Divulgaciones Matemáticas Vol. 7 No. 2 (1999), pp. 167–185

Control Óptimo Determinista via Programación Dinámica

Deterministic Optimal Control via Dynamic Programming

Guillermo Ferreyra (ferreyra@math.lsu.edu)
Department of Mathematics
Louisiana State University
Baton Rouge, LA 70803, U.S.A.
Jesús Pascal (pascal@luz.ve)
Departamento de Matemática y Computación
Facultad de Ciencias, Universidad del Zulia
Maracaibo, Venezuela.

Resumen

Este trabajo ofrece una introducción a la Teoría del Control Óptimo con varios ejemplos de aplicaciones prácticas de esta teoría a problemas de ciencias e ingeniería. Se hace énfasis en el método de la Programación Dinámica el cual, aplicado a problemas de control óptimo determinista, produce una ecuación diferencial en derivadas parciales no lineal de primer orden cuya solución es la función de valor del problema original. Se incluye una demostración no rigurosa de este resultado (la cual sin embargo puede volverse rigurosa utilizando conceptos más avanzados que los presentados en este trabajo). Finalmente se utiliza esta técnica para resolver explícitamente dos de los problemas planteados.
Palabras y Frases Clave: control óptimo, solución de viscosidad, programación dinámica.

Abstract

This paper offers an introduction to Optimal Control Theory and several examples of practical applications of this theory to problems in science end engineering. Emphasis is given to the dynamic programming method, which applied to deterministic optimal control problems

Recibido 1999/03/03. Aceptado 1999/05/23.
MSC (1991): Primary 49-01; Secondary 49L20, 49L25.

produce a first order nonlinear partial differential equation whose solution is the value function of the original problem. A non-rigorous proof of this fact is included (but the proof can be made rigorous using deeper concepts than those defined in this paper). Finally the technique is used to explicitly solve two of the stated problems.
Key words and phrases: optimal control, viscosity solutions, dynamic programming.

1 Introducción

La idea de *control* puede ser expresada como el proceso mediante el cual se ejerce una influencia sobre el comportamiento de un sistema dinámico (que varía con el tiempo) para alcanzar un propósito previamente fijado. Una clase importante de modelos de sistemas dinámicos controlados, a los cuales se les dice simplemente *sistemas de control*, es la representada por la ecuación diferencial en $\mathbb{R}^n$

$$\dot{x}(t) = f(x(t), u(t)), \quad x(t_0) = x_0, \tag{1}$$

donde la dinámica f es una función que satisface condiciones adecuadas y el control $u(\cdot)$ pertenece a una familia especial $\mathcal{U}$ de funciones con valores en un subconjunto U de $\mathbb{R}^n$. Una vez elegido un control $u \in \mathcal{U}$, el sistema (1) determina una trayectoria o estado $x(\cdot)$ con condición inicial x_0 en el momento t_0.

Por ejemplo, si se desea controlar la trayectoria de un avión, con condición inicial $x(t_0)$, para lograr una condición final $x(t_f)$, el estado del sistema $x(\cdot)$ podría representar la posición y velocidad del avión y el control $u(\cdot)$ representaría la fuerza o aceleración necesaria para lograr tal objetivo. Con esta formulación, este ejemplo representa un problema para la Teoría de Control, la cual hace énfasis en el análisis sobre las condiciones necesarias y suficientes para la existencia de los controles adecuados, y su computabilidad, asi como también de la existencia, unicidad, y estabilidad de la trayectoria que garantice el logro de dicho objetivo. Ahora bién, si además se desea lograr tal propósito en un tiempo mínimo, o con mínimo uso de combustible, entonces este es un problema de control óptimo. En tal caso, se quiere minimizar una funcional que depende del estado del sistema y del control llamada *funcional de costo*

$$J^{u(\cdot)}(t_0, x_0) = \ell(x(t_f)) + \int_{t_0}^{t_f} L(x(t), u(t))\, dt, \tag{2}$$

