Capítulo 6

Programación Dinámica

6.1 Principio de optimalidad de Bellman

El propósito de este capítulo es presentar una breve introducción a la programación dinámica, que es una alternativa a los métodos de control optimal discutidos en los temas precedentes.

La programación Dinámica fue desarrollada por R.E. Bellman a finales de los años 50 y puede utilizarse para resolver problemas de control de sistemas no lineales y dependientes del tiempo.

La programación dinámica está basada en el llamado "principio de optimalidad" de R. Bellman, según el cual en una trayectoria óptima que une dos puntos A y D cualquier tramo de extremos B y C es en sí una trayectoria óptima entre todas las posibles que van de B a C.

El principio de optimalidad juega un papel similar al jugado por el principio del mínimo de Pontryagin para los métodos variacionales de los sistemas de control. Sirve para *limitar el número de los posibles controles optimales que deben investigarse.* También implica que las estrategias control optimal pueden determinarse hacia atrás desde el estado final.

En nuestro caso se cumple: Si $\mathbf{u}^*$ es un control óptimo en el intervalo de tiempo $[t_0, t_1]$ y $\mathbf{x}^*$ es la trayectoria óptima correspondiente a $\mathbf{u}^*$, entonces la restricción de $\mathbf{u}^*$ a cualquier subintervalo $[t, t_1]$ es un control óptimo para el estado inicial $\mathbf{x}^*(t)$

6.2 Sistemas discretos

La aplicación más inmediata es en el caso de sistemas discretos. El problema se plantea como:

Dado un sistema

$$\mathbf{x}_{k+1} = f^k(\mathbf{x}_k, \mathbf{u}_k)$$

donde el superíndice k de f indica que puede variar con el tiempo, encontrar la secuencia de entrada

$$\{\mathbf{u}^*(0), \mathbf{u}^*(1), \ldots, \mathbf{u}^*(N)\}$$

que minimiza el índice de costo

$$J = \sum_{k=0}^{N} F^k(\mathbf{x}_k, \mathbf{u}_k)$$

Denominando

$$J_i = \sum_{k=i}^{N} F^k(\mathbf{x}_k, \mathbf{u}_k)$$

este costo será función del instante inicial i, del estado inicial $\mathbf{x}(i)$, y de la entrada $\mathbf{u}$ en el intervalo $[i, N]$, pero no dependerá de estados anteriores.

Aplicando el principio de Bellman, el óptimo J_i^* para las trayectorias que en el instante i se encuentran en el estado $\mathbf{x}(i)$, dependerá por tanto únicamente del estado inicial $\mathbf{x}(i)$ y del instante inicial i

$$J_i^* = \min_u \left\{ \sum_{k=i}^{N} F^k(\mathbf{x}_k, \mathbf{u}_k) \right\} = J_i^*(\mathbf{x}_i)$$

Para $i = N$ se cumple:

$$J_N^* = \min_{u_N} \left\{ F^N(\mathbf{x}_N, \mathbf{u}_N) \right\}$$

por lo que J_N^* sólo depende de N y de $\mathbf{x}_N$.

Si suponemos que J_{i+1}^* sólo depende de $i+1$ y de x_{i+1}, el principio de optimalidad de Bellman implica:

$$J_i^*(\mathbf{x}_i) = \min_{u_i} \left\{ F^i(\mathbf{x}_i, \mathbf{u}_i) + J_{i+1}^*(\mathbf{x}_{i+1}) \right\} \tag{6.1}$$

y además $\mathbf{x}_{i+1}$ sólo depende de $\mathbf{x}_i$ y de $\mathbf{u}_i$ por lo que al minimizar respecto de $\mathbf{u}_i$, el coste óptimo J_i^* dependerá únicamente de $\mathbf{x}_i$ y de i.

