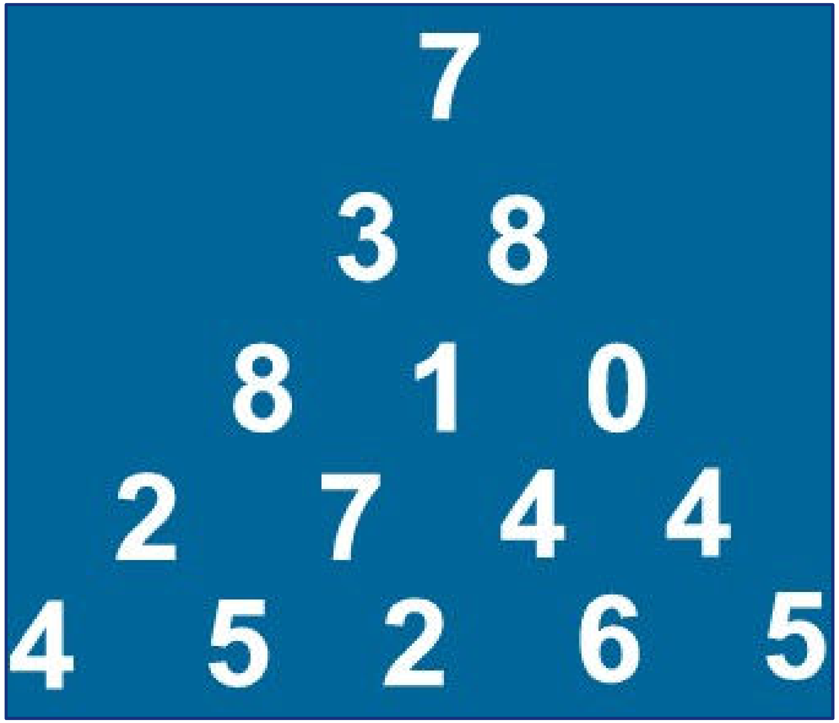
Pirâmide de Números

- Problema clássico das **Olimpíadas Internacionais de Informática de 1994**

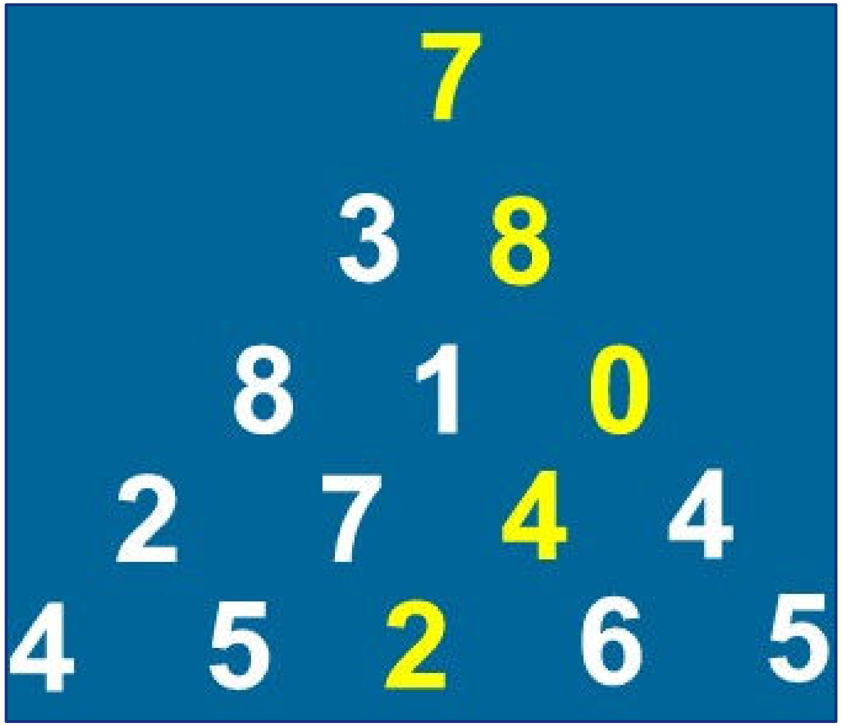


Problema

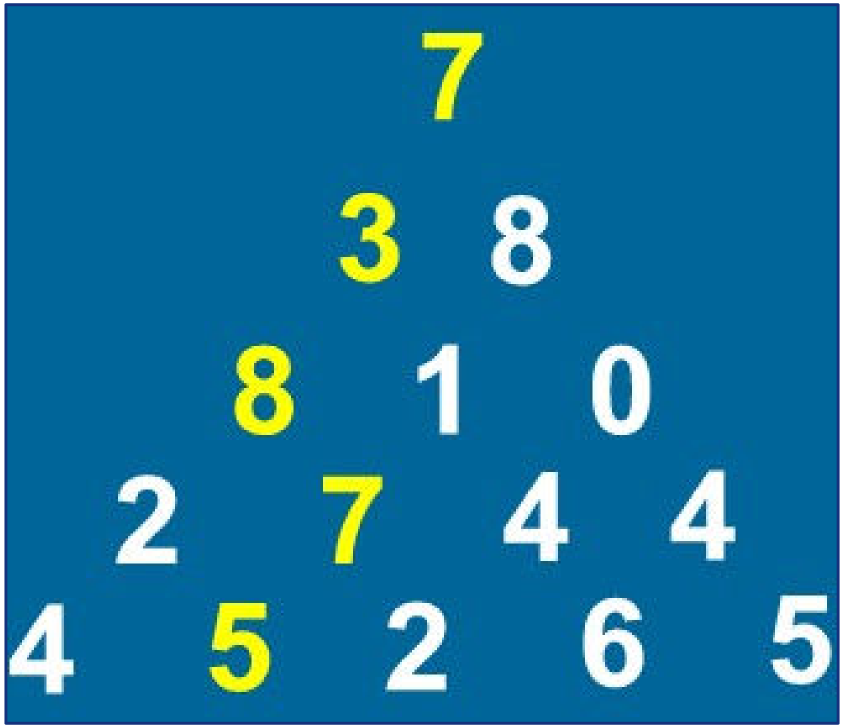
Calcular a rota, que começa no topo da pirâmide e acaba na base, com maior soma . Em cada passo podemos ir diagonalmente para baixo e para a esquerda ou para baixo e para a direita.

Pirâmide de Números

- Duas possíveis rotas



Soma = 21



Soma = 30

- **Limites:** todos os números da pirâmide são inteiros entre 0 e 99 e o número de linhas do triângulo é no máximo 100.

Pirâmide de Números

- Como resolver o problema?

- Ideia: **Força Bruta!**
  - avaliar todos os caminhos possíveis e ver qual o melhor.

- Mas quanto tempo demora isto?
  - Quantos caminhos existem?

Pirâmide de Números

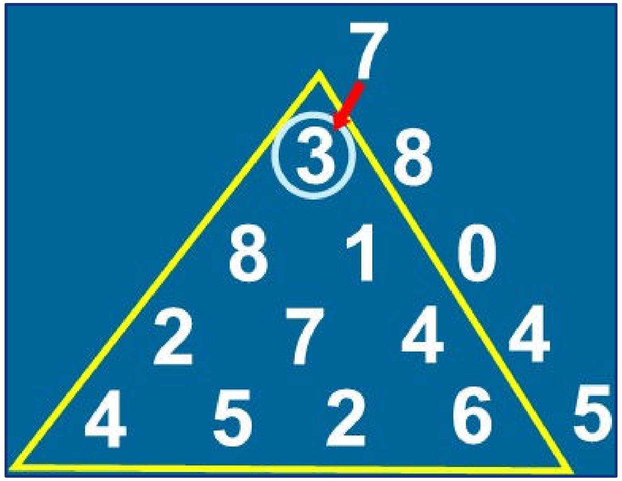
Análise da complexidade

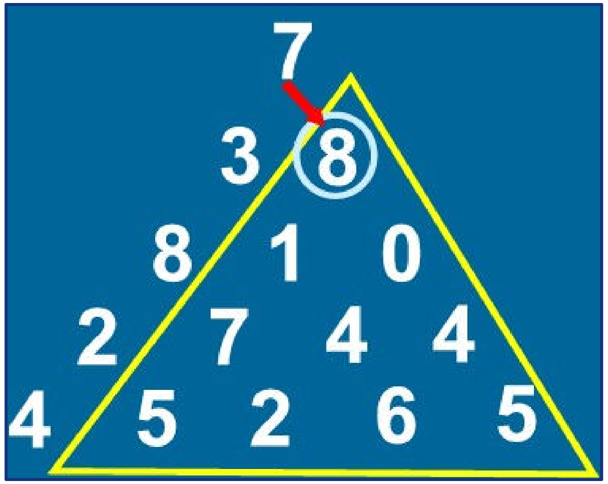
- Em cada linha podemos tomar **duas** decisões diferentes: esquerda ou direita
- Seja n a altura da pirâmide. Uma rota é constituída por **n–1** decisões diferentes.
- Existem $\mathbf{2^{n-1}}$ caminhos diferentes. Então, um programa que calculasse todas rotas teria complexidade temporal $\mathbf{O(2^n)}$: crescimento exponencial!
- Note-se que $\mathbf{2^{99} \approx 6{,}34 \times 10^{29}}$, que é um número demasiado grande!

633825300114114700748351602688

Pirâmide de Números

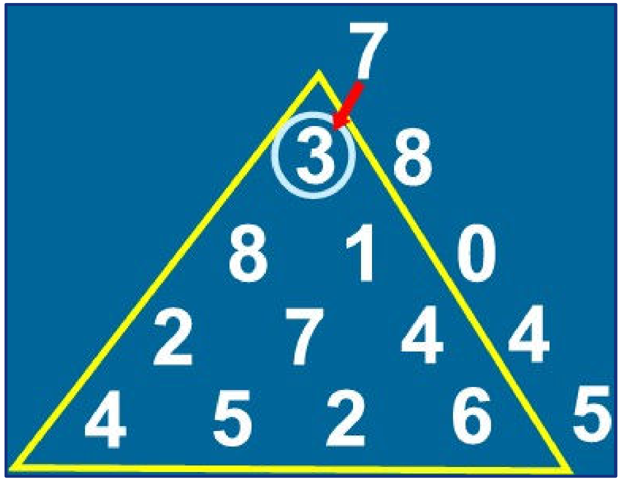
- Quando estamos no topo da pirâmide, temos duas decisões possíveis (esquerda ou direita):

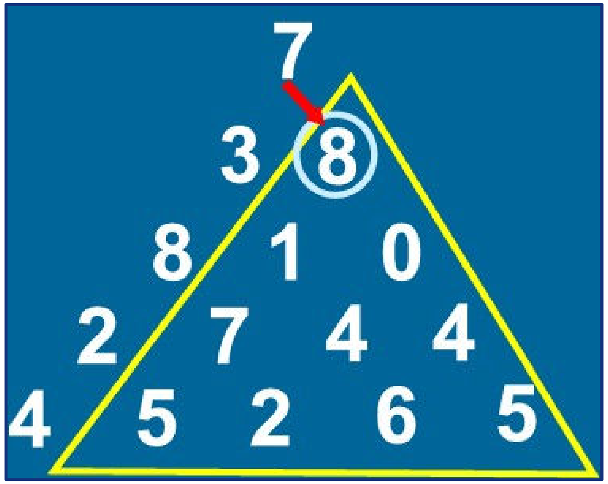




- Em cada um dos casos, temos de ter em conta todas as rotas das respectivas subpirâmides assinaladas a amarelo.

Pirâmide de Números





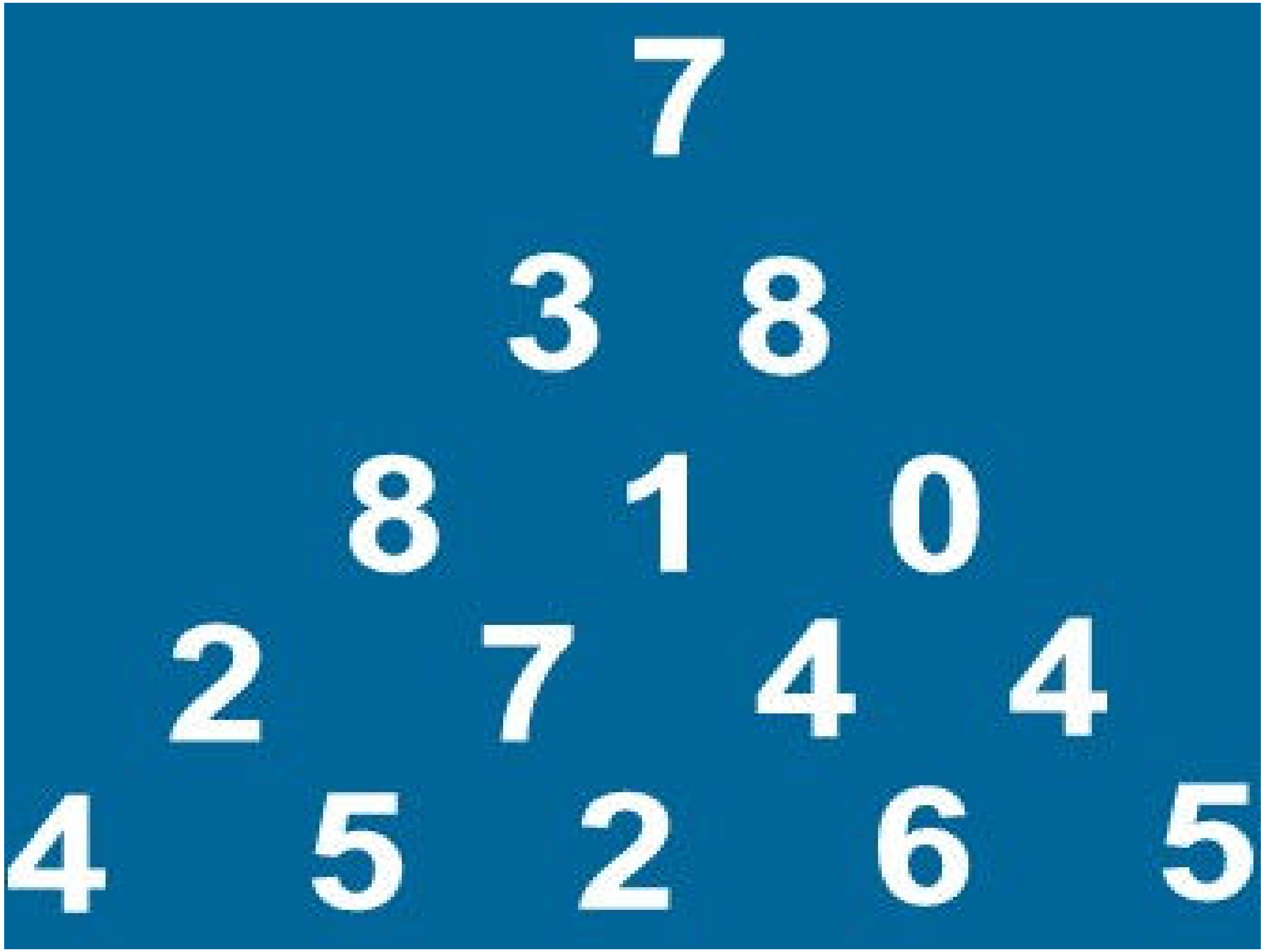
- Mas o que nos interessa saber sobre estas subpirâmides?

Apenas interessa o valor da sua melhor rota interna (que é um instância mais pequena do mesmo problema)!