donde L y ℓ son funciones que satisfacen condiciones adecuadas. La función L representa el costo incurrido por el desplazamiento $x(\cdot)$ y por la fuerza realizada $u(\cdot)$, mientras que la función ℓ representa la penalización por la desviación del estado $x(t_f)$ en el instante final t_f de un estado deseado x_f. En nuestro ejemplo, si queremos minimizar la cantidad de tiempo transcurrido t_f, debemos tomar $\ell = 0$, $L = 1$. Por otro lado, si deseamos minimizar el uso de combustible, podemos tomar $L(x, u) = u^2$.
Si un control u^* es tal que minimiza la funcional de costo, es decir, si

$$J^{u^*(\cdot)}(t_0, x_0) \leq J^{u(\cdot)}(t_0, x_0), \quad \forall u(\cdot) \in \mathcal{U},$$

entonces u^* se denomina *control óptimo*.
La Teoría de Control Optimo hace énfasis en el estudio de condiciones necesarias y suficientes para la existencia y unicidad del control óptimo, asi como también del desarrollo de metodologías para su determinación y computabilidad.

1.1 El principio del máximo de Pontryagin

En 1959, L.S. Pontryagin et al [14], presentaron condiciones necesarias de optimalidad, las cuales han sido llamadas el *Principio del Máximo de Pontryagin*, para el problema de optimización determinado por (1) y (2). Este resultado establece, bajo ciertas condiciones para la dinámica f y la familia de controles $\mathcal{U}$, que si $u^*(\cdot)$ es un control óptimo y $x^*(\cdot)$ es la solución de (1) que corresponde a $u^*(\cdot)$, entonces existen una constante $\lambda \leq 0$ y una función vectorial ϕ, tales que el vector (λ, ϕ) no es idénticamente nulo, ϕ es absolutamente continua y

$$\begin{aligned}
\dot{\phi}(t) &= -H_x(x^*(t), u^*(t), \phi(t)), \\
\dot{x}^*(t) &= H_\phi(x^*(t), u^*(t), \phi(t)), \\
H(x^*(t), u^*(t), \phi(t)) &= \max_{u \in U} H(x^*(t), u, \phi(t)),
\end{aligned}$$

donde el *Hamiltoniano* H está dado por

$$H(x, u, \phi) = \lambda\, L(x, u) + \phi \cdot f(x, u).$$

Aquí $\phi \cdot f$ denota el producto interno de ϕ y f. Para algunos problemas especiales de control el Principio del Máximo de Pontryagin enunciado arriba no aporta suficiente información para resolver el problema de control óptimo. Problemas de este tipo son los descriptos como *problemas de control óptimo singulares* y que han emergido en varias especialidades de ingeniería, ciencias

básicas, finanzas, etc.
Un control óptimo se dice *singular* si el determinante $\det(H_{uu})$ se anula en todo punto a lo largo de la trayectoria óptima. En caso contrario, el control óptimo se dice no singular. En particular, si el Hamiltoniano H es lineal con respecto a una o más de las funciones componentes del control, entonces dicho problema resulta ser singular.

1.2 El método de programación dinámica

En 1957, Richard Bellman presentó el Método de Programación Dinámica para resolver problemas de control óptimo. Este método consiste en reemplazar el problema de optimización (1), (2), el cual contiene una minimización en el espacio de dimensión infinita $\mathcal{U}$, por una ecuación diferencial en derivadas parciales no lineal, llamada *ecuación de programación dinámica*, o también *ecuación de Hamilton-Jacobi-Bellman* (HJB)

$$0 = V_t(t,x) + \inf_{u \in U} \{L(x,u) + V_x(t,x) \cdot f(x,u)\}, \quad t \in [t_0, t_f], \quad x \in \mathbb{R}^n, \quad (3)$$

que es satisfecha por una función denominada *función de valor*, la cual se define como

$$V(t_0, x_0) = \inf_{u(\cdot) \in \mathcal{U}} J^{u(\cdot)}(t_0, x_0). \quad (4)$$