La aplicación iterativa de 6.1 desde el final hacia atrás permite calcular u_k^*

6.2.1 Ejemplo

Supuesto el sistema

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

donde $u_k \in \{-1, 0, 1\}$, transferir el sistema desde el estado inicial $x_0 = 2$ al estado final $x_4 = 0$ minimizando el índice de coste

$$J = \sum_{k=0}^{4} \left(x_k^2 + 2u_k^2 \right)$$

Las posibles trayectorias se indican en la figura 6.1

Se tiene

$$x_4 = 0 \qquad u_4^* = 0 \qquad J_4^*(0) = 0$$

En el paso del 3 al 4 los únicos caminos posibles son:

$$\begin{array}{llll}
x_3 = 1 & u_3 = -1 & x_4 = 0 & J_3^*(1) = 3 \\
x_3 = 0 & u_3 = 0 & x_4 = 0 & J_3^*(0) = 0 \\
x_3 = -1 & u_3 = 1 & x_4 = 0 & J_3^*(-1) = 3
\end{array}$$

En el paso del 2 al 3 los únicos caminos posibles son:

$$\begin{array}{llll}
x_2 = 2 & u_2 = -1 & x_3 = 1 & J_2^*(2) = 9 \\
 & & & \\
x_2 = 1 & u_2 = 0 & x_3 = 1 & \\
x_2 = 1 & u_2 = -1 & x_3 = 0 & J_2^*(1) = 3 \\
 & & & \\
x_2 = 0 & u_2 = 1 & x_3 = 1 & \\
x_2 = 0 & u_2 = 0 & x_3 = 0 & J_2^*(0) = 0 \\
x_2 = 0 & u_2 = -1 & x_3 = -1 &
\end{array}$$