- Para o exemplo, a solução é 7 mais o máximo entre o valor da melhor rota de cada uma das subpirâmides

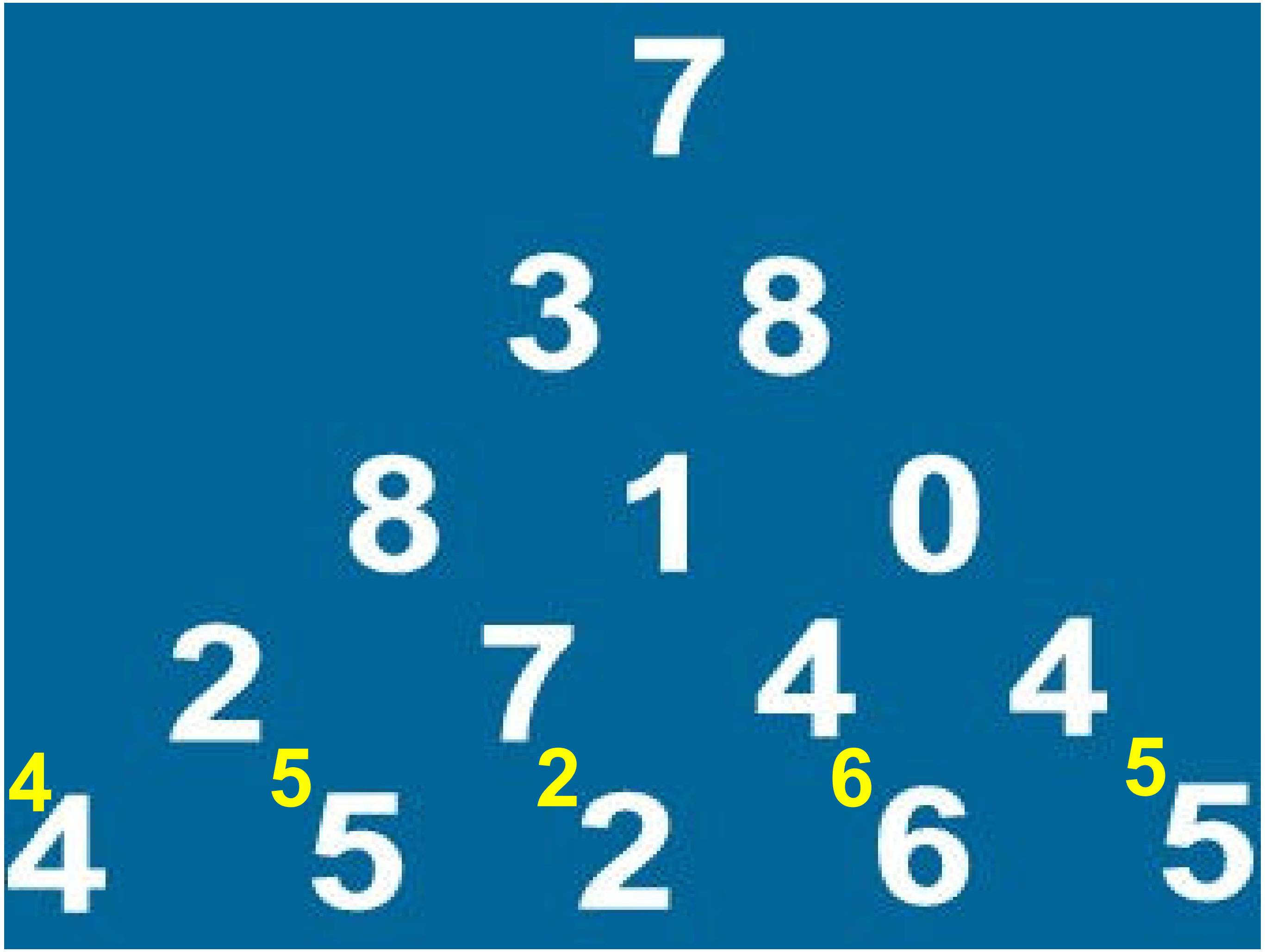
Pirâmide de Números

- Começar a partir do fim!



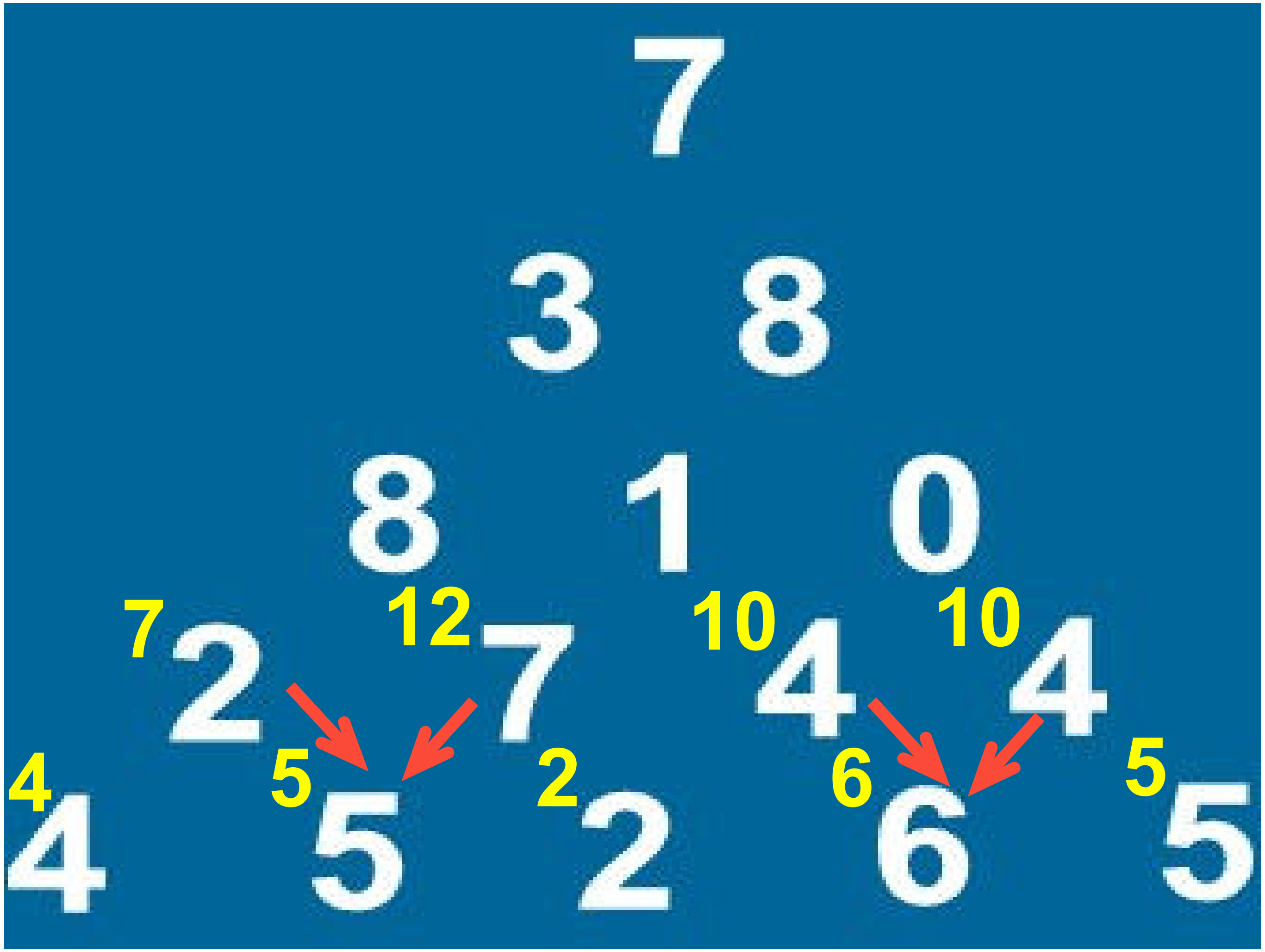
Pirâmide de Números

- Começar a partir do fim!



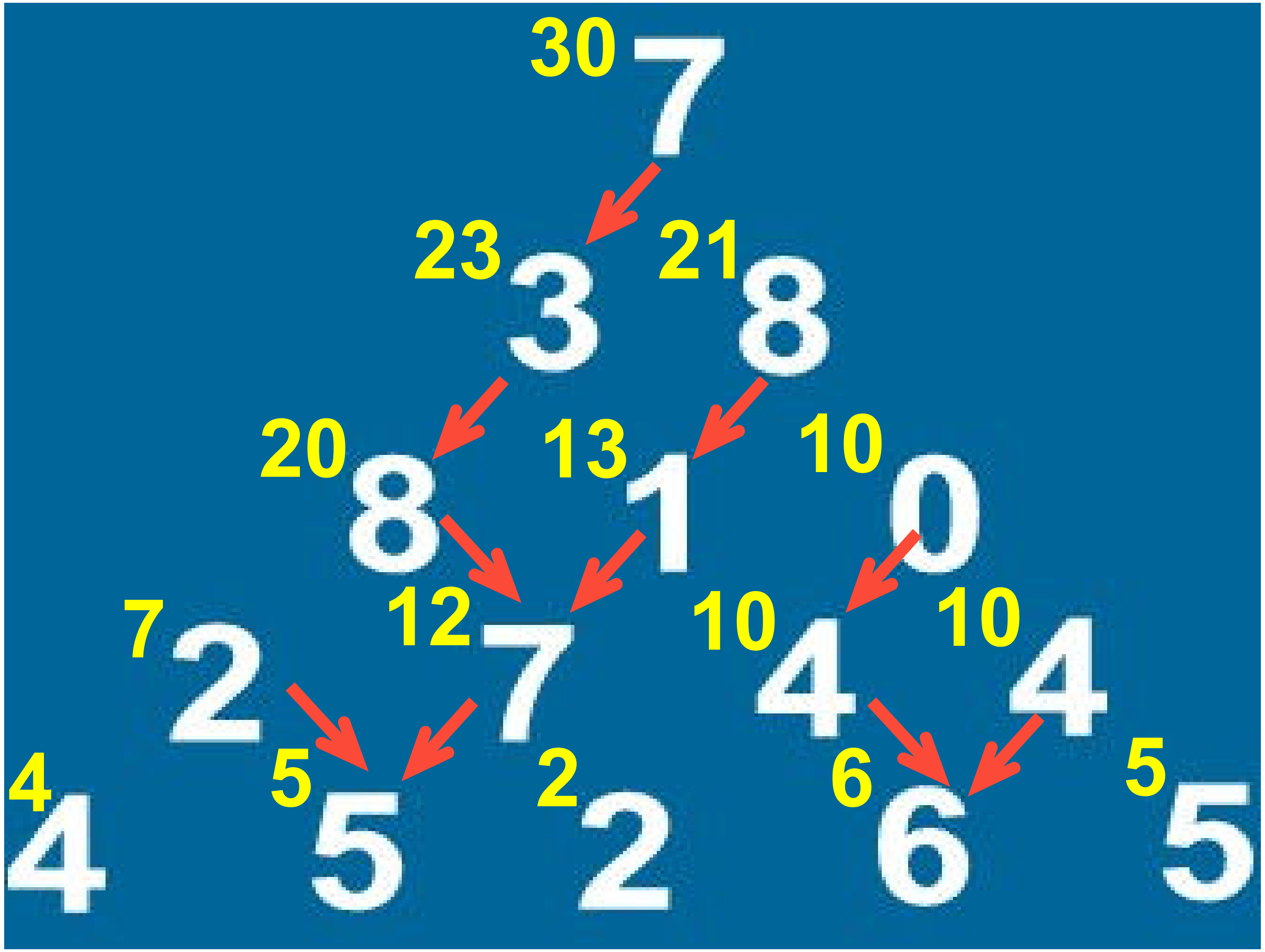
Pirâmide de Números

- Começar a partir do fim!



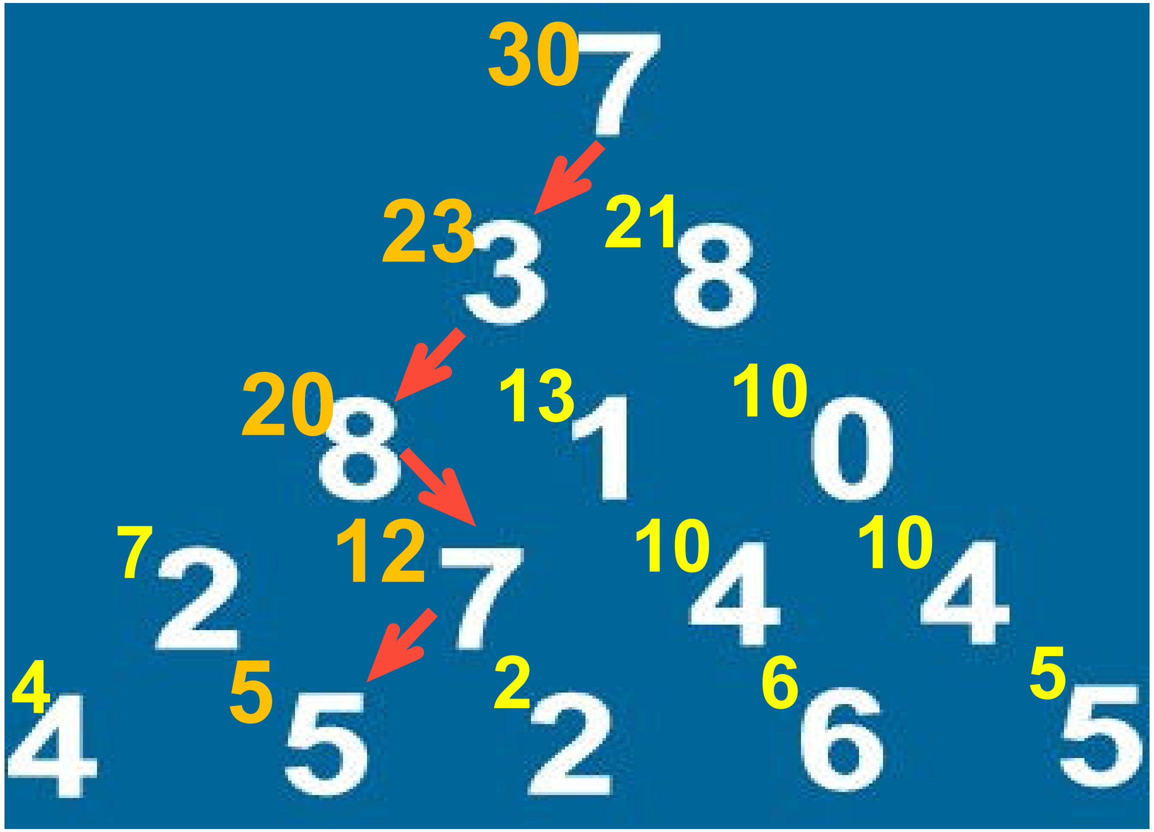
Pirâmide de Números

- Começar a partir do fim!



Pirâmide de Números

- Se fosse necessário saber constituição da melhor solução?
  - Basta usar a tabela já calculada!



vez dele B Fósforos	B (vez do B)	A	B fim
5	1	3	1
	2	2	1
	3	1	1
9	1	3	5
	2	2	5
	3	1	5
13	1	3	9
	2	2	9
	3	1	9
17	1	3	13
	2	2	13
	3	1	13
21	1	3	17
	2	2	17
	3	1	17
25	1	3	21
	2	2	21
	3	1	21
29	1	3	25
	2	2	25
	3	1	25
30		1	29

→ Dois jogadores A e B e 30 palitos.

→ Cada jogador pode tirar até 3 palitos.

→ Jogador A começa.

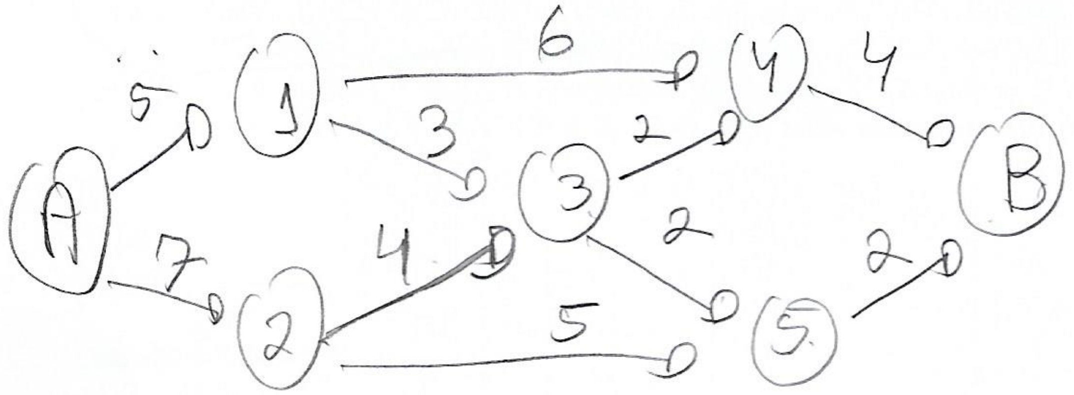
→ Perde quem tirar o último palito.