Durante dos décadas la Teoría de Control Optimo dedicó mayor énfasis al estudio, desarrollo y aplicación del Principio del Máximo de Pontryagin dado que este es válido en condiciones mucho más sencillas que las condiciones requeridas para la validez del resultado de Bellman. En efecto, para que la función de valor V satisfaga (3) en el sentido clásico de ecuaciones en derivadas parciales se necesita que V sea continuamente diferenciable, pero para muchos problemas de control óptimo, la función de valor no es diferenciable. Crandall y Lions [4] presentaron una noción más débil de solución de la ecuación de programación dinámica (3). Esta noción se conoce como la de *solución de viscosidad* de (3). Con esta noción el método de programación dinámica es aplicable en la mayoría de los casos de interés, aún cuando la función de valor no sea diferenciable. En efecto, se ha probado en condiciones muy generales que la función de valor es una solución de viscosidad de la ecuación de programación dinámica (3). Es decir, la ecuación de programación dinámica (3) es una condición necesaria que la función de valor debe satisfacer en el sentido de viscosidad. Para más detalles sobre este tema, leer [2], [3], [4] y [12].
La organización del presente trabajo es como sigue. En la sección 2, presentamos varios ejemplos para ilustrar con más detalles algunas aplicaciones

prácticas de la teoría de control óptimo en la solución de problemas específicos de ingeniería, de ciencias básicas y de economía. En la sección 3 se introducen las condiciones para la dinámica f del sistema de control y para la familia especial de controles $\mathcal{U}$, bajo las cuales será considerado el problema de control óptimo analizado en la sección 4 siguiente. En la sección 4 se introduce el Principio de Programación Dinámica y una demostración no rigurosa del Teorema 1 el cual establece que la función de valor V es solución de la ecuación de programación dinámica (3) asumiendo que dicha función es continuamente diferenciable. En la sección 5 aplicamos el método de programación dinámica para resolver dos problemas de control óptimo; el problema presentado en el Ejemplo 3 de la sección 2 y el presentado en el Ejemplo 5 de la sección 2. Este último es un problema de control óptimo singular.

2 Ejemplos

Ejemplo 1

Inestabilidad del brazo de un robot para soldaduras en una planta automotriz.
El movimiento del brazo de un robot para hacer soldaduras en una planta automotriz produce una oscilación del elemento soldador situado en la punta del brazo. Se quiere detener tal vibración rápidamente, es decir, se desea que el elemento soldador se detenga en una posición determinada, en el menor tiempo posible. Se asume que en forma libre el comportamiento del elemento soldador sería como el del oscilador armónico con fricción en el plano vertical. Esto es, si designamos por $\theta(t)$ la desviación en el instante t del ángulo deseado, el movimiento del elemento satisface la ecuación diferencial

$$\ddot{\theta} + \alpha\dot{\theta} + \omega^2\theta = 0,$$

donde α y ω son constantes que representan la fuerza de fricción y la constante de elasticidad del oscilador respectivamente. La posición inicial $\theta(0)$ y la velocidad inicial $\dot{\theta}(0)$ son conocidas. Como deseamos anular la oscilación lo mas rápido posible, es necesario diseñar un mecanismo que aplique un torque para detener tal oscilación en tiempo mínimo. Así se obtiene la ecuación diferencial que rige el movimiento del oscilador armónico con fricción y fuerza externa $u(t)$

$$\ddot{\theta}(t) + \alpha\dot{\theta}(t) + \omega^2\theta(t) = u(t). \tag{5}$$

En este caso el control $u(t)$ es llamado el control de torque. El problema de control óptimo entonces consiste en encontrar una función (control) $u^*(\cdot)$ de forma tal que dados $\theta(0)$ y $\dot{\theta}(0)$, la solución de (5) satisfaga

$$\theta(t_f) = 0 \quad y \quad \dot{\theta}(t_f) = 0,$$

y tal que el tiempo final t_f sea el mínimo posible. Esto es, dicha función $u^*(\cdot)$ debe minimizar el valor

$$t_f = \int_0^{t_f} 1\,dt.$$

Ejemplo 2

Superficie mínima de revolución.