obteniendo los valores de J_2^* a partir de la expresión

$$J_2(x_2) = \min_{u_2} \left\{ x_2^2 + 2u_2^2 + J_3^*(x_3) \right\}$$

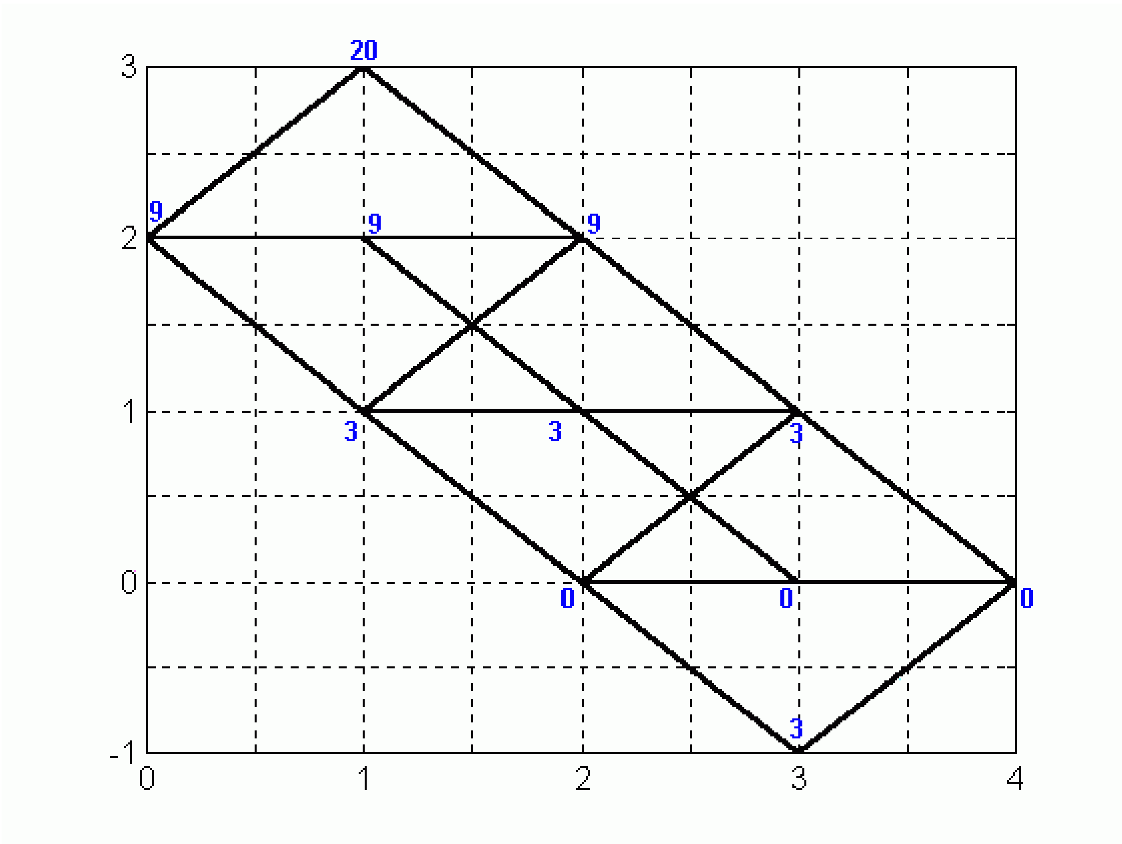

Figura 6.1: Sistema discreto

En el paso del 1 al 2 los únicos caminos posibles son

$$\begin{array}{llll} x_1 = 3 & u_1 = -1 & x_2 = 2 & J_1^*(3) = 20 \\ \\ x_1 = 2 & u_1 = 0 & x_2 = 2 & \\ x_1 = 2 & u_1 = -1 & x_2 = 1 & J_1^*(2) = 9 \\ \\ x_1 = 1 & u_1 = 1 & x_2 = 2 & \\ x_1 = 1 & u_1 = 0 & x_2 = 1 & \\ x_1 = 1 & u_1 = -1 & x_2 = 0 & J_1^*(1) = 3 \end{array}$$

obteniendo los valores de J_1^* a partir de la expresión

$$J_1(x_2) = \min_{u_1} \left\{ x_1^2 + 2u_1^2 + J_2^*(x_2) \right\}$$

En el paso del 0 al 1 los únicos caminos posibles son

$$\begin{array}{llll} x_0 = 2 & u_0 = 1 & x_1 = 3 & \\ x_0 = 2 & u_0 = 0 & x_1 = 2 & \\ x_0 = 2 & u_0 = -1 & x_1 = 1 & J_0^*(2) = 9 \end{array}$$

obteniendo este valor de J_0^* a partir de la expresión

$$J_0(x_0) = \min_{u_0} \left\{ x_0^2 + 2u_0^2 + J_1^*(x_1) \right\}$$

Los valores que toman el control y el estado en la trayectoria óptima son los indicados en la siguiente tabla

k	0	1	2	3	4
x_k^*	2	1	0	0	0
u_k^*	-1	-1	0	0	0

6.3 Sistemas Continuos

Podemos considerar los problemas de control optimal de un sistema continuo desde dos puntos de vista. Podemos discretizar, resolver utilizando control optimal discreto o podemos resolver el problema de control optimal continuo para obtener una salida continua. Estudiemos separadamente ambos casos

6.3.1 Control digital

Sea el sistema

$$\dot{\mathbf{x}} = f(t, \mathbf{x}, \mathbf{u}) \tag{6.2}$$

con el índice de coste siguiente

$$J(0) = \phi(x(t_1), t_1) + \int_0^{t_1} F(t, \mathbf{x}(t), \mathbf{u}(t))\, dt \tag{6.3}$$

Para discretizar el sistema con un periodo de τ segundos, podemos utilizar una aproximación de primer orden

$$\dot{\mathbf{x}}(k\tau) = \frac{(\mathbf{x}_{k+1} - \mathbf{x}_k)}{\tau} \tag{6.4}$$

para escribir 6.2 como

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \tau f(k\tau, \mathbf{x}_k, \mathbf{u}_k) \tag{6.5}$$

donde $\mathbf{x}_k = \mathbf{x}(k\tau)$, $\mathbf{u}_k = \mathbf{u}(k\tau)$. Definiendo

$$f^k(\mathbf{x}_k, \mathbf{u}_k) = \mathbf{x}_k + \tau f(k\tau, \mathbf{x}_k, \mathbf{u}_k) \tag{6.6}$$

la ecuación puede escribirse como

$$\mathbf{x}_{k+1} = f^k(\mathbf{x}_k, \mathbf{u}_k) \tag{6.7}$$

exactamente igual que en el caso discreto.

Para discretizar el índice de coste, podemos escribir

$$J(0) = \phi(x(t_1), t_1) + \sum_{k=0}^{N-1} \int_{k\tau}^{(k+1)\tau} F(t, \mathbf{x}(t), \mathbf{u}(t))\, dt \tag{6.8}$$

donde

$$N = \frac{t_1}{\tau} \tag{6.9}$$

Utilizando una aproximación de primer orden para cada integral obtenemos

$$J(0) = \phi(x(t_1), t_1) + \sum_{k=0}^{N-1} \tau F(k\tau, \mathbf{x}_k, \mathbf{u}_k) \tag{6.10}$$