Ganha quem começa o jogo

Programação Dinâmica Discreta

$$W_k(x_k) = \min_a \left\{ c(x_k, a) + W_{k+1}(x_{k+1}, a) \right\}$$

Exemplo 1: Rota mais rápida entre 2 pontos A e B



$W(4) = 4$
$W(5) = 2$

$$W(3) = \min\{2 + W(4), 2 + \underline{W(5)}\} = \min\{6, 4\} = 4$$

$3 \to 5 \to B$

$$W(1) = \min\{6 + W(4), 3 + \underline{W(3)}\} = \{\overset{10}{6+4}, \overset{7}{3+4}\} = 7$$

$1 \to 3 \to 5 \to B$

$$W(2) = \min\{4 + W(3),\ 5 + W(5)\}$$

$$= \min\{4+4,\ \underline{5+2}\} = 7$$

$2 \to 5 \to B$

$$W(A) = \min\{5 + W(1),\ 7 + W(2)\}$$

$$= \min\{\underline{5+7},\ 7+7\} = 12$$

<u>$A \to 1 \to 3 \to 5 \to B$</u> rota ótima

<u>Exemplo 2</u>: Carregamento de um navio com unidades de N itens, item i pesa P_i e vale V_i, $i = 1, \ldots, N$. Capacidade máxima é P. Quanto se deve carregar de cada item de modo a maximizar o valor e não exceder o peso máximo.

	i	p_i	v_i
	1	2	65
$N=3$	2	3	80
	3	1	30

$P = 5$

$$\max \; v_1 x_1 + \dots + v_N x_N$$

$$p_1 x_1 + \dots + p_N x_N \leq P$$

$$x_i \geq 0, \text{ inteiro}$$

estágio j → item j

a) Estágio j → associado à variável x_j

b) Estado j (y_j) → peso disponível no navio no estágio j, com $y_1 = P$, $y_j = 0, \dots, P$, $j = 2, \dots, N$.

c) Ações x_j no estágio j → número de unidades do item j, variando de 0 a $\left[\frac{P}{p_j}\right]$ → o maior inteiro

(4)

$$W_N(y_N) = \max_{x_N = 0, \dots, \left[\frac{y_N}{p_N}\right]} \{ v_N x_N \}$$

$$W_n(y_n) = \max_{x_n = 0, \dots, \left[\frac{y_n}{p_n}\right]} \{ v_n x_n + W_{n+1}(y_n - p_n x_n) \}$$

n: item n; y_n: espaço disponível

No nosso exemplo,

$$W_3(y_3) = \max_{x_3 = 0, \dots, \left[\frac{y_3}{1}\right]} 30 \cdot x_3$$

$$= \max_{x_3 = 0, \dots, y_3} 30 x_3 = 30 y_3 \quad , y_3 = 0, \dots, 5$$

y_3	0	1	2	3	4	5
$W_3(y_3)$	0	30	60	90	120	150

$$W_2(y_2) = \max_{x_2 = 0, \dots, \left[\frac{y_2}{3}\right]} \left\{ 80x_2 + W_3(y_2 - 3x_2) \right\}$$

$$= \max_{x_2 = 0, \dots, \left[\frac{y_2}{3}\right]} \left\{ 80x_2 + 30(y_2 - 3x_2) \right\}$$

$$= \max_{x_2 = 0, \dots, \left[\frac{y_2}{3}\right]} \left\{ -10x_2 + 30y_2 \right\}$$

$$= 30y_2 \quad (x_2 = 0)$$

$$W_1(5) = \max_{x_1 = 0, \dots, \underbrace{\left[\frac{5}{2}\right]}_{2,5}} \left\{ 65x_1 + W_2(5 - 2x_1) \right\}$$

$$= \max_{x_1 = 0, \dots, 2} \left\{ 65x_1 + 30(5 - 2x_1) \right\}$$

$$= \max_{x_1 = 0, \dots, 2} \left\{ 5x_1 \right\} + 150 = 160$$

$$x_1 = 2$$

(6)

$x_1 = 2,\ x_2 = 0 \Rightarrow y_3 = 5 - 2.2 = \underline{1} \Rightarrow x_3 = \underline{1}$

<u>Peso</u>: $2.2 + 0.3 + 1.1 = 5$ ok

<u>Valor</u>: $2.65 + 0.80 + 1.30 = 160$

T. (1)

4.1. A equação de Bellman

Considere o modelo

$$x_{k+1} = f(x_k, u_k, w_k), \quad \{w_k\}$$ sequência de vetores aleatórios w_k, w_j independentes $k \neq j$.

x_0 independente de $\{w_k\}$

Procuramos controles da forma $u_k = u_k(x_k)$.

O objetivo é minimizar $u = (u_0, \ldots, u_{T-1})$

$$C_T(u) = E\left(\sum_{k=0}^{T-1} g(x_k, u_k) + h(x_T)\right)$$

A função valor $W_j(x)$ no instante j para este problema é o valor mínimo de

$$E_{j,x}\left(\sum_{k=j}^{T-1} g(x_k, u_k) + h(x_T)\right)$$

onde $E_{j,x}$ representa o valor esperado dado que o processo começa de $x_j = x$.

Se $x_j = x$ e o controle $u_j = v$ é aplicado, o próximo estado é

$x_{j+1} = g(x, v, w_j)$

e o custo mínimo restante para o resto do problema a partir de $j+1$ até T é $W_{j+1}(x_{j+1})$. Note que $W_{j+1}(x_{j+1})$ é uma variável aleatória, com custo esperado dado por ← o valor esperado em relação a w_j.

$$E\left(W_{j+1}\left(g(x, v, w_j)\right)\right)$$

A equação estocástica de Bellman para este problema é

$$\begin{cases} W_j(x) = \min_{v} \left\{ g(x,v) + E\left[W_{j+1}\left(f(x,v,w_j) \right) \right] \right\} \\ W_T(x) = h(x) \end{cases}$$

Assim podemos obter a sequência de funções $W_T, W_{T-1}, \ldots, W_0$ voltando no tempo.

<u>Proposição 4.1.1</u>: Suponha que $\overline{W}_N, \overline{W}_{N-1}, \ldots, W_0$ sejam dados como acima e que $u_j^1(x)$ seja o valor de V que minimiza $W_j(x)$. Então o controle via realimentação $u_k^* = u_k^1(x_k)$ minimiza o custo $C_N(u)$ sobre a classe de todos os controles de realimentação.

<u>Prova</u>: Seja $u_k(x_k)$ um controle de realimentação qualquer e seja x_k o processo dado pela equação de estado visto no início do capítulo. Então

$$W_N(x_N) - W_0(x_0) = \sum_{k=0}^{N-1} \left(W_{k+1}(x_{k+1}) - W_k(x_k) \right)$$

$$E\left\{W_N(x_N) - W_0(x_0)\right\} = \sum_{k=0}^{N-1} E\left\{ W_{k+1}(x_{k+1}) - W_k(x_k) \right\}$$

Dado x_k, $W_k(x_k)$ é conhecido e x_{k+1} é dado por

$$x_{k+1} = f(x_k, u_k(x_k), w_k)$$

Notando que

$$E\left(\overline{W}_{k+1}(x_{k+1}) - W_k(x_k)\right) = E\left(E\left(W_{k+1}(x_{k+1}) - W_k(x_k) \mid x_k\right)\right)$$

temos que, devido a independência de $W_{k+1}(x_{k+1})$ e x_k, ou em w_k

$$E\left(\overline{W}_{k+1}(x_{k+1}) \mid x_k\right) = E\left(\overline{W}_{k+1}\left(f(x_k, u(x_k), w_k)\right)\right)$$

onde o valor esperado acima é calculado em cima da de W_k com x_k fixo. Temos então que

$$\bar{W}_k(x_k) = E\left(W_k(x_k) \mid x_k\right) \leq \| D(k) x_k + F(k) u_k(x_k) \|^2 + E\left(W_{k+1}\left(A(k) x_k + B(k) u_k(x_k) + C(k) w_k\right)\right) = \| D(k) x_k + F(k) u_k(x_k) \|^2 + E\left(W_{k+1}(x_{k+1}\right.$$

$g(x_k, u_k)$; $f(x_k, u_k, k)$

$$\therefore E\left(W_{k+1}(x_{k+1}) - W_k(x_k) \mid x_k\right) \geq - \underbrace{\| D(k) x_k + F(k) u_k(x_k) \|^2}_{g(x_k, u_k)}$$

logo,

$$E\left(W_N(x_N) - W_0(x_0)\right) \geq -E\left(\sum_{k=0}^{N-1} \| D(k) x_k + F(k) u_k(x_k) \|^2\right)$$

$h(x_N)$

e como $W_N(x_N) = x_N' Q x_N$, obtemos que

$$E\left(W_0(x_0)\right) \leq E\left(\sum_{k=0}^{N-1} \| D(k) x_k + F(k) u_k(x_k) \|^2 + x_N' Q x_N\right)$$

$$\therefore E\left(W_0(x_0)\right) \leq C_N(u)$$

Por outro lado, os mesmos argumentos acima valem com igualdade ao invés de desigualdade para u^1. Logo,

$$E\left(W_0(x_0)\right) = C_N(u^1)$$

e portanto u^1 é ótimo.

Carteira Auto-Financiável

$$V(t) = u_0(t) + u(t)$$

$u_0(t)$ → o valor investido no livre de risco

$u(t)$ → o valor investido no ativo de risco

r_f → a rentabilidade livre de risco

$R(t)$ = rentabilidade do ativo de risco

$$V(t+1) = (1+r_f)\,u_0(t) + (1+R(t))\,u(t)$$

$$\Rightarrow u_0(t) = V(t) - u(t)$$

$$V(t+1) = (1+r_f)(V(t) - u(t)) + (1+R(t))\,u(t)$$

$$= (1+r_f)V(t) + (R(t) - r_f)\,u(t)$$

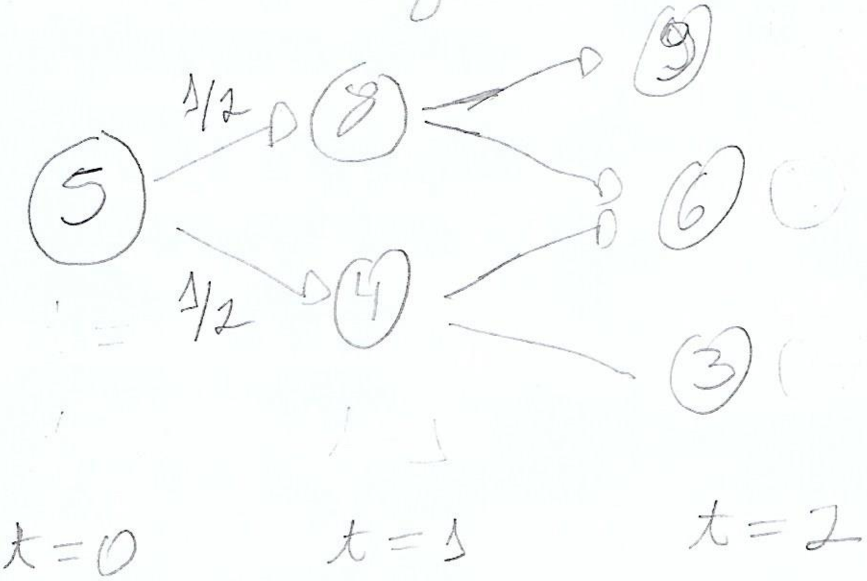
Problema: Achar $u(0), \dots, u(T-1)$ de modo a

$$\max E\big(U(V(T))\big), \quad V(0) = v_0$$

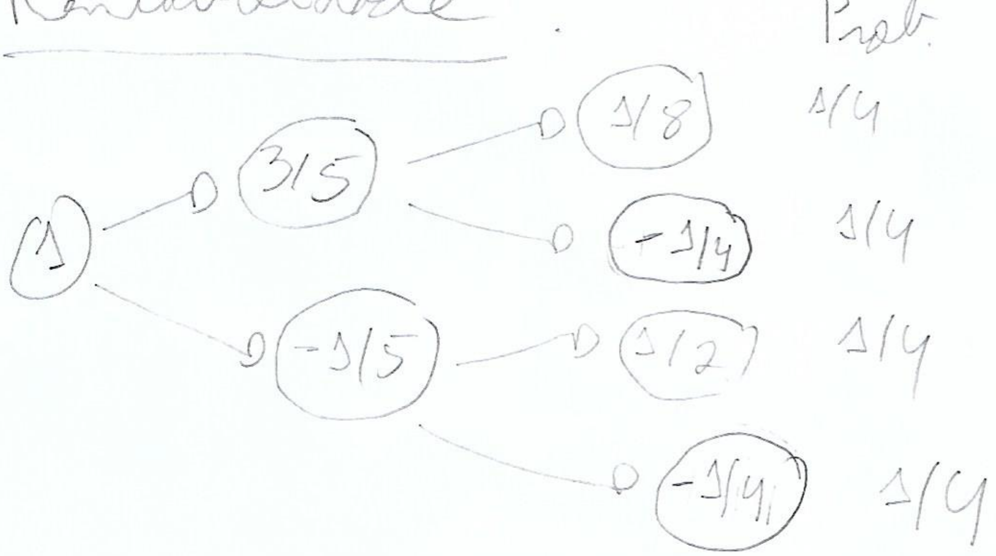
Equação de Bellman

$$\begin{cases} W_t(v) = \max\limits_{u} E\Big(W_{t+1}\big((1+r_f)v + (R(t) - r_f)u\big)\Big) \\ W_T(v) = U(v) \end{cases}$$

Exemplo: Considere os valores de $S(t)$ como a seguir



Rentabilidade:



$\frac{8}{5} = 1 + \frac{3}{5}$

$\frac{4}{5} = 1 - 1/5$

$\frac{9}{8} = 1 + \frac{1}{8}$

$\frac{6}{8} = 1 - \frac{1}{4}$

$\frac{6}{4} = 1 + \frac{1}{2}$

$\frac{3}{4} = 1 - \frac{1}{4}$

$0 \leq w(t) = \frac{u(t)}{V(t)}$ → % no ativo de risco (sem alavancagem)

$$W_t(v) = \max_{w \geq 0} E\left(W_{t+1}\left(v\left(1 + r_f + (R(t) - r_f)w\right)\right)\right)$$

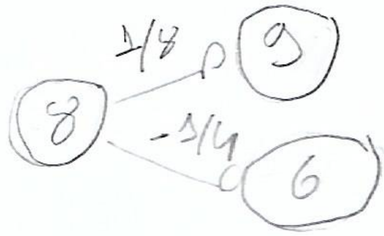
③

Vamos considerar: $\begin{cases} U(v) = \ln(v) \\ r_g = 0 \end{cases}$

O objetivo é:

$$\max_{\substack{w(0)\geq 0 \\ w(1)\geq 0}} E\left(\ln(V(2))\right)$$

$t = 1$; nó de subida →



$$(4)\quad W_{2,1}(v) = \max_{w_2 \geq 0} E\left(\ln\left(v\left(1 + R(1)\,w_2\right)\right)\right) =$$

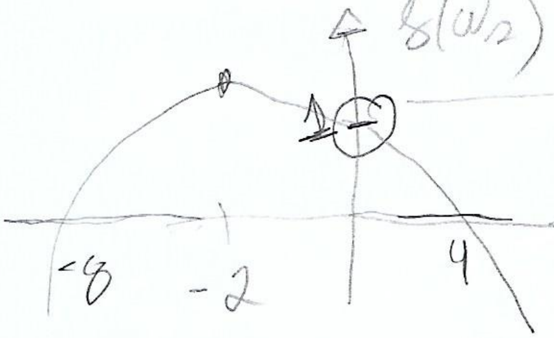
$$\max_{w_2 \geq 0}\left\{ \frac{1}{2}\ln\left(v\left(1+\frac{1}{8}w_2\right)\right) + \frac{1}{2}\ln\left(v\left(1-\frac{1}{4}w_2\right)\right)\right\} =$$

$$\frac{1}{2}\max_{w_2 \geq 0}\left\{ \ln\left(1+\frac{1}{8}w_2\right) + \ln\left(1-\frac{1}{4}w_2\right)\right\} + \ln(v) =$$

$$\frac{1}{2}\max_{w_2 \geq 0}\left\{ \ln\left(\left(1+\frac{1}{8}w_2\right)\left(1-\frac{1}{4}w_2\right)\right)\right\} + \ln(v)$$

$$g(w_2) = \left(1+\frac{1}{8}w_2\right)\left(1-\frac{1}{4}w_2\right) = 1 - \frac{1}{8}w_2 - \frac{1}{32}w_2^2$$