Se quiere determinar una curva plana, continua, C^1 a trozos, con extremos $(0, A)$ y (T, B) fijos, que al ser rotada alrededor del eje horizontal produzca una superficie de revolución con área mínima. Designemos por $x(t)$ una curva plana, continua y C^1 a trozos en el intervalo $[0, T]$, con $x(0) = A \quad y \quad x(T) = B$. Asumamos que $x(t) > 0$. Entonces el área de la superficie de revolución determinada por $x(t)$ está dada por la expresión

$$2\,\pi \int_0^T x(t)\,ds(t),$$

donde $x(t)$ es el radio y $s(\cdot)$ es la longitud del arco determinado por $x(\cdot)$. Designemos $u(t) = \dot{x}(t)$. El problema consiste entonces en minimizar la expresion

$$J^{u(\cdot)} = 2\,\pi \int_0^T x(t)\,\sqrt{1 + u(t)^2}\,dt.$$

En [1] puede verse con detalle la solución de este problema. La curva que origina la superficie de revolución con área mínima resulta ser una ”catenaria”, también llamada curva de los cables colgantes.

Ejemplo 3

Reacción química. Regulador lineal cuadrático.

En una reacción química la calidad del producto final depende del valor del pH. Por ello se desea controlar el pH de dicha reacción química variando la cantidad u de uno de los ingredientes. Designemos por $x(t)$ al valor del pH en el instante t. Se supone que en cada instante la velocidad de cambio del pH

es proporcional a la suma del valor del pH en dicho instante más la cantidad u del ingrediente de control. Entonces

$$\dot{x}(t) = \alpha\, x(t) + \beta\, u(t).$$

Aquí suponemos

(i) α y β son constantes positivas conocidas.

(ii) $x(0) = x_0$ es el valor del pH en el instante inicial.

(iii) El intervalo de tiempo $0 \leq t \leq T$ es fijo.

(iv) La cantidad del ingrediente de control u que puede ser utilizada no es acotada.

Supongamos que x_d es el valor del pH deseado. Introduciendo la nueva variable $y(t) = x(t) - x_d$ podemos asumir, sin pérdida de generalidad, que deseamos controlar la calidad del producto de forma tal que $|x(t)|$ sea lo más pequeño posible. Luego, $x(t)$ y $u(t)$ juegan dos papeles económicos opuestos. Supuestamente el uso de un control $u(t) \neq 0$ produce una reducción en $|x(t)|$, pero tanto $x(t)$ como $u(t)$ no nulos resultan onerosos para la empresa. Entonces el problema consiste en minimizar la expresión

$$J^{u(\cdot)}(0, x_0) = \int_0^T (K\, x^2(t) + u^2(t))\, dt,$$

donde K es una constante de proporcionalidad conocida.
En la sección 5 regresaremos a este ejemplo presentando su solución vía programación dinámica.

Ejemplo 4

Consumo e inversión.
La riqueza $x(t)$ de un individuo es invertida en depósitos bancarios que producen renta k. Al mismo tiempo, el individuo consume parte de su riqueza a razón $u(t) \geq 0$. El individuo desea maximizar su capacidad de consumo a lo largo del tiempo. Luego, se tiene el siguiente sistema de control

$$\dot{x}(t) = k\, x(t) - u(t), \quad x(0) = x_0.$$

El problema de control óptimo consiste en escoger la función $u(\cdot)$ de tal forma que la utilidad del consumo total, sobre un período de tiempo previamente determinado, sea máxima. Esto es, queremos maximizar una expresión como

$$\int_0^T e^{-\rho t}\sqrt{u(t)}\,dt.$$

Además, es lógico imponer la condición $x(T) = 0$. Haciendo un análisis rapido, se pueden ver las siguientes alternativas:

(i) Consumir con tasa $u(t) = k\,x(t)$, en cuyo caso preservamos el capital inicial x_0 dado que tenemos $\dot{x} = 0$.