Definiendo las siguientes funciones

$$\begin{aligned} J_0 &= J(0) \\ \phi^S(N, \mathbf{x}_N) &= \phi(\mathbf{x}(N\tau), N\tau) \\ F^k(\mathbf{x}_k, \mathbf{u}_k) &= \tau F(k\tau, \mathbf{x}_k, \mathbf{u}_k) \end{aligned} \tag{6.11}$$

la ecuación 6.10 se transforma en

$$J_0 = \phi^S\left(N, \mathbf{x}_N\right) + \sum_{k=0}^{N-1} F^k\left(\mathbf{x}_k, \mathbf{u}_k\right) \tag{6.12}$$

A continuación, podemos utilizar la programación dinámica para calcular $\mathbf{u}_k^*$ como en la sección anterior. El control digital que hay que aplicar al sistema actual está dado por

$$\mathbf{u}\left(t\right) = \mathbf{u}_k^* \qquad k\tau \leq t < \left(k+1\right)\tau \tag{6.13}$$

Para utilizar programación dinámica, los valores del estado y del control deben primero ser *cuantizados,* es decir, restringidos a un conjunto finito de valores admisibles. Cuanto más fina es la cuantización más preciso será el control digital.

6.3.2 Control continuo y la ecuación de Hamilton-Jacobi-Bellman

Dado un sistema, en general multivariante, definido por la ecuación diferencial

$$\dot{\mathbf{x}}\left(t\right) = f\left(t, \mathbf{x}\left(t\right), \mathbf{u}\left(t\right)\right) \tag{6.14}$$

se trata de encontrar un control $\mathbf{u}^*$ en el intervalo de tiempo $[t_0, t_1]$ que minimiza el índice de costo

$$J = \phi\left(\mathbf{x}\left(t_1\right), t_1\right) + \int_{t_0}^{t_1} F\left(\tau, \mathbf{x}\left(\tau\right), \mathbf{u}\left(\tau\right)\right) d\tau \tag{6.15}$$

y que conduce el sistema desde un estado inicial $\mathbf{x}\left(t_0\right)$ hasta un estado final y satisfaciendo

$$\psi\left(\mathbf{x}\left(t_1\right), t_1\right) = 0 \tag{6.16}$$

Para una función dada ψ. Veamos en primer lugar la forma que el principio de optimalidad de Bellman toma para este problema

Supongamos que t es el instante de tiempo actual y $t+\Delta t$, es un instante de tiempo posterior cercano a t, entonces el coste hasta $\mathbf{x}\left(t_1\right)$ puede escribirse como:

$$J\left(t, \mathbf{x}\right) = \phi\left(\mathbf{x}\left(t_1\right), t_1\right) + \int_{t+\Delta t}^{t_1} F\left(\tau, \mathbf{x}, \mathbf{u}\right) d\tau + \int_{t}^{t+\Delta t} F\left(\tau, \mathbf{x}, \mathbf{u}\right) d\tau \tag{6.17}$$

podemos decir por tanto

$$J\left(t, \mathbf{x}\right) = \int_{t}^{t+\Delta t} F\left(\tau, \mathbf{x}, \mathbf{u}\right) d\tau + J\left(t+\Delta t, \mathbf{x}+\Delta\mathbf{x}\right) \tag{6.18}$$

donde $\mathbf{x}+\Delta\mathbf{x}$, es el estado en el instante $t+\Delta t$, cuando se utilizan el estado actual $\mathbf{x}\left(t\right)$ y el control actual $\mathbf{u}\left(t\right)$ en la ecuación 6.14. Notar que como aproximación hasta el primer orden

$$\Delta\mathbf{x} = f\left(t, \mathbf{x}, \mathbf{u}\right)\Delta t \tag{6.19}$$

La ecuación 6.18 describe todos los costes posibles para ir desde el estado en el instante t hasta el estado en t_1. De acuerdo con el principio de optimalidad de Bellman, el único candidato para $J^*\left(t, \mathbf{x}\right)$ son aquellos costes que son óptimos para llevar el sistema desde $t+\Delta t$ hasta t_1. Supongamos que el coste optimal $J^*\left(t+\Delta t, \mathbf{x}+\Delta\mathbf{x}\right)$ es conocido para todos los posibles estados $\mathbf{x}+\Delta\mathbf{x}$. Supongamos también que el control optimal ha sido determinado en el intervalo $[t+\Delta t, t_1]$ para cada $\mathbf{x}+\Delta\mathbf{x}$,. Entonces solamente queda por seleccionar el control actual en el intervalo $[t, t+\Delta t]$. De aquí

$$J^*\left(t, \mathbf{x}\right) = \min_{\substack{u(\tau)\\ t\leq\tau\leq t+\Delta t}} \left\{ \int_{t}^{t+\Delta t} F\left(\tau, \mathbf{x}, \mathbf{u}\right) d\tau + J^*\left(t+\Delta t, \mathbf{x}+\Delta\mathbf{x}\right) \right\} \tag{6.20}$$