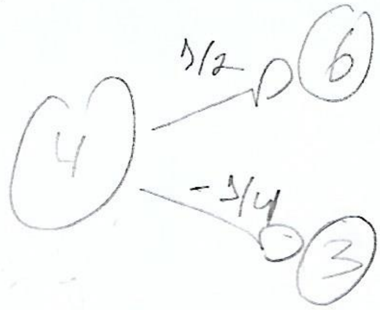
$$\frac{d f(w_2)}{d_2} = -\frac{1}{8} - \frac{1}{16} w_2 = 0 \Leftrightarrow w_2 = -2$$ (4)



máximo para $W_2 \geq 0$ em $W_2 = 0$

$$W_{2,1}(v) = \ln(v), \quad w_2^{*} = 0$$

(d) $t = 1$, nó de descida



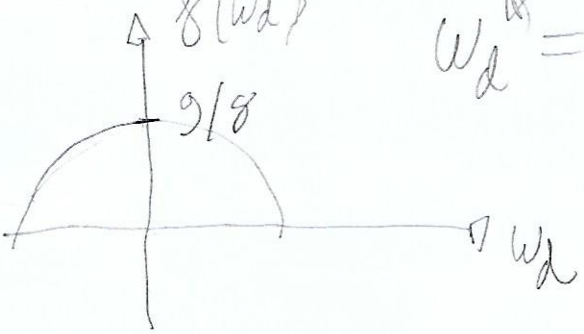
$$W_{d,1}(v) = \max_{w_d \geq 0} E\left(\ln\left(v(1+R(1))w_d\right)\right) =$$

$$\frac{1}{2} \max_{w_d \geq 0} \left\{ \ln\left(v\left(1+\tfrac{1}{2}w_d\right)\right) + \ln\left(v\left(1-\tfrac{1}{4}w_d\right)\right) \right\} =$$

$$= \frac{1}{2} \max_{w_d \geq 0} \left\{ \ln\left(\left(1+\tfrac{1}{2}w_d\right)\left(1-\tfrac{1}{4}w_d\right)\right) \right\} + \ln(v)$$

$$f(w_d) = \left(1+\frac{1}{2}w_d\right)\left(1-\frac{1}{4}w_d\right) = 1 + \frac{1}{4}w_d - \frac{1}{8}w_d^2$$

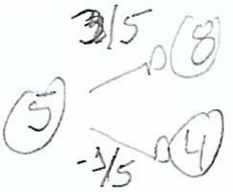
$$\frac{d f(w)}{dw} = \frac{1}{4} - \frac{1}{4} w = 0 \Leftrightarrow w = 1$$



$w_d^* = 1$

$$W_{d,1}(v) = \frac{1}{2}\left(\ln\left(9/8\right)\right) + \ln(v) = \ln\left(3/\sqrt{8}\right) + \ln(v)$$

t = 0



$$W_0(v) = \frac{1}{2} \max_{w_0 \geq 0} \left\{ \ln\left(v\left(1 + \frac{3}{5} w_0\right)\right) + \ln\left(v\left(1 - \frac{1}{5} w_0\right)\right) \right\}$$

$$+ \frac{1}{2} \ln\left(\frac{3}{\sqrt{8}}\right) = \frac{1}{2} \max_{w_0 \geq 0} \left\{ \ln\left(v\left(1 + \frac{3}{5} w_0\right)\left(1 - \frac{1}{5} w_0\right)\right) \right\} +$$

$$+ \frac{1}{2} \ln\left(3/\sqrt{8}\right) + \ln(v) \qquad g(w_0) = \left(1 + \frac{3}{5} w_0\right)\left(1 - \frac{1}{5} w_0\right)$$

$$g(w) = 1 + \frac{2}{5} w_0 - \frac{3}{25} w_0^2 \qquad \frac{d g(w)}{dw} = \frac{2}{5} - \frac{6}{25} w_0 = 0$$

$$\Rightarrow w_0^* = 5 \cdot \frac{2}{6} = \frac{5}{3}, \quad g(w_0^*) = 1 + \frac{2}{3} - \frac{1}{3} = \frac{4}{3}$$

$$W_0(v) = \frac{1}{2}\left(\ln\left(4/3\right) + \ln\left(3/\sqrt{8}\right)\right) + \ln(v)$$

$$= \frac{1}{2} \ln\left(\frac{4}{2\sqrt{2}}\right) + \ln(v) = \frac{1}{2} \ln\sqrt{2} + \ln(v)$$

$$= \ln(v) + \frac{1}{4} \ln(2)$$

$w_0^* = \frac{5}{3}$

$w_{rf}^* = -\frac{2}{3}$

v_0

$t = 0$

$v_u = v_0\left(1 + \frac{5}{3}\frac{3}{5}\right) = 2v_0$

$w_u^* = 0$

$w_{rf}^* = 1$

$v_{uu} = 2v_0$

$v_{ud} = 2v_0$

$v_d = v_0\left(1 - \frac{5}{3}\frac{1}{5}\right) = \frac{2}{3}v_0$

$w_d^* = 1$

$w_{rf}^* = 0$

$v_{du} = \left(1 + \frac{1}{2}\right)\left(\frac{2}{3}v_0\right) = v_0$

$v_{dd} = \left(1 - \frac{1}{4}\right)\left(\frac{2}{3}v_0\right) = \frac{1}{2}v_0$

$$W^*(v_0) = \frac{1}{4}\left\{\ln(2v_0) + \ln(2v_0) + \ln(v_0) + \ln\left(\frac{1}{2}v_0\right)\right\}$$

$$= \ln(v_0) + \frac{1}{4}\left\{2\ln(2) + \ln\left(\frac{1}{2}\right)\right\} = \ln(v_0) + \frac{1}{4}\ln(2)$$

CHAPTER 6

Optimal control for state-space models

This chapter concerns optimal control problems for the state-space models discussed in Chapters 2 and 3. The state and observation processes x_k and y_k are given respectively by the equations

$$x_{k+1} = A(k)x_k + B(k)u_k + C(k)w_k \tag{6.0.1}$$

$$y_k = H(k)x_k + G(k)w_k \tag{6.0.2}$$

where w_k is a white-noise sequence. We now wish to choose the control sequence u_k so that the system behaves in some desirable way. We have to settle two questions at the outset, namely what sort of controls are to be allowed (or, are *admissible*) and what the control objective is.

The simplest class of controls is that of *open-loop* controls which are just deterministic sequences $u_0, u_1, \ldots$, chosen *a priori*. In this case the observation equation (6.0.2) is irrelevant since the system dynamics are entirely determined by the state equation (6.0.1). As we shall see in Section 6.1, open-loop controls are in some sense adequate for non-stochastic problems ($w_k \equiv 0$). Generally, however, it is better to use some form of *feedback control*. Such a control selects a value of u_k on the basis of measurements or observations of the system. We have *complete observations* if the state vector x_k can be measured directly, and, since the future evolution of the system depends only on its current state and future controls and noise, the natural form of control is then *state feedback*: $u_k = u_k(x_k)$. The functions $u_1(\cdot), u_2(\cdot), \ldots$ are sometimes described as a *control policy* since they constitute a decision rule: *if* the state at time k is x, *then* the control applied will be $u = u_k(x)$. Again, the observations y_k are irrelevant in this situation. In the case of *noisy measurements* or *partial observations*, however, x_k cannot be measured directly and only the sequence $y_0, y_1, \ldots, y_k$ is available. Feedback control now means that u_k is determined on the basis of the available measurements: $u_k = u_k(y_0, y_1, \ldots, y_k)$. In this case, since y_k is not the state of the system, one generally does better by allowing dependence on all past observations, not just on the current observation y_k. Finally, we shall assume throughout that the control *values* are unconstrained. It would be perhaps more realistic to restrict the values of the controls by introducing constraints of the form $|u_k| \leq 1$. While this causes no theoretical difficulties, it would make the calculation of explicit control policies substantially more difficult.

We now turn to the control objective. In classical control system design the objectives are qualitative in nature: one specifies certain stability and transient response characteristics, and any design which meets the specification will be regarded as satisfactory. The 'pole shifting' controllers considered in Chapter 7 follow this general philosophy. Here, however, our formulation is in terms of *optimal control.* The idea is as follows: the class of admissible controls is specified precisely and a scalar performance criterion or cost function $C(u)$ is associated with each control. We can then ask which control achieves the minimum cost; this control is *optimal.* Once the three ingredients (system dynamics, admissible controls and cost criterion) are specified, determination of the optimal control is in principle a purely mathematical problem involving no 'engineering judgement'. Indeed, optimal control theory has often been criticized precisely on these grounds. It may well be that a control which is theoretically optimal is subjectively quite unsatisfactory. If it is, this will be because the system model is inadequate or because the cost criterion fails to take account of all the relevant features of the problem. On the other hand, a more realistic model or a criterion which *did* include all the relevant features might well lead to an impossibly complicated optimization problem. As usual, the true situation is a trade-off between realistic modelling and mathematical tractability, and this is where the engineering judgement comes in.

In this chapter we shall study *linear regulator* problems, where the cost criterion is given by

$$C_N(u) = E\left[\sum_{k=0}^{N-1} \|Dx_k + Fu_k\|^2 + x_N^{\mathrm{T}} Q x_N\right]. \tag{6.0.3}$$

The number N of stages in the problem is called the *time horizon* and we shall consider both the finite-horizon ($N < \infty$) and infinite-horizon ($N = \infty$) cases. Further discussion of the cost function $C_N(u)$

will be found in Section 6.1. It implies a general control objective of regulating the state x_k to 0 while not using too much control energy as measured by the quantity $u_k^T F^T F u_k$. Note that the quantity in square brackets in (6.0.3) is a random variable and we obtain a scalar cost function (as required for optimization) by taking its expected value, which is practical terms means that we are looking for a control policy which gives the minimum average cost over a long sequence of trials.

The optimization problem represented by equations (6.0.1)–(6.0.3) is known as the LQG problem since it involves a *linear* system (6.0.1), (6.0.2), a *quadratic* cost criterion (6.0.3) and *gaussian* or normal white-noise disturbances in the state-space model. (For reasons explained below, $\{w_k\}$ is assumed here to be a sequence of independent normal random variables rather than a 'wide-sense' white noise as generally considered in previous chapters.) It is sufficiently general to be applicable in a wide variety of cases and the optimal control is obtained in an easily implemented form. It also has, as we shall see, close relations with the Kalman filter.

In addition to the standard linear regulator as defined above we shall study the same problem with *discounted costs*:

$$C^{\rho}(u) = E\left[\sum_{k=1}^{N-1} \rho^k \| Dx_k + Fu_k \|^2 + \rho^N x_N^T Q x_N\right]$$

where ρ is a number, $0 < \rho < 1$. There are important technical reasons for introducing the *discount factor* ρ, but there is also a financial aspect to it. Suppose that money can be invested at a constant interest rate $r\%$ per annum and one has to pay bills of £a_0, £$a_1, \ldots$ each year starting at the present time. What capital is needed to finance these bills entirely out of investment income? Since £1 now is worth £$(1 + 0.01\,r)^k$ in k years' time, the amount required is $\sum_k a_k \rho^k$ where $\rho = (1 + 0.01\,r)^{-1}$ and this is one's total debt *capitalized at its present value*. In particular, a constant debt of £a/year in perpetuity can be financed with a capital of

$$£ \sum_{k=0}^{\infty} a\rho^k = £a/(1 - \rho).$$

An important feature of this result is that while the total amount of debt is certainly infinite, it nevertheless has a finite capital value. Similarly, in the control problems, the discount factor enables us to attach a finite cost (and therefore consider optimization) in cases where without discounting the cost would be $+\infty$ for all control policies. Of course it is not realistic to assume that interest rates will remain constant for all time, and a more subjective interpretation of $C^{\rho}(u)$ is simply to say that it attaches small importance to costs which have to be paid at some time in the distant future.

In the three sections of this chapter we discuss the linear regulator problem in three stages. First, in Section 6.1 we consider the deterministic case when $w_k = 0$. Many of the 'structural features' of the LQG problem are already present in this case, and the optimal control turns out to be linear feedback: $u_k = -M(k)x_k$ for a precomputable sequence of matrices $M(k)$. This same control is shown in Section 6.2 to be optimal also in the stochastic case with complete observations, the effect of the noise being simply to increase the cost. Finally we consider the 'full' LQG problem in Section 6.3 and show that the optimal control is now $-M(k)\hat{x}_{k|k-1}$ where $\hat{x}_{k|k-1}$ is the best estimate of the state given the observations, generated by the Kalman filter. This results demonstrates the so-called 'certainty-equivalence' principle: if the state cannot be observed directly, estimate it and use the estimate as if it were the true state. We also discuss an idea of somewhat wider applicability known as the 'separation principle'.

6.1 The deterministic linear regulator

6.1.1 Finite time horizon

In this section we consider control of the linear system

$$x_{k+1} = A(k)x_k + B(k)u_k \tag{6.1.1}$$

for $k = 0, 1, \ldots, N$ with a given initial condition x_0. We wish to choose a control sequence $u = (u_0, u_1, \ldots, u_{N-1})$ so as to minimize the cost†

$$J_N(u) = \sum_{k=0}^{N-1} \|D(k)x_k + F(k)u_k\|^2 + x_N^{\mathrm{T}} Q x_N. \tag{6.1.2}$$

Here $D(k)$, $F(k)$ are matrices of dimensions $p \times n$, $p \times m$ respectively and Q is a non-negative definite symmetric $n \times n$ matrix. It will be assumed throughout that the $m \times m$ matrices $F^{\mathrm{T}}(k)F(k)$ are strictly positive definite, which implies in particular that we must have $p \geq m$.

We shall also study various infinite-time problems related to (6.1.1)–(6.1.2), i.e. consider what happens as $N \to \infty$.

†We denote the cost by J_N in the deterministic case, reserving C_N for the average cost in the stochastic problem.

The cost function $J_N(u)$ is somewhat different from that conventionally employed in treatments of this subject. The more usual form of cost function is

$$\tilde{J}_N(u) = \sum_{k=0}^{N-1} (x_k^{\mathrm{T}} Q(k) x_k + u_k^{\mathrm{T}} R(k) u_k) + x_N^{\mathrm{T}} Q x_N$$

where $Q(k)$, $R(k)$ are symmetric non-negative definite matrices (strictly positive definite in the case of $R(k)$). This has more intuitive appeal since the terms involving x_k penalize deviation of x_k from 0 while $\Sigma u_k^{\mathrm{T}} R(k) u_k$ is a measure of control energy. Thus the control problem is to steer x_k to zero as quickly as possible without expending too much control energy; energy expenditure can be penalized more or less heavily by appropriate specification of the matrices $R(k)$. This cost function is, however, a special case of (6.1.2): take $p = n + m$ and

$$D(k) = \left[\begin{array}{c} Q^{1/2}(k) \\ \hline 0 \end{array}\right] \qquad F(k) = \left[\begin{array}{c} 0 \\ \hline R^{1/2}(k) \end{array}\right]$$

where $Q^{1/2}(k)$, $R^{1/2}(k)$ are any 'square roots' of $Q(k)$, $R(k)$, i.e. satisfy $(Q^{1/2}(k))^{\mathrm{T}} Q^{1/2}(k) = Q(k)$ (and similarly for $R^{1/2}(k)$). Such square roots always exist for non-negative definite symmetric matrices, as shown in Appendix D, Proposition D.1.3.

We prefer the cost function (6.1.2) because of its extra generality, but more importantly because it connects up naturally with the formulation of the Kalman filter given in Chapter 3. This will become apparent below.

