível)* o sistema de equações passa a ser o seguinte:

$$\begin{cases} \bar{g}_2(1) = (1 \times 8) + 0{,}8 \times \bar{g}_2(1) \\ \bar{g}_2(2) = (\frac{3}{4} \times 8 + \frac{1}{4} \times 4) + 0{,}8(\frac{3}{4} \times \bar{g}_2(1) + \frac{1}{4} \times \bar{g}_2(2)) \end{cases}$$

Resolvendo o sistema de equações lineares temos:

$$\bar{g}_2(1) = 40 \quad \text{e} \quad \bar{g}_2(2) = 38{,}8$$

Testando estes valores nas *equações de otimalidade*:

$$\textit{mínimo}\{\ ((\frac{1}{4} \times 4 + \frac{3}{4} \times 12) + 0{,}8(\frac{1}{4} \times 40 + \frac{3}{4} \times 38{,}8))\ ,\ 40\ \} = 40 = \bar{g}_2(1)$$

$$\textit{mínimo}\{\ 38{,}8\ ,\ ((\frac{1}{2} \times 2 + \frac{1}{2} \times 12) + 0{,}8(\frac{1}{2} \times 40 + \frac{1}{2} \times 38{,}8))\ \} = 38{,}5 < \bar{g}_2(2)$$

$$\Rightarrow \bar{u}_3(2) = D$$

A nova *estratégia* passa a ser $(\bar{u}_3(1) = B\ ,\ \bar{u}_3(2) = D)$.

3ª Iteração

Com a nova *estratégia (política admissível)* o sistema de equações passa a ser:

$$\begin{cases} \bar{g}_3(1) = (1 \times 8) + 0{,}8 \times \bar{g}_3(1) \\ \bar{g}_3(2) = (\frac{1}{2} \times 2 + \frac{1}{2} \times 12) + 0{,}8(\frac{1}{2} \times \bar{g}_3(1) + \frac{1}{2} \times \bar{g}_3(2)) \end{cases}$$

Resolvendo o sistema de equações lineares temos:

$$\bar{g}_3(1) = 40 \quad \text{e} \quad \bar{g}_3(2) = 38{,}35$$

Testando estes valores nas *equações de otimalidade*:

$$\textit{mínimo}\{\ ((\frac{1}{4} \times 4 + \frac{3}{4} \times 12) + 0{,}8(\frac{1}{4} \times 40 + \frac{3}{4} \times 38{,}5))\ ,\ 40\ \} = 40 = \bar{g}_3(1)$$

$$\textit{mínimo}\{\ ((\frac{3}{4} \times 8 + \frac{1}{4} \times 4) + 0{,}8(\frac{3}{4} \times 40 + \frac{1}{4} \times 38{,}5)\ ,\ 38{,}35)\ \} = 38{,}35 = \bar{g}_3(2)$$

Com a precisão adotada na solução das equações lineares, neste caso (10^{-1}), o algoritmo termina obtendo em três iterações a *estratégia ótima efetiva* $(u^*(1) = B\ ,\ u^*(2)) = D$ e com os *valores presentes esperados* $g^*(1) = 40$ e $g^*(2) \approx 38{,}33$.

6. Referências Bibliográficas

ARDUINO, A., *Programação Dinâmica*, PDD-1/72, Universidade Federal do Rio de Janeiro, COPPE/UFRJ, junho 1972.

BATHER, J. A., *Decision Theory: An Introduction to Dynamic Programming and Sequential Decisions*, John Wiley and Sons, 2000.

BERTSEKAS, D. P., *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall Inc., Englewood Cliffs, NJ, 1987.

BERTSEKAS, D. P., *Dynamic Programming and Optimal Control*, 2nd Edition, Vols. I and II, Athena Scientific, 2001.

BRONSON, R. and NAADIMUTHU, G., *Operations Research*, Second Edition, Schaum's Outline Series, McGraw-Hill, 1997.

CORMEN, T. H., LEISERSON, C. E. and RIVEST, R. L., *Introduction to Algorithms*, Second Edition, McGraw-Hill Book Company, 2001.

DREYFUS, S. E. and LAW, A. M., *The Art and Theory of Dynamic Programming*, Academic Press, 1977.

EDGAR, T. F. and HIMMELBLAU, D. M., *Optimization of Chemical Processes*, McGraw-Hill Book Company, 1989.

EDMONDS, J., *Matroids an the Greedy Algorithm*, Mathematical Programming, (1) 127-136, 1971.

GOLDBARG, M. C. e LUNA, H. P., *Otimização Combinatória e Programação Linear – Modelos e Algoritmos*, Editora Campus, 2000.

HILLIER, F. S. and LIEBERMAN, G. J., *Introduction to Operations Research*, Seventh Edition, McGraw-Hill, 2001.

MITTEN, L.G., *Composition Principles for Synthesis of Optimal Multistage Processes*, Operations Research, 12, 610-619, 1964.

NEMHAUSER, G. L., *Introduction to Dynamic Programming*, John Wiley and Sons, New York, 1966.

PUTERMAN, M. L., *Markov Decision Processes: Discrete Stochastic Programming*, Wiley-Interscience, 1994.

RAGSDALE, C. T., *Spreadsheet Modeling and Decision Analysis*, Third Edition, South-Western College Publishing, 2001.

ROSS, S. M., *Introduction to Stochastic Dynamic Programming*, Academic Press, New York, 1983.

ROSS, S. M., *Introduction to Probability Models*, 7th Edition, Academic Press, 2000.

WAGNER, H. M., *Principles of Operations Research with Applications to Managerial Decisions*, Prentice-Hall Inc., 1969.