Este es el *principio de optimalidad para sistemas continuos.*

Si el integrando en la expresión 6.20 es continuo en el intervalo $[t, t+\Delta t]$, y en particular si F y $\mathbf{u}$ son continuas, se puede aplicar el teorema del valor medio del cálculo integral, en virtud del cual existe un $s \in [t, t+\Delta t]$ tal que

$$\int_{t}^{t+\Delta t} F(\tau, \mathbf{x}, \mathbf{u})\, d\tau = F(s, \mathbf{x}(s), \mathbf{u}(s)) \cdot \Delta t$$

y si J^* es desarrollable en serie de Taylor en t, se tiene

$$J^*(t+\Delta t, \mathbf{x}+\Delta \mathbf{x}) = J^*(t, \mathbf{x}) + \left(\frac{\partial J^*}{\partial \mathbf{x}}\right)^T \cdot \Delta \mathbf{x} + \frac{\partial J^*}{\partial t} \cdot \Delta t + o\left((\Delta t)^2\right)$$

donde

$$\frac{\partial J^*}{\partial \mathbf{x}} = \left(\frac{\partial J^*(t, \mathbf{x})}{\partial x_1}, \frac{\partial J^*(t, \mathbf{x})}{\partial x_2}, \cdots, \frac{\partial J^*(t, \mathbf{x})}{\partial x_n}\right)$$

$$\frac{\partial J^*}{\partial t} = \frac{\partial J^*(t, \mathbf{x}(t))}{\partial t}$$

$$f(t, \mathbf{x}, \mathbf{u}) = f(t, \mathbf{x}(t), \mathbf{u}(t))$$

con lo que la expresión 6.20 se transforma en

$$J^*(t, \mathbf{x}(t)) = \min_{\substack{u(\tau) \\ t \le \tau \le t+\Delta t}} \left\{F(s, \mathbf{x}(s), \mathbf{u}(s)) \cdot \Delta t + J^*(t, \mathbf{x}) + \left(\left(\frac{\partial J^*}{\partial x}\right)^T \cdot \Delta \mathbf{x} + \frac{\partial J^*}{\partial t}\Delta t\right) + o\left((\Delta t)^2\right)\right\}$$

Utilizando 6.19 y como J^* y $\partial J^* / \partial t \cdot \Delta t$ son independientes de $\mathbf{u}(\tau)$, $\tau \in [t, t+\Delta t]$

$$0 = \frac{\partial J^*}{\partial t}\Delta t + \min_{\substack{u \\ t \le \tau \le t+\Delta t}} \left\{F(s, \mathbf{x}(s), \mathbf{u}(s)) \cdot \Delta t + \left(\frac{\partial J^*}{\partial x}\right)^T \cdot f(t, \mathbf{x}, \mathbf{u}) \cdot \Delta t + o\left((\Delta t)^2\right)\right\}$$

y si hacemos $\Delta t \to 0$, se obtiene

$$0 = \frac{\partial J^*(t, \mathbf{x}(t))}{\partial t} + \min_{u(t)} \left\{F(t, \mathbf{x}, \mathbf{u}) + \left(\frac{\partial J^*}{\partial \mathbf{x}}\right)^T \cdot f(t, \mathbf{x}, \mathbf{u})\right\} \tag{6.21}$$

conocida como Ecuación de Hamilton-Jacobi-Bellman (HJB). Se resuelve hacia atrás en el tiempo a partir de $t = t_1$, y poniendo $t_0 = t_1$ en 6.15, la condición de contorno es

$$J^*(t_1, \mathbf{x}(t_1)) = \phi(t_1, \mathbf{x}(t_1)) \text{ sobre la hipersuperficie } \psi(t_1, \mathbf{x}(t_1)) = 0 \tag{6.22}$$