The control problem (6.1.1)–(6.1.2) can in principle be regarded as an unconstrained minimization problem. For a given sequence $u = (u_0, u_1, \ldots, u_{N-1})$ and initial condition x_0, the corresponding x_k sequence can be computed from the state equations (6.1.1):

$$\begin{aligned} x_1 &= A(0)x_0 + B(0)u_0 \\ x_2 &= A(1)x_1 + B(1)u_1 \\ &= A(1)A(0)x_0 + A(1)B(0)u_0 + B(1)u_1, \quad \text{etc.} \end{aligned}$$

Substituting in (6.1.2), we obtain $J_N(u)$ explicitly as a function of the mN-vector $u = \mathrm{col}\{u_0, u_1, \ldots, u_{N-1}\}$ and one could now use 'standard' hill-climbing techniques to find the vector u^* which minimizes $J_N(u)$. This would, however, be a very unsatisfactory way of solving the problem. Not only is the dimension mN very large even for innocuous-looking problems, but also we have thrown away an essential feature of the problem, namely its dynamic structure, and therefore calculation of the optimal u^* would give us very little insight into what is really happening in the optimization process.

A solution method which uses in an essential way the dynamic nature of the problem is R. Bellman's technique of dynamic programming. Introduced by Bellman in the mid-1950s, dynamic programming has been the subject of extensive research over the years and the associated literature is now enormous. We propose to discuss it here only to the extent necessary to solve the problem at hand. The basic idea is, like many good ideas, remarkably simple, and is known as Bellman's *principle of optimality*. Suppose that u^* is an optimal control for the linear regulator problem (6.1.1)–(6.1.2), that is to say,

$$J_N(u^*) \leq J_N(u)$$

for all other controls $u = (u_0, u_1, \ldots, u_{N-1})$. Let $x_0^* = x_0, x_1^*, \ldots, x_N^*$ be the corresponding state trajectory given by (6.1.1) with $u_k = u_k^*$. Now fix an integer j, $0 \leq j < N$, and consider the 'intermediate' problem of minimizing

$$J_{N,j}(u^{(j)}) = \sum_{k=j}^{N-1} \|D(k)x_k + F(k)u_k\|^2 + x_N^{\mathrm{T}} Q x_N$$

over controls $u^{(j)} = (u_j, u_{j+1}, \ldots, u_{N-1})$, subject to the dynamics (6.1.1) as before with the 'initial condition'

$$x_j = x_j^*.$$

The intermediate problem is thus to optimize the performance of the system over the last $N-j$ stages, starting at a point x_j^* which is on the optimal trajectory for the *overall* optimization problem. The principle of optimality states that *the control* $u^{*(j)} = (u_j^*, u_{j+1}^*, \ldots, u_{N-1}^*)$ *is optimal for the intermediate problem.* Put another way, if u^* is optimal for the overall problem then $u^{*(j)}$ is optimal over the last $N-j$ stages starting at x_j^*. The reason for this is fairly clear: if $u^{*(j)}$ were *not* optimal for the intermediate problem then there would be some sequence $\tilde{u}^{(j)} = (\tilde{u}_j, \tilde{u}_{j+1}, \ldots, \tilde{u}_{N-1})$ such that

$$J_{N,j}(\tilde{u}^{(j)}) < J_{N,j}(u^{*(j)}).$$

Now consider the control u^0 defined as follows:

$$u_k^0 = \begin{cases} u_k^* & k < j \\ \tilde{u}_k & k \geq j \end{cases}$$

and let x_k^0 be the corresponding trajectory. Then $x_k^0 = x_k^*$ for $k \le j$ and hence

$$\begin{aligned} J_N(u^0) &= \sum_{k=0}^{j-1} \|D(k)x_k^* + F(k)u_k^*\|^2 + J_{N,j}(\tilde{u}^{(j)}) \\ &< \sum_{k=0}^{j-1} \|D(k)x_k^* + F(k)u_k^*\|^2 + J_{N,j}(u^{*(j)}) \\ &= J_N(u^*). \end{aligned} \tag{6.1.3}$$

But this contradicts the supposition that u^* is optimal. Thus $u^{*(j)}$ must be optimal for the intermediate problem, as claimed.

In the preceding argument, the system started in a fixed but arbitrary state x_0. However, there is nothing special about the initial time zero: the same argument implies that if $\{x_k^*, u_k^*, k \ge j\}$ is an optimal control-trajectory sequence for the intermediate problem starting at $x_j = x$ (arbitrary) then $\{x_k^*, u_k^*, k \ge j'\}$ is optimal for the further intermediate problem starting at $x_{j'} = x_{j'}^*$ for any j' between j and $N-1$.

The principle of optimality is turned into a practical solution technique as follows. Let $V_j(x)$ be the minimum cost for the intermediate problem starting at $x_j = x$. This is known as the *value function* at time j. Then taking $j' = j+1$, the above argument indicates that V_j ought to satisfy

$$V_j(x) = \min_v \left[\|D(j)x + F(j)v\|^2 + V_{j+1}(A(j)x + B(j)v) \right] \tag{6.1.4}$$

the minimum being taken over all m-vectors v. Essentially, this comes from calculations similar to (6.1.3) above. If $x_j = x$ and control $u_j = v$ is applied, then:

(a) The cost paid at time j is $\|D(j)x + F(j)v\|^2$.
(b) The next state is $x_{j+1} = A(j)x + B(j)v$.

Thus $V_{j+1}(A(j)x + B(j)v)$ is the minimal cost for the rest of the problem if control value v is applied at stage j. So certainly

$$V_j(x) \le \|D(j)x + F(j)v\|^2 + V_{j+1}(A(j)x + B(j)v) \tag{6.1.5}$$

and this holds for any value of v. On the other hand, if $\{x_k^*, u_k^*\}$ is optimal over the last $N-j$ stages starting at $x_j^* = x$, then the principle of optimality indicates that

$$V_l(x_l^*) = \sum_{k=l}^{N-1} \|D(k)x_k^* + F(k)u_k^*\|^2 + x_N^{*\mathrm{T}} Q x_N^*$$

where l is either j or $j+1$, and this shows since $x_j^* = x$ that

$$V_j(x) = \|D(j)x + F(j)u_j^*\|^2 + V_{j+1}(A(j)x + B(j)u_j^*). \tag{6.1.6}$$

Now (6.1.5) and (6.1.6) together imply that (6.1.4) holds.

Equation (6.1.4) is known as the *Bellman equation* and is the basic entity in discrete-time dynamic programming since it enables the optimal control u^* to be determined. Note that *at the terminal time N* the value function is

$$V_N(x) = x^T Q x, \tag{6.1.7}$$

since no further control is possible and one has no choice but to pay the terminal cost of $x^T Q x$. Applying (6.1.4) with $j = N-1$ gives

$$V_{N-1}(x) = \min_v [\|D(N-1)x + F(N-1)v\|^2 + (A(N-1)x + B(N-1)v)^T Q (A(N-1)x + B(N-1)v)]$$

and hence determines $V_{N-1}(x)$. Now using (6.1.4) again we can calculate $V_{N-2}, V_{N-3}, \ldots, V_0$. By definition, $V_0(x_0)$ is then the minimal cost for the overall problem starting at state x_0. From (6.1.5) and (6.1.6), the optimal control u_j^* is just the value of v that achieves the minimum in (6.1.4) with $x = x_j^*$.

Before proceeding any further let us consolidate the discussion so far. We have used the principle of optimality to obtain the Bellman equation (6.1.4) and this suggests the procedure outlined above for obtaining an optimal control. Having arrived at this procedure, however, we can verify that it is correct by a simple and self-contained argument; this will be given below. Thus the principle of optimality is actually only a heuristic device which tells us why we would expect the Bellman equation to take the form it does; it does not appear in the final formulation of any results. One could present the theory without mentioning the principle of optimality at all, but this would involve pulling the Bellman equation out of the hat, and readers would be left wondering – at least, we *hope* they would be left wondering – where it came from.

Theorem 6.1.1 (Verification theorem)

Suppose $V_{N-1}(x), V_{N-2}(x), \ldots, V_0(x)$ satisfy the Bellman equation (6.1.4) with terminal condition (6.1.7). Suppose that the minimum in (6.1.4) is achieved at $v = u_j^0(x)$, i.e.

$$\|D(j)x + F(j)u_j^0(x)\|^2 + V_{j+1}(A(j)x + B(j)u_j^0(x))$$
$$\leq \|D(j)x + F(j)v\|^2 + V_{j+1}(A(j)x + B(j)v)$$

for all m-vectors v. Now define (x_k^*, u_k^*) recursively as follows:

$$x_0^* = x_0 \tag{6.1.8}$$

$$\left.\begin{aligned} u_k^* &= u_k^0(x_k^*) \\ x_{k+1}^* &= Ax_k^* + Bu_k^* \end{aligned}\right\} \quad k = 0, 1, \ldots, N-1. \tag{6.1.9}$$

Then $u^* = (u_0^*, \ldots, u_{N-1}^*)$ is an optimal control and the minimum cost is $V_0(x_0)$.

PROOF Let $u = (u_0, \ldots, u_{N-1})$ be *any* control and $x_0, \ldots, x_N$ the corresponding trajectory, always with the same initial point x_0. Then from (6.1.4) we have

$$V_j(x_k) \leq \|D(j)x_j + F(j)u_j\|^2 + V_{j+1}(x_{j+1}). \tag{6.1.10}$$

Hence

$$V_N(x_N) - V_0(x_0) = \sum_{k=0}^{N-1} (V_{k+1}(x_{k+1}) - V_k(x_k))$$
$$\geq - \sum_{k=0}^{N-1} \|D(j)x_j + F(j)u_j\|^2. \tag{6.1.11}$$

Since $V_N(x_N) = x_N^{\mathrm{T}} Q x_N$ this shows that

$$V_0(x_0) \leq J_N(u). \tag{6.1.12}$$

On the other hand, by definition, equality holds in (6.1.10) and hence in (6.1.11) when $x_j = x_j^*$, $u_j = u_j^*$, so that

$$V_0(x_0) = J_N(u^*). \tag{6.1.13}$$

Now (6.1.12), (6.1.13) say that u^* is optimal and that the minimal cost is $V_0(x_0)$. □

Two remarks are in order at this point:

1. Note that the optimal control is obtained in *feedback form*, i.e. x_k^* is generated by

$$x_{k+1}^* = A(k)x_k^* + B(k)u_k^0(x_k^*)$$

where $u_k^0(\cdot)$ is a pre-determined function. (See Fig. 6.1(a).) One could in principle obtain the same cost $V_0(x_0)$ by calculating the u_k^* sequence

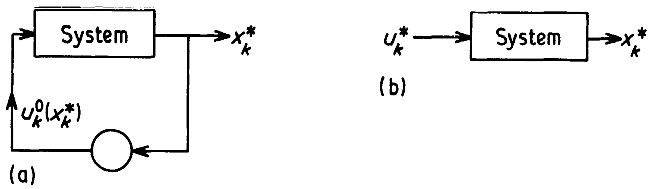


Fig. 6.1 (*a*) *Feedback control*; (*b*) *Open loop control.*

explicitly and applying it in open loop (Fig. 6.1(b)) but such a procedure has serious disadvantages. Using the dynamic programming approach, we have in fact not only solved the original overall control problem but have solved all the intermediate problems as well: an argument identical to that given above shows that the control u_k^* generated by (6.1.9) with any initial condition $x_j^* = x$ is optimal for the control problem over the last $N - j$ stages starting at $x_j = x$. Thus if for some reason the system gets 'off course' the feedback controller continues to act optimally for the remaining stages of control. On the other hand, the values u_k^* calculated for the open-loop control of Fig. 6.1(b) are based on a specific starting point x_0 and if this is erroneous or if an error occurs at some intermediate point then the u_k^* sequence will no longer be optimal.

2. Nothing so far depends on the quadratic nature of the cost function (6.1.2). Similar results would be obtained for any scalar cost function of the form

$$J'_N(u) = \sum_{k=0}^{N-1} l(k, x_k, u_k) + g(x_N). \qquad (6.1.14)$$

We have seen above that the basic step in solving the optimal control problem is to calculate the value functions $V_{N-1}(x), \ldots, V_0(x)$. With general cost functions $J'(u)$ as in (6.1.14) this involves an immense amount of work since the whole function $V_k(\cdot)$ has to be calculated and not just the value $V_k(x)$ at some specific point x. The advantage of the quadratic cost (6.1.2) is that the value functions take a simple parametric form and can be computed in an efficient way. Indeed, the value functions are themselves quadratic forms, as the following result shows.

Theorem 6.1.2

The solution of the Bellman equation (6.1.4), (6.1.7) for the linear regular problem (6.1.1), (6.1.2) is given by

$$V_k(x) = x^{\mathrm{T}} S(k) x \qquad k = 0, 1, \ldots, N \tag{6.1.15}$$

where $S(0), \ldots, S(N)$ are symmetric non-negative definite matrices defined by (6.1.20) below. The optimal feedback control is

$$u_j^1(x) = -M(j)x$$

where

$$\begin{aligned} M(j) = [&B^{\mathrm{T}}(j)S(j+1)B(j) + F^{\mathrm{T}}(j)F(j)]^{-1} \\ &\cdot [B^{\mathrm{T}}(j)S(j+1)A(j) + F^{\mathrm{T}}(j)D(j)]. \end{aligned} \tag{6.1.16}$$

We see that the optimal controller has a very simple structure, namely *linear feedback* of the state variables. The notation u_j^1 for optimal control is used for consistency with the discounted cost case to be discussed below.

PROOF Note that the result is certainly true at $k = N$ since $V_N(x) = x^{\mathrm{T}}Qx$. To show that it holds for $k < N$ we use backwards induction: supposing (6.1.15) holds for $k = j + 1$ we show that it holds for $k = j$. Taking $V_{j+1}(x) = x^{\mathrm{T}}S(j+1)x$, the Bellman equation (6.1.4) becomes

$$\begin{aligned} V_j(x) = \min_v [&\|D(j)x + F(j)v\|^2 + (x^{\mathrm{T}}A^{\mathrm{T}}(j) + v^{\mathrm{T}}B^{\mathrm{T}}(j)) \\ &\cdot S(j+1)(A(j)x + B(j)v)]. \end{aligned} \tag{6.1.17}$$

The quantity in square brackets on the right-hand side is equal to

$$\begin{aligned} &v^{\mathrm{T}}(B^{\mathrm{T}}S(j+1)B + F^{\mathrm{T}}F)v + 2x^{\mathrm{T}}(A^{\mathrm{T}}S(j+1)B + D^{\mathrm{T}}F)v \\ &\quad + x^{\mathrm{T}}(A^{\mathrm{T}}S(j+1)A + D^{\mathrm{T}}D)x \end{aligned} \tag{6.1.18}$$

where we temporarily write $B(j) = B$, etc. Now if R is a symmetric positive definite matrix and a an m-vector then

$$(v + a)^{\mathrm{T}}R(v + a) = v^{\mathrm{T}}Rv + 2a^{\mathrm{T}}Rv + a^{\mathrm{T}}Ra$$

i.e.

$$v^{\mathrm{T}}Rv + 2a^{\mathrm{T}}Rv = (v + a)^{\mathrm{T}}R(v + a) - a^{\mathrm{T}}Ra.$$

Clearly this expression is minimized over v at $v = -a$ and the minimum value is $-a^{\mathrm{T}}Ra$. In order to identify this with the first two terms in (6.1.18) we require

$$R = B^{T}S(j+1)B + F^{T}F$$
$$Ra = (B^{T}S(j+1)A + F^{T}D)x.$$

Now by assumption $F^{T}F$, and hence R, is strictly positive definite, and therefore a is specified by

$$a = R^{-1}(B^{T}S(j+1)A + F^{T}D)x.$$

Thus the right-hand side of (6.1.17) is equal to

$$x^{T}[A^{T}S(j+1)A + D^{T}D - (A^{T}S(j+1)B + D^{T}F) \\ R^{-1}(B^{T}S(j+1)A + F^{T}D)]x. \tag{6.1.19}$$

Hence $V_j(x) = x^{T}S(j)x$ where $S(j)$ is given by the expression in the square brackets in (6.1.19) and $S(j) \geq 0$ by (6.1.17). Thus $V_k(x)$ is a quadratic form, as in (6.1.15), for all $k = 0, 1, \ldots, N$. Note from the above analysis (specifically from (6.1.19)) that the matrices $S(k)$ can be computed recursively backwards in time starting with $S(N) = Q$. In fact, writing out (6.1.19) in full we see that the $S(k)$ are generated by

$$\begin{aligned} S(N) &= Q \\ S(j) &= A^{T}(j)S(j+1)A(j) + D^{T}(j)D(j) - (A^{T}(j)S(j+1)B(j) \\ &\quad + D^{T}(j)F(j))(B^{T}(j)S(j+1)B(j) + F^{T}(j)F(j))^{-1} \\ &\quad \cdot(B^{T}(j)S(j+1)A(j) + F^{T}(j)D(j)) \\ j &= N-1, N-2, \ldots, 0. \end{aligned} \tag{6.1.20}$$

Applying the dynamic programming results, the optimal feedback control is the value of v that achieves the minimum in (6.1.16), and this is equal to $-a$, so that

$$u_j^1(x) = -[B^{T}(j)S(j+1)B(j) + F^{T}(j)F(j)]^{-1} \\ \cdot [B^{T}(j)S(j+1)A(j) + F^{T}(j)D(j)]x.$$

This completes the proof. □

Filtering/control duality

A very important feature of the above result is its close connection to the Kalman filter discussed in Section 3.3. Equation (6.1.20) is a Riccati equation of exactly the same type as that appearing in the Kalman filter equations, with the distinction that (6.1.20) evolves backwards from a terminal condition at time N whereas the filtering Riccati equation (3.3.6) for the estimation error covariance $P(j)$ evolves forward from an initial condition at $j = 0$. The Kalman gain $K(j)$ is related to $P(j)$ in exactly the same way that the control gain $M(j)$ is related to $S(j)$, except for transposition. Specifically, the correspondence between the two problems is as shown in Table 6.1.