(ii) Consumir menos inicialmente, $u(t) < k\,x(t)$, de manera que el capital crezca y posteriormente se pueda incrementar el consumo.

(iii) Consumir más inicialmente cuando el valor temporal del dinero $e^{-\rho t}$ es mayor.

Ejemplo 5

Modelo de propaganda de Vidale-Wolfe,(1957).
Una firma quiere maximizar sus ganancias mediante un control adecuado de los gastos de propaganda. Consideremos que la firma E vende un producto determinado con tasas de venta $x(t)$. En un mercado competitivo las tasas de venta de la firma E corresponden a una porción del mercado total que asumimos es contante e igual a 1. Las tasas de venta del mismo producto, pero con otras marcas, que realiza la competencia resultan ser de $1 - x(t)$. La propaganda de la empresa E está dirigida hacia la porción del mercado $1 - x(t)$ que mantiene la competencia.
Designemos por $u(t)$ la tasa de gastos en promoción y publicidad del producto por parte de la empresa E. Entonces el modelo para las ventas resultantes como respuesta a la propaganda realizada, según Vidale-Wolfe, es

$$\dot{x} = \alpha\,(1 - x)\,u - \beta\,x,$$

con las condiciones siguientes:

(i) el intervalo de tiempo $0 \leq t \leq T$ es previamente fijado,

(ii) $x(0) = x_0$ es el nivel inicial de las tasas de ventas de la empresa E, con $0 < x_0 < 1$,

(iii) $\alpha\,y\,\beta$ son constantes positivas conocidas,

(iv) la tasa de gastos en propaganda es no-negativa, es decir, $u \geq 0$.

Este modelo ha sido obtenido por deducción empírica. Obsérvese que la tasa de variación de $x(t)$ es positivamente afectada por los gastos $u(t)$ en promoción del producto. Por otro lado, cuanto mayor es $x(t)$, menor es la tasa de crecimiento $\dot{x}(t)$. El término $-\beta\, x$ corresponde al hecho de que si $u(\cdot) \equiv 0$, entonces las ventas decaerán exponencialmente.
El problema de control óptimo consiste en maximizar la siguiente expresión que tiene en cuenta los ingresos y los egresos

$$\int_0^T e^{-\rho t}[\gamma\, x(t) - u(t)]\, dt.$$

donde γ y ρ son constantes positivas conocidas. El factor $e^{-\rho t}$ es llamado el valor temporal del dinero.
En la sección 5 regresaremos a este ejemplo presentando su solución vía programación dinámica. Generalizaciones de este problema aparecen en [5]-[10].

3 Control óptimo

3.1 El sistema de control

Consideremos el sistema de control sobre el intervalo $[t_0, T]$

$$\dot{x} = f(x(t), u(t)), \quad t_0 \leq t \leq T, \tag{6}$$

con el valor inicial $x(t_0) = x_0 \in \mathbb{R}^n$.
La función $u(\cdot)$ es llamada *control* y en general pertenece a una familia de funciones $\mathcal{U}$, la cual se define de una manera adecuada para que tenga sentido en una aplicación determinada y para garantizar la existencia y unicidad de la solución de (6).
La función f es llamada la *dinámica del sistema de control* (6) y también debe gozar de condiciones suficientes a los efectos de garantizar la existencia y unicidad de la solución de (6) una vez que $u(\cdot)$ es elegido en $\mathcal{U}$.
A continuación exponemos las condiciones bajo las cuales estarán definidas en el presente trabajo tanto la función f como la familia $\mathcal{U}$.
(i) La familia $\mathcal{U}$ se define como

$$\mathcal{U} = L([t_0, T], U),$$

el espacio de todas las funciones definidas en el intervalo $[t_0, T]$, medibles según Lebesgue, y cuyo rango está contenido en U, un subconjunto compacto de $\mathbb{R}^n$, llamado *el conjunto de valores del control*.