Si definimos la función Hamiltoniana como

$$H(t, \mathbf{x}, \mathbf{u}, \mathbf{p}) = F(t, \mathbf{x}, \mathbf{u}) + \mathbf{p}^T f(t, \mathbf{x}, \mathbf{u}) \tag{6.23}$$

la ecuación HJB podrá escribirse como

$$\frac{\partial J^*(t, \mathbf{x})}{\partial t} + \min_{u} \{H(t, \mathbf{x}, \mathbf{u}, J_x^*)\} = 0 \tag{6.24}$$

6.3.3 Ejemplo

Dado el sistema definido por

$$x'(t) = u(t)$$

minimizar el índice de costo

$$J = \int_{t_0}^{t_1} \left(ax^2 + bu^2\right) d\tau$$

donde a y b son constantes. Se ha de cumplir la ecuación 6.21 que aquí es

$$0 = \frac{\partial J^*(t,x)}{\partial t} + \min_{u(t)} \left\{ ax^2 + bu^2 + \frac{\partial J^*(t,x)}{\partial x} u \right\}$$

Condición necesaria de mínimo respecto de $u(t)$ para la expresión del segundo miembro es

$$2bu^* + \frac{\partial J^*(t,x)}{\partial x} = 0 \Longleftrightarrow u^*(t) = -\frac{1}{2b} \cdot \frac{\partial J^*(t,x(t))}{\partial x} \tag{6.25}$$

con lo que la ecuación 6.21 de Hamilton-Jacobi quedará

$$\frac{\partial J^*(t,x)}{\partial t} + ax^2 - \frac{1}{4b} \cdot \left(\frac{\partial J^*(t,x(t))}{\partial x} \right)^2 = 0 \tag{6.26}$$

con la condición de contorno

$$J^*(t_1, x(t_1)) = 0$$

Caso particular

Si los datos iniciales son:

$$\begin{array}{ll} t_0 = 0 & x(0) = 0 \\ t_1 = 5 & x(5) = 1 \\ a = 0 & b = 1 \end{array}$$

$$J = \int_0^5 u^2 d\tau$$

La ecuación de Hamilton-Jacobi (6.26) queda en este caso particular

$$\frac{\partial J^*(t,x(t))}{\partial t} - \frac{1}{4} \cdot \left(\frac{\partial J^*(t,x(t))}{\partial x} \right)^2 = 0 \tag{6.27}$$

con la condición de contorno

$$J^*(5,1) = 0 \tag{6.28}$$

Se puede ensayar

$$J^*(t,x) = \Phi(t) \cdot (x-1)^2 \tag{6.29}$$

Por 6.27 se debe cumplir

$$\Phi'(t) \cdot (x-1)^2 = \frac{1}{4} [2 \cdot \Phi(t) \cdot (x-1)]^2$$

de donde se deduce que

$$\Phi'(t) = \Phi(t)^2$$

cuyas soluciones son de la forma

$$\Phi(t) = -\frac{1}{t+A}$$

con A constante. Por ello la expresión 6.29 queda:

$$J^*(t,x) = -\frac{1}{t+A} \cdot (x-1)^2$$

La expresión del control óptimo se puede deducir de 6.25, siendo:

$$u^*(t) = -\frac{1}{2} \cdot \frac{\partial J^*(t,x(t))}{\partial x} = \frac{x(t)-1}{t+A} \tag{6.30}$$

Por la dinámica del sistema se cumple

$$x'(t) = u(t)$$

luego cuando la entrada es u^* el estado, que corresponde a la trayectoria óptima, cumple la ecuación diferencial

$$x'(t) = \frac{x(t) - 1}{t + A}$$

o la equivalente

$$\frac{x'(t)}{x(t) - 1} = \frac{1}{t + A}$$

cuyas soluciones son de la forma

$$x(t) = Bt + (AB + 1) \tag{6.31}$$

siendo B una constante.

Particularizando 6.31 para los valores inicial y final se obtiene

$$\begin{aligned} AB + 1 &= 0 \\ 5B + AB + 1 &= 1 \end{aligned}$$

de donde se deduce que

$$A = -5 \qquad B = \frac{1}{5}$$

que llevadas a las expresiones 6.30 y 6.31 determinan el control óptimo y el estado del sistema en cada instante

$$u^*(t) = \frac{1}{5} \qquad x^*(t) = \frac{t}{5}$$