Table 6.1

Filtering	*Control*
(time) j	$N - j$
$A(j)$	$A^{\mathrm{T}}(j)$
$H(j)$	$B^{\mathrm{T}}(j)$
$C(j)$	$D^{\mathrm{T}}(j)$
$G(j)$	$F^{\mathrm{T}}(j)$
$P(j)$	$S(j)$
$K(j)$	$M^{\mathrm{T}}(j)$

This means that if we take the filtering Riccati equation (3.3.6), make the time substitution $j \to N - j$ and relabel A, H, C, G as $A^{\mathrm{T}}, B^{\mathrm{T}}, D^{\mathrm{T}}, F^{\mathrm{T}}$ respectively, then we get precisely (6.1.20). The same relabelling applied to the expression (3.3.5) for $K(j)$ produces $M^{\mathrm{T}}(j)$. Thus the Riccati equations (6.1.20) and (3.3.6) are the same in all but notation. This will be very important when we come to consider various properties of the Riccati equation, since its solution can be regarded interchangeably as the value function for a control problem or the error covariance for a filtering problem, and various facts can be deduced from one or other of these interpretations.

Discounted costs

Let us now specialize to the time-invariant system

$$x_{k+1} = Ax_k + Bu_k \tag{6.1.21}$$

(i.e. $A(k) = A$, $B(k) = B$ for all k) and consider minimizing a discounted cost of the form

$$J_N^{\rho}(u) = \sum_{k=0}^{N-1} \rho^k \|Dx_k + Fu_k\|^2 + \rho^N x_N^{\mathrm{T}} Q x_N \tag{6.1.22}$$

where D, F, Q are fixed matrices and ρ is the discount factor $(0 < \rho \leq 1)$. This is actually a special case of the preceding problem (take

$D(k) = \rho^{k/2}D$, $F(k) = \rho^{k/2}F$ and replace Q by $\rho^N Q$); but there is another way of looking at it which provides a little more insight. Write

$$\begin{aligned} J_N^\rho(u) &:= \sum_{k=0}^{N-1} \rho^k \|Dx_k + Fu_k\|^2 + \rho^N x_N^{\mathrm{T}} Q x_N \\ &= \sum_{k=0}^{N-1} \|D\rho^{k/2}x_k + F\rho^{k/2}u_k\|^2 + \rho^N x_N^{\mathrm{T}} Q x_N \\ &= \sum_{k=0}^{N-1} \|Dx_k^\rho + Fu_k^\rho\|^2 + x_N^{\rho\mathrm{T}} Q x_N^\rho \end{aligned} \tag{6.1.23}$$

where we have defined

$$\begin{aligned} x_k^\rho &:= \rho^{k/2} x_k \\ u_k^\rho &:= \rho^{k/2} u_k. \end{aligned} \tag{6.1.24}$$

Multiplying (6.1.21) by $\rho^{(k+1)/2}$ gives

$$\rho^{(k+1)/2} x_{k+1} = \rho^{1/2} A \rho^{k/2} x_k + \rho^{1/2} B \rho^{k/2} u_k$$

i.e.

$$x_{k+1}^\rho = A^\rho x_k^\rho + B^\rho u_k^\rho \tag{6.1.25}$$

where $A^\rho := \rho^{1/2}A$, $B^\rho := \rho^{1/2}B$. But (6.1.23)–(6.1.25) constitute a time-invariant linear regulator problem in standard non-discounted form. The optimal control is therefore

$$\begin{aligned} u_k^\rho &= -(B^{\rho\mathrm{T}} S^\rho(k+1) B^\rho + F^{\mathrm{T}}F)^{-1}(B^{\rho\mathrm{T}} S^\rho(k+1) A^\rho + F^{\mathrm{T}}D) x_k^\rho \\ &=: -M^\rho(k) x_k^\rho \end{aligned}$$

where $S^\rho(k)$ is the solution of (6.1.20) with A replaced by $\rho^{1/2}A$ and B replaced by $\rho^{1/2}B$. In view of (6.1.24) the optimal control u_k is expressed in terms of the 'real' state x_k by

$$u_k = -M^\rho(k) x_k.$$

Thus the discounted cost problem is solved simply by taking the *undiscounted* problem and making the substitutions $A \to \rho^{1/2}A$, $B \to \rho^{1/2}B$.

6.1.2 Infinite-time problems

In this section we will continue to assume that the system and costs are time-invariant, i.e. the matrices A, B, D, F do not depend on the time, k.

In many control problems no specific terminal time N is involved and one wishes the system to have good 'long-run' performance. This suggests replacing (6.1.2) by a cost

$$J_\infty(u) = \sum_{k=0}^{\infty} \|Dx_k + Fu_k\|^2. \qquad (6.1.26)$$

It is not obvious that the problem of minimizing $J_\infty(u)$ subject to the dynamics (6.1.1) makes sense: it might be the case that $J_\infty(u) = +\infty$ for all controls u. Note, however, that the problem *does* make sense as long as there is at least one control u such that $J_\infty(u) < \infty$. A simple sufficient condition for this is that the pair (A, B) be *stabilizable*, i.e. there exists an $m \times n$ matrix M such that $A - BM$ is stable. Taking for u the feedback control $\bar{u}_k = -Mx_k$, the system dynamics become

$$x_{k+1} = (A - BM)x_k.$$

Now since $A - BM$ is stable, it follows from Proposition D.3.1, Appendix D, that there exist constants $c > 0$ and $a \in (0, 1)$ such that

$$\|x_k\| \leq ca^k \|x_0\|$$

Since $\|(D - FM)x\| \leq K\|x\|$ for some constant K, the cost using control $\bar{u}$ is

$$\begin{aligned} J_\infty(\bar{u}) &= \sum_{k=0}^{\infty} \|(D - FM)x_k\|^2 \\ &\leq K^2 \sum_{k=0}^{\infty} \|x_k\|^2 \\ &\leq c^2K^2 \|x_0\|^2 \sum_{k=0}^{\infty} a^{2k} \\ &= c^2K^2 \|x_0\|^2/(1 - a^2). \end{aligned}$$

Thus with any stabilizing control, the norm of x_k decays sufficiently fast to give a finite total cost. We will therefore assume henceforth that the pair (A, B) is stabilizable.

If $V_k(x)$ is the value function at time k for the infinite-time problem then it seems likely that V_k does not actually depend on k, since, there being no 'time horizon' and the coefficients being time-invariant, the problem facing the controller is the same at time k as at time zero, except for some change in the initial state. Recalling the Bellman equation (6.1.4), this suggests that the value function $V \equiv V_k$ should satisfy

$$V(x) = \min_v [\|Dx + Fv\|^2 + V(Ax + Bv)]. \qquad (6.1.27)$$

Note this is no longer a recursion but is an implicit equation which may or may not be satisfied by a particular function V.

Proposition 6.1.3

Suppose that V is a solution of (6.1.27) such that V is continuous and $V(0) = 0$,† and that $u^1(x)$ achieves the minimum on the right, i.e. for all vectors v,

$$\| Dx + Fu^1(x) \|^2 + V(Ax + Bu^1(x)) \le \| Dx + Fv \|^2 + V(Ax + Bv).$$

Suppose also that u^1 is a stabilizing control in the sense that $\| x_k \| \to 0$ as $k \to \infty$, where x_k is the trajectory corresponding to u^1, i.e.

$$x_{k+1} = Ax_k + Bu^1(x_k).$$

Then $u^1(x)$ is optimal in the class of stabilizing controls. Equation (6.1.27) has the quadratic solution $V(x) = x^{\mathrm{T}}Sx$ if and only if S satisfies the algebraic Riccati equation (6.1.29) below, and in this case the corresponding control is

$$u^1(x) = -Mx$$

where

$$M = (B^{\mathrm{T}}SB + F^{\mathrm{T}}F)^{-1}(B^{\mathrm{T}}SA + F^{\mathrm{T}}D) \tag{6.1.28}$$

PROOF Let $\{x_k, u_k\}$ be any control/trajectory pair such that $\| x_k \| \to 0$ as $k \to \infty$ and write

$$\begin{aligned} V(x_N) - V(x_0) &= \sum_{k=0}^{N-1} V(x_{k+1}) - V(x_k) \\ &\ge \sum_{0}^{N-1} \|Dx_k + Fu_k\|^2 \qquad \text{(from (6.1.27))}. \end{aligned}$$

Thus

$$V(x_0) \le \sum_{k=0}^{N-1} \|Dx_k + Fu_k\|^2 + V(x_N).$$

Now by the assumptions on V and x_k, $V(x_N) \to 0$ as $N \to \infty$ and hence

$$V(x_0) \le \sum_{k=0}^{\infty} \|Dx_k + Fu_k\|^2 = J_\infty(u).$$

The same calculations hold with $=$ replacing $\ge$ when $u = u^1$, and this

†A natural requirement since if $x = 0$ the control $u_k = 0$ is plainly optimal.

shows that

$$V(x_0) = J_\infty(u^1) = \min_u J_\infty(u).$$

Thus u^1 is optimal in the class of stabilizing controls (those for which $\|x_k\| \to 0$ as $k \to \infty$, x_k being the corresponding trajectory.)

Since the value function for the finite-horizon problem is a quadratic form, let us try a solution to (6.1.27) of the form $x^T S x$ where S is a symmetric non-negative definite matrix. From (6.1.19), the minimum value on the right of (6.1.27) is then

$$x^T[A^T SA + D^T D - (A^T SB + D^T F)(B^T SB + F^T F)^{-1}(B^T SA + F^T D)]x$$

and $V(x) = x^T Sx$ is therefore a solution of (6.1.27) if and only if S satisfies the so-called *algebraic Riccati equation* (ARE):

$$S = A^T SA + D^T D - (A^T SB + D^T F)(B^T SB + F^T F)^{-1}(B^T SA + F^T D). \tag{6.1.29}$$

If S satisfies this then certainly $V(x) = x^T Sx$ is continuous and $V(0) = 0$. The corresponding minimizing u^1 is given as before by

$$u^1(x) = -Mx$$

where

$$M = (B^T SB + F^T F)^{-1}(B^T SA + F^T D).$$ □

If the matrix $A - BM$ is stable then $\|x_k\| \to 0$ as $k \to \infty$ where

$$x_{k+1} = Ax_k + Bu^1(x_k) = (A - BM)x_k.$$

The above proof thus shows that if S satisfies (6.1.29) and $A - BM$ is stable then the control $u^1(x_k) = -Mx_k$ is optimal in the class of all stabilizing controls. An important feature of this result is that the optimal control is *time-invariant* (does not depend explicitly on k), although time varying controls are not in principle excluded.

It is evident from Proposition 6.1.3 that the infinite time problem hinges on properties of the algebraic Riccati equation. These are somewhat technical and a full account will be found in Appendix B. Let us summarize the main results. The conditions required on the coefficient matrices A, B, D, F are as follows:

(a) The pair (A, B) is stabilizable.
(b) The pair $(\hat{D}, \hat{A})$ is detectable, where (6.1.30)

$$\hat{A} = A - B(F^T F)^{-1}F^T D$$
$$\hat{D} = [I - F(F^T F)^{-1}F^T]D.$$

The first of these conditions is a natural one since, as remarked before, it ensures the existence of at least one control giving finite cost. The motivation for condition (b) is less obvious, though it does seem clear that *some* condition involving D and F, in particular concerning the relation between states x_k and 'output' Dx_k, is required to justify limiting attention to stabilizing controls. Condition (b) takes the simpler form

(b′) (D, A) is detectable,

when $F^{\mathrm{T}}D = 0$; this is the case alluded to at the beginning of this section, in which the cost takes the form

$$\|Dx_k + Fu_k\|^2 = x_k^{\mathrm{T}}D^{\mathrm{T}}Dx_k + u_k^{\mathrm{T}}F^{\mathrm{T}}Fu_k.$$

Under conditions (6.1.30), the argument given in Appendix B shows that there is a unique non-negative definite matrix S satisfying the algebraic Riccati equation, that $A - BM$ is stable, where M is given by (6.1.28), and that the control $u^1(x_k) = -Mx_k$ is optimal in the sense of minimizing $J_\infty(u)$ over *all* control–trajectory pairs (x_k, u_k) satisfying the dynamic equation (6.1.1). (The less precise argument summarized in Proposition 6.1.3 only shows that $u^1(x)$ minimizes $J_\infty(u)$ over all such pairs satisfying $\|x_k\| \to 0$ as $k \to \infty$.)

The relation between the finite and infinite-time problems is also elucidated in Appendix B. In fact it is shown that under conditions (a) and (b),

$$S = \lim_{k \to \infty} S(-k) \tag{6.1.31}$$

where $S(-1), S(-2), \ldots$ is the sequence of matrices produced by the Riccati equation (6.1.19) with $S(0) = Q$ where Q is an arbitrary non-negative definite matrix. Now $x^{\mathrm{T}}S(-k)x$ is the minimal cost for the k-stage control problem (6.1.1)–(6.1.2) with terminal cost $x_k^{\mathrm{T}}Qx_k$. In view of (6.1.31) we see that as the time horizon recedes to infinity, the cost of the finite-horizon problem approaches that of the infinite horizon problem, whatever the terminal cost matrix Q. Q is unimportant because $\| x_k \|$ will be very small for large k when the optimal control is applied.

Generally, in the finite-horizon case, the optimal control $u_k = -M(k)x_k$ is time-varying. If, however, one selects $Q = S$ as the terminal cost, where S satisfies the algebraic Riccati equation, then $S(k) = S$ for all k, so that the time-invariant control $u_k = -Mx_k$ is optimal, and this is the same control that is optimal for the infinite-horizon problem. The situation is somewhat analogous to that of a transmission line terminated by a matched impedance. With this termination the line is indistinguishable from one of infinite length. In the control case, if the terminal cost is $x_k^{\mathrm{T}} S x_k$ the controller is indifferent between paying it and stopping, or continuing optimally *ad infinitum*. In either case the total cost is the same, so it is reasonable to describe S as the 'matched' terminal cost matrix.