(ii) La función $f : \mathbb{R}^n \times U \longrightarrow \mathbb{R}^n$ es considerada como una función de dos variables, a saber, $(x, u) \in \mathbb{R}^n \times U$, donde x se denomina *la variable de estado*, y u se denomina *la variable de control.* La función f se asume continua y satisface la condición de Lipschitz como función de la variable de estado $x \in \mathbb{R}^n$, uniformemente en u. Esto es, existe una constante $K > 0$ tal que

$$|f(x,u) - f(y,u)| \leq K\,|x - y|, \quad \forall x, y \in \mathbb{R}^n,\ u \in U.$$

Dado un valor inicial $x(t_0) = x_0$ y un control $u(\cdot) \in \mathcal{U}$, las condiciones anotadas arriba garantizan la existencia y unicidad de la solución $x(\cdot)$ del sistema de control (6) asociada a dicho valor inicial y al mencionado control. Entonces $x(\cdot)$ es solución de la ecuación integral

$$x(t) = x_0 + \int_{t_0}^{t} f(x(s), u(s))\, ds, \quad t \in [t_0, T],$$

donde $x(t)$ es llamado el estado del sistema de control (6) en el instante t . En particular, $x(\cdot)$ resulta ser una función absolutamente continua que resuelve (6) en el intervalo $[t_0, T]$, excepto en un conjunto de medida cero.
Debemos observar, que en las condiciones anotadas arriba, los controles resultan ser uniformemente acotados, dado que el conjunto de valores del control U es compacto. Sin embargo, existen otros problemas de control para los cuales el conjunto de valores del control U no es acotado. Es decir, hay formas mas generales de plantear un sistema de control orientadas hacia la solución de otros problemas que ameriten esa generalidad.

3.2 La funcional de costo

Dada una función $L : \mathbb{R}^n \times U \longrightarrow \mathbb{R}$ continua, podemos definir una funcional llamada *costo* de la manera siguiente

$$u(\cdot) \in \mathcal{U} \longrightarrow J^{u(\cdot)}(t_0, x_0) = \int_{t_0}^{T} L(x(s), u(s))\, ds, \tag{7}$$

donde $x(\cdot)$ es la solución de (6) asociada al control $u(\cdot)$ con valor inicial $x(t_0) = x_0$. La función L es llamada *Lagrangiano.*

3.3 El problema de control óptimo

El problema de control óptimo consiste en minimizar sobre todos los controles $u(\cdot) \in \mathcal{U}$ la funcional (7). Entonces, dado un instante inicial t_0 y un valor inicial $x(t_0) = x_0$, el problema de control óptimo consiste en determinar el valor de la expresión

$$\inf_{u(\cdot)\in\mathcal{U}} \{ \int_{t_0}^{T} L(x(s), u(s))\, ds \},$$

y en determinar, en caso de que exista, un control $u^*(\cdot)$, llamado *control óptimo*, que realice este valor mínimo. En ese caso, la correspondiente solución $x^*(\cdot)$ es llamada *trayectoria óptima*, y se tiene

$$\begin{aligned} J^{u(\cdot)}(t_0, x_0) &\geq J^{u^*(\cdot)}(t_0, x_0), \quad \forall u(\cdot) \in \mathcal{U}, \\ \frac{d}{dt} x^*(t) &= f(x^*(t), u^*(t)) \quad a.e., \quad x^*(t_0) = x_0. \end{aligned}$$

Observación

Si el problema de control óptimo consiste en maximizar sobre todos los controles $u(\cdot) \in \mathcal{U}$ la funcional (7) (ahora llamada de *réditos*) entonces el control óptimo es el que maximiza

$$J^{(t_0,x_0)}(\cdot),$$

y se tiene

$$J^{u(\cdot)}(t_0, x_0) \leq J^{u^*(\cdot)}(t_0, x_0), \quad \forall u(\cdot) \in \mathcal{U}.$$

La existencia y unicidad del control óptimo son problemas importantes estudiados en [11] y [13].