Finally, let us consider the infinite-time discounted cost problem, where the cost function is

$$J_\infty^\rho(u) = \sum_{k=0}^{\infty} \rho^k \|Dx_k + Fu_k\|^2.$$

Proceeding exactly as in the finite-horizon discounted case, we conclude that the optimal control is

$$u_k^\rho(x_k) = -M^\rho x_k.$$

Here

$$M^\rho = (B^{\rho\mathrm{T}} S^\rho B^\rho + F^{\mathrm{T}} F)^{-1}(B^{\rho\mathrm{T}} S^\rho A^\rho + F^{\mathrm{T}} D)$$

and S^ρ is the solution of the algebraic Riccati equation with A and B replaced by A^ρ and B^ρ respectively, where

$$A^\rho = \rho^{1/2} A, \qquad B^\rho = \rho^{1/2} B.$$

The conditions for existence of a solution S^ρ to the modified equation are the appropriately modified version of (6.1.30) above, namely

(c) (A^ρ, B^ρ) is stabilizable.
(d) $(\hat{D}, \hat{A}^\rho)$ is detectable $(\hat{A}^\rho = \rho^{1/2}\hat{A})$. (6.1.32)

Note that if U is any $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ then the eigenvalues of $\rho^{1/2}U$ are $\rho^{1/2}\lambda_1, \ldots, \rho^{1/2}\lambda_n$ since if x_i is an eigenvector corresponding to λ_i then

$$\rho^{1/2} U x_i = \rho^{1/2} \lambda_i x_i. \tag{6.1.33}$$

Thus $A^\rho - B^\rho M = \rho^{1/2}(A - BM)$ is stable if $A - BM$ is stable. Similarly $\hat{A}^\rho - (\rho^{1/2}N)\hat{D} = \rho^{1/2}(\hat{A} - N\hat{D})$ is stable if $\hat{A} - N\hat{D}$ is stable. Thus conditions (6.1.30) imply conditions (6.1.32), so that S^ρ exists for any $\rho \le 1$ if conditions (6.1.30) are satisfied. However, taking $U = A$ and $U = \hat{A}$ in (6.1.33) we see that, for sufficiently small ρ, A^ρ and $\hat{A}^\rho$ are both stable and, *a fortiori*, (A^ρ, B^ρ) and $(\hat{D}, \hat{A}^\rho)$ are stabilizable and detectable respectively. Thus an optimal solution to the discounted cost infinite-time problem *always* exists if the discount factor ρ sufficiently small. An optimal control with finite cost can, however, be obtained without discounting if the rather mild conditions (6.1.30) are met. This contrasts with the situation in the stochastic case considered in the next section, where discounting is always necessary to obtain finite costs in infinite-time problems.

This concludes our discussion of the deterministic optimal regulator problem. We need it as a stepping-stone to the stochastic case and also to isolate the duality relationships which connect the Riccati equations which arise here and in the Kalman filter. In Appendix B, the asymptotic behaviour of the Riccati equation is investigated by methods which rely heavily on its control-theoretic interpretation. But, thanks to the duality properties, these results apply equally to tell us something about asymptotic behaviour of the estimation error in the Kalman filter.

In recent years, techniques based on the linear/quadratic optimal regulator have become an important component of multivariable control system design methodology. It is outside the scope of this book to discuss such questions, but some references will be found in the Notes and References at the end of this chapter. The essential advantage of the linear/quadratic framework in this connection is that arbitrary dimensions m and p of input u_k and output Dx_k are allowed, whereas techniques which attempt to generalize the classical single-input, single-output methods are seriously complicated by the combinatorial fact that there are rp transfer functions to consider, one from each input to each output. A subsidiary advantage of the linear/quadratic framework is that time-varying systems are handled with relative ease.

6.2 The stochastic linear regulator

In this section we consider problems of optimal regulation when the state equation includes additive noise, as in the state-space stochastic model discussed in Section 2.4. Thus x_k satisfies

$$x_{k+1} = A(k)x_k + B(k)u_k + C(k)w_k \tag{6.2.1}$$

where $\{w_k\}$ is a sequence of l-vector random variables with mean 0 and covariance I. We will assume in this section that w_k and w_j are *independent* (rather than merely uncorrelated) for $k \neq j$. The initial state x_0 is a random vector independent of w_k with mean and covariance m_0, P_0 respectively. We suppose that the state x_k can be measured directly by the controller, so that controls will be feedback functions of the form $u_k = u_k(x_k)$. The objective is to minimize the cost criterion

$$C_N(u) = E\left[\sum_{k=0}^{N-1} \|D(k)x_k + F(k)u_k\|^2 + x_N^T Q x_N\right].$$

The value function $W_j(x)$ at time j for this problem is the minimum value of

$$E_{j,x}\left[\sum_{k=j}^{N-1} \|D(k)x_k + F(k)u_k\|^2 + x_N^T Q x_N\right]$$

where $E_{j,x}$ denotes the expectation given that the process starts off at $x_j = x$ (a fixed vector in $\mathbb{R}^n$). If $x_j = x$ and the control value $u_j = v$ is applied then the next state is

$$x_{j+1} = A(j)x + B(j)v + C(j)w_j$$

and, by definition, the minimal remaining cost for the rest of the problem from time $j + 1$ to N is $W_{j+1}(x_{j+1})$. This, however, is now a random variable since x_{j+1} is determined partly by w_j. The *expected* minimal remaining cost is obtained by averaging this over the distribution of w_j, giving a value of

$$EW_{j+1}(A(j)x + B(j)v + C(j)w_j).$$

Thus the minimum expected cost starting at $x_j = x$, if control $u_j = v$ is used, is the sum of this and the cost $\|D(j)x + F(j)v\|^2$ paid at time j. This suggests that $W_j(x)$ should satisfy the stochastic Bellman equation

$$W_j(x) = \min_v [\|D(j)x + F(j)v\|^2 + EW_{j+1}(A(j)x + B(j)v + C(j)w_j)] \tag{6.2.2}$$

where again E means averaging over the distribution of w_j with x, v fixed. At the final time N no further control or noise enters the system, so that

$$W_N(x) = x^T Q x. \tag{6.2.3}$$

As before, (6.2.2)–(6.2.3) determine a sequence of functions $W_N, W_{N-1}, \ldots, W_0$ by backwards recursion. And, also as before, we do not rely on the above heuristic argument to conclude that these functions are indeed the value functions for the control problem, but provide independent direct verification.

Proposition 6.2.1

Suppose that $W_N, \ldots, W_0$ are given by (6.2.2), (6.2.3) and that $u_j^0(x)$ is the value of v that achieves the minimum in (6.2.2). Then the feedback control $u_k^* = u_k^1(x_k)$ minimizes the cost $C_N(u)$ over the class of all feedback control policies.

PROOF Let $u_k(x_k)$ be an arbitrary feedback control and let x_k be the process given by (6.2.1) with $u_k = u_k(x_k)$. Then

$$W_N(x_N) - W_0(x_0) = \sum_{k=0}^{N-1} (W_{k+1}(x_{k+1}) - W_k(x_k))$$

so that

$$E[W_N(x_N) - W_0(x_0)] = \sum_{k=0}^{N-1} E[W_{k+1}(x_{k+1}) - W_k(x_k)] \quad (6.2.4)$$

In calculating the expectations on the right we are entitled to introduce any intermediate conditional expectation. We therefore write

$$E[W_{k+1}(x_{k+1}) - W_k(x_k)] = E\{E[W_{k+1}(x_{k+1}) - W_k(x_k)|x_k]\}. \quad (6.2.5)$$

Now, given x_k, $W_k(x_k)$ is known and x_{k+1} is given by

$$x_{k+1} = A(k)x_k + B(k)u_k(x_k) + C(k)w_k.$$

The first two terms on the right are known and the third is a random vector independent of x_k. The conditional expectation of $W_{k+1}(x_{k+1})$ is therefore given by

$$E[W_{k+1}(x_{k+1})|x_k] = EW_{k+1}(A(k)x_k + B(k)u_k(x_k) + C(k)w_k)$$

where the expectation on the right is taken over the distribution of w_k for fixed x_k. Now, using (6.2.2) we obtain

$$\begin{aligned} E[W_{k+1}(x_{k+1}) - W_k(x_k)|x_k] &= EW_{k+1}(A(k)x_k + B(k)u_k(x_k) \\ &\quad + C(k)w_k) - W_k(x_k) \\ &\geq -\|D(k)x_k + F(k)u_k(x_k)\|^2. \end{aligned} \quad (6.2.6)$$

Combining (6.2.4)–(6.2.6) shows that

$$E[W_N(x_N) - W_0(x_0)] \geq -E\sum_{k=0}^{N-1} \|D(k)x_k + F(k)u_k(x_k)\|^2$$

and hence, since $W_N(x_N) = x_N^T Q x_N$, that

$$EW_0(x_0) \le C_N(u). \tag{6.2.7}$$

On the other hand, the same argument holds with equality instead of inequality in (6.2.6) when $u_k(x) = u_k^1(x)$, so that

$$EW_0(x_0) = C_N(u^1). \tag{6.2.8}$$

Now (6.2.7) and (6.2.8) say that u^1 is optimal. □

The proof actually shows a little more than is claimed in the proposition statement. Indeed, since $W_0(x_0)$ is only a function of x_0, the expectation in (6.2.8) only involves the (arbitrary) distribution of the initial state x_0. In particular, if x_0 takes a fixed value, say $\bar{x}_0$, with probability one, then the corresponding optimal cost is just $W_0(\bar{x}_0)$. Thus $W_0(x_0)$ should be interpreted as the *conditional* optimal cost given the initial state x_0. The *overall* optimal cost is then obtained by averaging over x_0, as in (6.2.8). A similar interpretation applies to W_k, namely $W_k(x)$ is the optimal cost over stages $k, k+1, \ldots, N$ conditional on an initial state $x_k = x$.

The solution of (6.2.2) is related in a simple way to that of the 'deterministic' Bellman equation (6.1.4). In fact,

$$W_k(x) = x^T S(k) x + \alpha_k$$

where $S(N) = Q$, $S(N-1), \ldots, S(0)$ are given by the Riccati equation (6.1.20) as before, and α_k is a constant, to be determined below. Note that if $W_{k+1}(x) = x^T S(k+1)x + \alpha_{k+1}$ then for fixed x, v,

$$\begin{aligned}
&EW_{k+1}(A(k)x + B(k)v + C(k)w_k) \\
&= (A(k)x + B(k)v)^T S(k+1)(A(k)x + B(k)v) \\
&\quad + 2E(A(k)x + B(k)v)^T S(k+1)C(k)w_k \\
&\quad + Ew_k^T C^T(k)S(k+1)C(k)w_k + \alpha_{k+1} \\
&= (A(k)x + B(k)v)^T S(k+1)(A(k)x + B(k)v) \\
&\quad + \operatorname{tr}[C^T(k)S(k+1)C(k)] + \alpha_{k+1}
\end{aligned}$$

where the last line follows from the facts that $Ew_k = 0$, $\operatorname{cov}(w_k) = I$. Notice that the final expression is identical to that obtained in the deterministic case except for the term $\operatorname{tr}[C^T(k)S(k+1)C(k)] + \alpha_{k+1}$, which does not depend on x or v and hence does not affect the minimization on the right-hand side of (6.2.2). Thus if $W_{k+1}(x) = x^T S(k+1)x + \alpha_{k+1}$ then the induction argument as used in the deterministic case shows that

$$W_k(x) = x^{\mathrm{T}}S(k)x + \alpha_{k+1} + \mathrm{tr}[C^{\mathrm{T}}(k)S(k+1)C(k)].$$

But $W_N(x) = x^{\mathrm{T}}Qx$, i.e. $\alpha_N = 0$, so working backwards from $k = n$ we see that

$$\alpha_k = \sum_{j=k}^{N-1} \mathrm{tr}[C^{\mathrm{T}}(j)S(j+1)C(j)].$$

Summarizing, we have the following result.

Theorem 6.2.2

For the stochastic linear regulator with complete observations, the optimal control is

$$u_k^1(x_k) = -M(k)x_k$$

where $M(k)$ is given by (6.1.16), i.e. is the same as in the deterministic case. The minimal cost is

$$C_N(u^0) = m_0^{\mathrm{T}}S(0)m_0 + \mathrm{tr}[S(0)P_0] + \sum_{k=0}^{N-1} \mathrm{tr}[C^{\mathrm{T}}(k)S(k+1)C(k)]. \tag{6.2.9}$$

PROOF The optimality of u^1 follows from Proposition 6.2.1. As to the cost, we note that

$$W_0(x) = x^{\mathrm{T}}S(0)x + \alpha_0$$

is the conditional minimal cost given that the process starts at $x_0 = x$. Taking the expectation over the distribution of x_0, and using Proposition 1.1.3(b), we obtain (6.2.9). □

Note that only the mean m_0 and covariance P_0 of the initial state are needed to compute the optimal cost, so it is not necessary to suppose that x_0 is normally distributed. The important feature of the above result is that the matrices $S(k)$ and $M(k)$ do not depend on the noise coefficients $C(k)$, so that in particular *the optimal control is the same as in the deterministic case.* Thus adding noise to the state equation as in (6.2.1) makes no difference to the optimal policy, but simply makes that policy more expensive. Indeed, if the system starts at a fixed state x_0 (so that $m_0 = x_0$ and $P_0 = 0$) then the additional cost is precisely

$$\sum_{k=0}^{N-1} \mathrm{tr}[C^{\mathrm{T}}(k)S(k+1)C(k)].$$

Let us now consider the *discounted cost case.* We will assume for simplicity of notation that the coefficient matrices A, B, D, F are time invariant but, with later applications in mind, time variation will be retained for $C(k)$. Thus the problem is to minimize

$$E\left(\sum_{k=0}^{N-1} \rho^k \|Dx_k + Fu_k\|^2 + \rho^N x_N^{\mathrm{T}} Q x_N\right).$$

We use the same device as before, namely rewriting the cost as

$$E\left(\sum_{k=0}^{N-1} \|Dx_k^\rho + Fu_k^\rho\|^2 + x_N^{\rho\mathrm{T}} Q x_N^\rho\right) \tag{6.2.10}$$

where $x_k^\rho = \rho^{k/2} x_k$, $u_k^\rho = \rho^{k/2} u_k$. Multiplying (6.2.1) by $\rho^{(k+1)/2}$ shows that x_k^ρ, u_k^ρ satisfy

$$x_{k+1}^\rho = A^\rho x_k^\rho + B^\rho u_k^\rho + C^\rho(k) w_k \tag{6.2.11}$$

where $A^\rho = \rho^{1/2} A$, $B^\rho = \rho^{1/2} B$, $C^\rho(k) = \rho^{(k+1)/2} C(k)$. Now (6.2.10) and (6.2.11) give the problem in non-discounted form. As noted above, the optimal control does not depend on $C^\rho(k)$; applying our previous results it is given by

$$u_k^\rho(x) = -M^\rho(k)x$$

where $M^\rho(k)$ is defined as in Section 6.1 above. The corresponding cost is, from (6.2.9)

$$\begin{aligned} C_N^\rho(u^\rho) &= m_0^{\mathrm{T}} S^\rho(0) m_0 + \mathrm{tr}[S^\rho(0) P_0] + \sum_{k=0}^{N-1} \mathrm{tr}[C^{\rho\mathrm{T}}(k) S^\rho(k+1) C^\rho(k)] \\ &= m_0^{\mathrm{T}} S^\rho(0) m_0 + \mathrm{tr}[S^\rho(0) P_0] + \sum_{k=0}^{N-1} \rho^{k+1} \mathrm{tr}[C^{\mathrm{T}}(k) S^\rho(k+1) C(k)]. \end{aligned}$$

The importance of the discount factor becomes apparent when we consider infinite-horizon problems. Suppose that conditions (6.1.30) are met and that S^ρ is the solution to the algebraic Riccati equation with coefficient matrices A^ρ, B^ρ. Such a solution exists for any $\rho \leq 1$. Now consider the N-stage problem as above, with terminal cost matrix $Q = S^\rho$. This is the 'matched impedance' case, discussed at the end of Section 6.1, for which $S^\rho(k) = S^\rho$ for all k. Thus the optimal control is the time-invariant feedback

$$u^\rho(x_k) = -M^\rho x_k \tag{6.2.12}$$

and the cost over N stages is

$$C_N^\rho(u^\rho) = m_0^{\mathrm{T}} S^\rho m_0 + \mathrm{tr}[S^\rho P_0] + \sum_{k=0}^{N-1} \rho^{k+1}\, \mathrm{tr}[C^{\mathrm{T}}(k) S^\rho C(k)]. \tag{6.2.13}$$

Note that if $\rho = 1$ (no discounting) and $C(k) \equiv C$ is constant, then $C_N^\rho \to \infty$ as $N \to \infty$ and hence the infinite-time problem has no solution (all controls give cost $+\infty$). This is not surprising. The reason that finite costs could be obtained in the deterministic case was that $\|x_k\|$ converged to zero sufficiently fast that

$$\sum_{k=0}^{\infty} \|x_k\|^2$$

was finite. However, in the present case $\|x_k\|$ does *not* converge to zero because at each stage it is being perturbed by the independent noise term Cw_k, and the controller has continually to battle against this disturbance to keep $\|x_k\|$ as small as possible. If, however, $\rho < 1$, then

$$\lim_{N \to \infty} C_N^\rho = m_0^{\mathrm{T}} S^\rho m_0 + \mathrm{tr}[S^\rho P_0] + \frac{\rho}{1-\rho}\mathrm{tr}[C^{\mathrm{T}} S^\rho C]. \tag{6.2.14}$$

Thus any amount of discounting, however little, leads to a finite limiting cost. One can show, by methods exactly analogous to those used in the previous section, that the time-invariant control u^ρ given by (6.2.12) does in fact minimize the cost

$$C_\infty^\rho(u) = E\left(\sum_{k=0}^{\infty} \rho^k \|Dx_k + Fu_k\|^2\right) \tag{6.2.15}$$

and that the minimal cost is precisely the expression given in (6.2.14). As to the conditions required, recall that if (A, B) is stabilizable then (A^ρ, B^ρ) is stabilizable for any $\rho \le 1$; thus

(a) If conditions (6.1.30) are satisfied then the infinite time discounted problem is well-posed, and has the above solution, for any $\rho < 1$.
(b) If either of conditions (6.1.30) fails then we must take $\rho < \rho_0$ where ρ_0 is such that (A^ρ, B^ρ), $(\hat{D}, \hat{A}^\rho)$ are stabilizable and detectable respectively for any $\rho < \rho_0$. Generally, $\rho_0 < 1$.

If $C(k)$ is not constant then exactly similar results apply as long as

$$\sum_{k=0}^{\infty} \rho^{k+1} \operatorname{tr}[C^{\mathrm{T}}(k)S^{\rho}C(k)] < \infty$$

and this will certainly be the case for any $\rho < 1$ as long as the elements of $C^{\mathrm{T}}(k)$ are uniformly bounded, i.e. there is some constant c_1 such that for all i, j, k,

$$|C(k)_{ij}| \leq c_1.$$

This, in turn, is always true if the $C(k)$ sequence is convergent, i.e. there is a matrix C such that $C(k) \to C$ as $k \to \infty$. The same control is optimal but there is in general no closed-form expression, as in (6.2.14), for the minimal cost, which is now

$$m_0^{\mathrm{T}} S^{\rho} m_0 + \operatorname{tr}[S^{\rho} P_0] + \sum_{k=0}^{\infty} \rho^{k+1} \operatorname{tr}[C^{\mathrm{T}}(k) S^{\rho} C(k)]. \qquad (6.2.16)$$

Let us now consider minimizing the *average cost per unit time*,

$$C_{\mathrm{av}}(u) = \lim_{N \to \infty} \frac{1}{N} E\left[\sum_{k=0}^{N-1} \|Dx_k + Fu_k\|^2\right]. \qquad (6.2.17)$$

As before we assume that all coefficients are constant except for the noise matrices $C(k)$ which are supposed to be convergent: $C(k) \to C$ as $k \to \infty$. This is needed in the next section.

The limit in (6.2.17) may or may not exist for any particular control u, but it certainly does exist for all *constant, stabilizing* controls, i.e. controls of the form $u_k^K = -Kx_k$ where $\bar{A} := A - BK$ is stable. For then the closed-loop system is

$$x_{k+1} = \bar{A}x_k + C(k)w_k$$

and we know by a slight extension of results in Section 2.4 that $Q(k) := \operatorname{cov}(x_k) \to Q$ where Q satisfies

$$Q = \bar{A}Q\bar{A}^{\mathrm{T}} + CC^{\mathrm{T}}.$$

Thus

$$\begin{aligned} C_{\mathrm{av}}(u^K) &= \lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N} \operatorname{tr}[(D - FK)Q(k)(D - FK)^{\mathrm{T}}] \\ &= \operatorname{tr}[(D - FK)Q(D - FK)^{\mathrm{T}}]. \end{aligned}$$

If the pair (A, B) is stabilizable then a stabilizing K exists and the problem of minimizing $C_{av}(u)$ is meaningful. We now show that $C_{av}(u)$ is minimized by the control $u_k = -Mx_k$ where M is given by (6.1.28). This is the same control policy that is optimal for the deterministic infinite-time problem.

Theorem 6.2.3

Suppose conditions (6.1.30) hold. Then, among all controls u for which $C_{av}(u)$ exists and $E\|x_k\|^2$ remains bounded, the minimal cost is achieved by the control $u_k^1(x) = -Mx$ where M is given by (6.1.28). The minimal value of the cost is

$$C_{av}(u^1) = \mathrm{tr}[C^{\mathrm{T}}SC]$$

where S is the unique solution of the algebraic Riccati equation (6.1.29).

PROOF It is shown in Appendix B that $A - BM$ is stable, so that $J_{av}(u^1)$ exists. Let S be the solution of the ARE (6.1.29) and consider the N-stage problem of minimizing

$$C_N(u) = E\left[\sum_{k=0}^{N-1} \|Dx_k + Fu_k\|^2 + x_N^{\mathrm{T}}Sx_N\right]$$

This is the 'matched terminal cost' problem for which, from Theorem 6.2.2, control u^1 is optimal. Thus for any control u,

$$C_N(u) \geq C_N(u^1) = m_0^{\mathrm{T}}Sm_0 + \mathrm{tr}[SP_0] + \sum_{k=0}^{N-1} \mathrm{tr}[C^{\mathrm{T}}(k)SC(k)]. \tag{6.2.18}$$

Thus

$$\lim_{N\to\infty} \frac{1}{N} C_N(u) \geq \lim_{N\to\infty} \frac{1}{N} C_N(u^1) = C_{av}(u^1)$$

as long as the left-hand limit exists. But if $C_{av}(u)$ exists and $E\|x_k\|^2$ is bounded, then

$$\lim_{N\to\infty} \frac{1}{N} C_N(u) = C_{av}(u) + \lim_{N\to\infty} \frac{1}{N} E[x_N^{\mathrm{T}}Sx_N] = C_{av}(u).$$

This shows that u^1 is optimal. From (6.2.18) its cost is

$$C_{av}(u^1) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \mathrm{tr}[C^{\mathrm{T}}(k)SC(k)] = \mathrm{tr}[C^{\mathrm{T}}SC]. \qquad \square$$

The control $u_k^1 = -Mx_k$ is not the only optimal control for the average cost per unit time problem. Indeed, for any integer j we can write

$$C_{\text{av}}(u) = \lim_{N\to\infty} \frac{1}{N} E\left[\sum_{k=0}^{j-1} \|Dx_k + Fu_k\|^2\right] + \lim_{N\to\infty} \frac{1}{N} E\left[\sum_{k=j}^{N-1} \|Dx_k + Fu_k\|^2\right].$$

Now for any given control u_k,

$$E\left[\sum_{0}^{j-1} \|Dx_k + Fu_k\|^2\right]$$

is a fixed number not depending on N. Thus the first limit is zero, and since $(N-j)/N \to 1$ as $N \to \infty$,

$$C_{\text{av}}(u) = \lim_{N\to\infty} \frac{1}{N-j} E\left[\sum_{k=j}^{N-1} \|Dx_k + Fu_k\|^2\right].$$

The expression on the right is the average cost from time j onwards starting in state x_j, and its minimal value does not depend at all on what controls u_k were used for $k < j$. Thus any control of the form

$$u_k = \begin{cases} \text{arbitrary}, & k < j \\ -Mx_k, & k \geq j \end{cases}$$

is optimal. Thus the average cost criterion is only relevant when one is mainly concerned with 'long-run performance'; the idea is that the system settles down to a statistically stationary state in which an average of precisely $\text{tr}[C^{\text{T}}SC]$ is added to the cost at each stage, and this is minimal. There is, however, nothing in the cost criterion which specifies just how long this settling-down period is supposed to last. The discounted cost formulation has the opposite effect: it emphasizes performance during some initial interval the length of which is effectively specified by the discount factor. In this case the optimal control is unique. Another advantage of discounted costs is that the stabilizability/detectability conditions can always be met by sufficiently rapid discounting, whereas with average costs little can be said if the original system matrices (A, B, D) do not satisfy these conditions.

Exemplo:

$$x_{k+1} = a x_k + b u_k$$

$$\min \sum_{t=0}^{T-1} \left(x_t^2 q + u_t^2 r \right) + x_T^2 q$$

$$P_k = a^2 P_{k+1} + q - \frac{(a b P_{k+1})^2}{r + b^2 P_{k+1}}$$

$$= a^2 \left(P_{k+1} - \frac{b^2 P_{k+1}^2}{b^2 P_{k+1} + r} \right) + q$$

$$= a^2 \left(\frac{b^2 P_{k+1}^2 + P_{k+1} r - b^2 P_{k+1}^2}{b^2 P_{k+1} + r} \right) + q$$

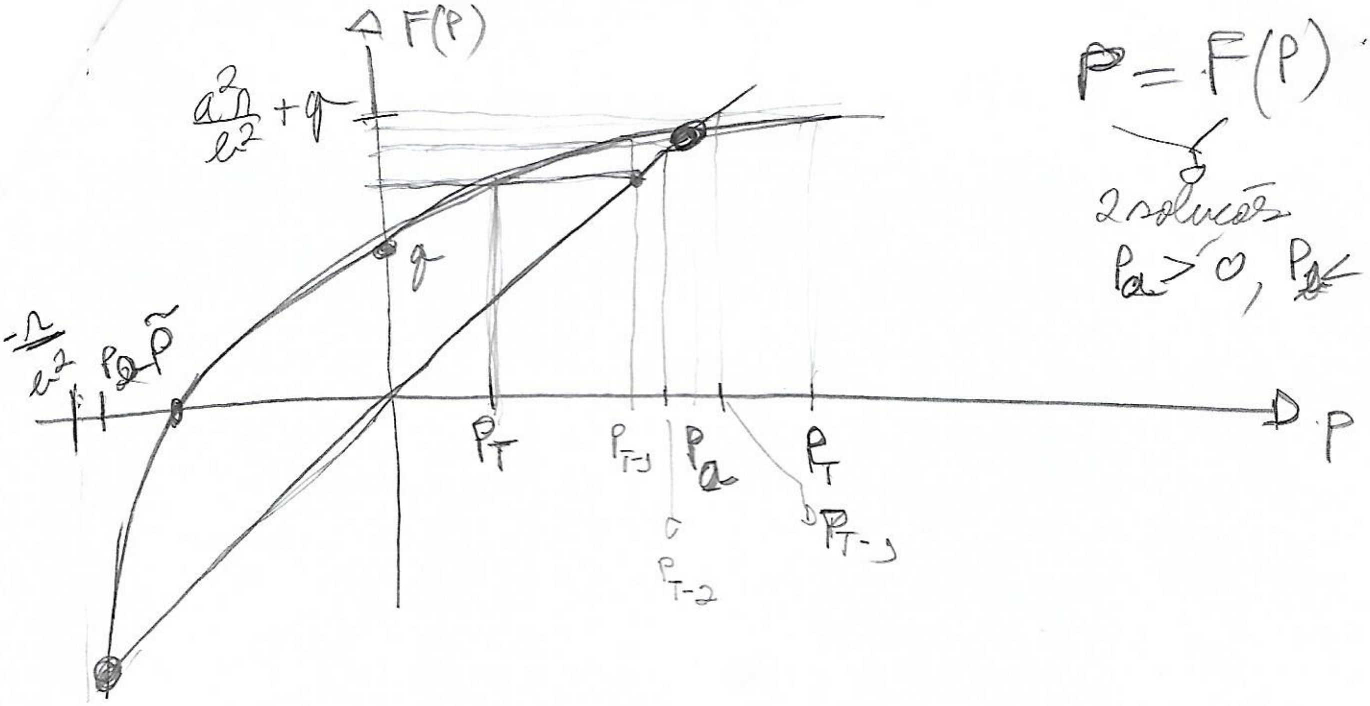
$$= \frac{a^2 r P_{k+1}}{b^2 P_{k+1} + r} + q = F(P_{k+1})$$

$$F(P) = \frac{a^2 r P}{b^2 P + r} + q$$

(a, b) estabilizável
(q, a) detetável

$\to$ $b^2 P + r = 0 \Rightarrow P = -\frac{r}{b^2}$

$S_{k+1} = F(S_k)$



$P = F(P)$

2 soluções

$P_a > 0$, $P_b < 0$

$$F(P) = 0 \Leftrightarrow a^2 r P + b^2 q P + qr = 0 \Leftrightarrow \tilde{P} = \frac{-qr}{a^2 r + b^2 q}$$

Graficamente, temos que

$P(K) = F(P_{K+1})$ é tal que

$P(K) \xrightarrow{K \to -\infty} P_a$,

qualquer que seja $P_T \geq 0$.

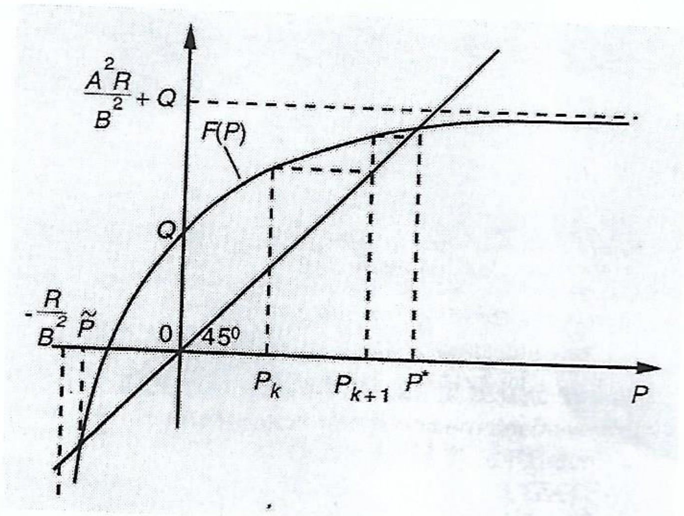


Figure 4.1.2 Graphical proof of Prop. 4.1 for the case of a scalar stationary system (one-dimensional state and control), assuming that $A \neq 0$, $B \neq 0$, $Q > 0$, and $R > 0$. The Riccati equation (1.13) is given by

$$P_{k+1} = A^2\left(P_k - \frac{B^2P_k^2}{B^2P_k + R}\right) + Q,$$

which can be equivalently written as

$$P_{k+1} = F(P_k),$$

where the function F is given by

$$F(P) = \frac{A^2RP}{B^2P + R} + Q.$$

Because F is concave and monotonically increasing in the interval $(-R/B^2, \infty)$, as shown in the figure, the equation $P = F(P)$ has one positive solution P^* and one negative solution $\tilde{P}$. The Riccati iteration $P_{k+1} = F(P_k)$ converges to P^* starting anywhere in the interval $(\tilde{P}, \infty)$ as shown in the figure.

where both minimizations are subject to the system equation constraint $x_{i+1} = Ax_i + Bu_i$. Furthermore, for a fixed x_0 and for every k, $x_0'P_k(0)x_0$ is bounded from above by the cost corresponding to a control sequence that forces x_0 to the origin in n steps and applies zero control after that. Such a sequence exists by the controllability assumption. Thus the sequence $\{x_0'P_k(0)x_0\}$ is nondecreasing with respect to k and bounded from above, and therefore converges to some real number for every $x_0 \in \Re^n$. It follows that the sequence $\{P_k(0)\}$ converges to some matrix P in the sense that each of the sequences of the elements of $P_k(0)$ converges to the correspond-