3.4 La función de valor

Considerando al tiempo inicial y a la condición inicial para el estado como variables, podemos definir una función $V : [t_0, T] \times \mathbb{R}^n \longrightarrow \mathbb{R}$, llamada la *función de valor*, de la siguiente forma

$$V(t, x) = \inf_{u(\cdot)\in\mathcal{U}} J^{u(\cdot)}(t, x). \tag{8}$$

El nombre '*valor*' viene de la interpretación económica del problema de minimizar el ' *costo* ' $J^{\cdot}(t, x)$. En este problema de control óptimo, $V(t, x)$ es lo mínimo que J puede costar dados el instante inicial t y el estado inicial x. Es claro que si $u^*(\cdot)$ es un control óptimo, entonces

$$V(t, x) = J^{u^*(\cdot)}(t, x).$$

Además se tiene la condición terminal $V(T, x) = 0, \quad \forall x \in \mathbb{R}^n$.

Observación

Si el problema de control óptimo consiste en maximizar la funcional (7), entonces la función de valor es definida por

$$V(t,x) = \sup_{u(\cdot)\in\mathcal{U}} J^{u(\cdot)}(t,x). \tag{9}$$

4 El método de programación dinámica

La idea fundamental del método de programación dinámica es que la función de valor V satisface una ecuación funcional llamada el '*principio de programación dinámica'*

$$V(t,x) = \inf_{u(\cdot)\in\mathcal{U}} \{\int_t^{t+\delta} L(x(s),u(s))\,ds \;+\; V(t+\delta, x(t+\delta))\},\ \forall\delta > 0. \tag{10}$$

Una demostración de este principio puede leerse en [12].
Cuando la función de valor V es suficientemente diferenciable, la versión infinitesimal del principio de programación dinámica es la ecuación no lineal en derivadas parciales de primer orden llamada ' *ecuación de Hamilton-Jacobi-Bellman'* (HJB)

$$V_t(t,x) + H(x, V_x(t,x)) = 0, \quad (t,x) \in [t_0, T] \times \mathbb{R}^n, \tag{11}$$

donde

$$H(x,p) = \inf_{u\in U} \{f(x,u)\cdot p + L(x,u)\}\,.$$

El método de programación dinámica reduce entonces el problema de control óptimo, que es un problema de minimización en el espacio de dimensión infinita $\mathcal{U}$, a resolver una ecuación no lineal en derivadas parciales de primer orden. Esta ecuación contiene toda la información relevante para configurar, a traves del '*procedimiento de síntesis*', un control óptimo de tipo 'feed-back' para el problema de control original.

Teorema 1. *Consideremos el problema de control óptimo determinado por* (6)*,* (7) *y* (8)*. Asumamos que la función de valor V es C^1. Entonces la función de valor V es solución de la ecuación no lineal en derivadas parciales de primer orden HJB* (11)*.*

Por razones didácticas presentaremos una demostración no rigurosa de este teorema. Esta demostración se puede hacer rigurosamente aún para el caso en que la función de valor no es diferenciable usando una noción de solución débil llamada 'solución de viscosidad'. Para ver detalles sobre esta noción referirse a [2], [3], [4] y [12].

Demostración *(Esbozo):* Consideremos el problema de control (6), (7), y (8) y asumamos que la función de valor V es C^1. Sea $(t,x) \in [t_0,T] \times \mathbb{R}^n$. Sea $u(\cdot)$ un control arbitrario, y sea $x(\cdot)$ la correspondiente solución con valor inicial $x(t) = x$. El Teorema de Taylor nos permite escribir

$$\begin{aligned} V(t+\delta, x(t+\delta)) &= V(t,x(t)) + V_t(t,x(t))\delta + V_x(t,x(t)) \cdot \dot{x}(t)\delta + o(\delta), \\ &= V(t,x) + V_t(t,x)\delta + V_x(t,x) \cdot f(x,u(t))\delta + o(\delta). \quad (12) \end{aligned}$$

Reemplazando $V(t+\delta, x(t+\delta))$ en el principio de programación dinámica (10) y cancelando $V(t,x)$ se obtiene

$$\begin{aligned} 0 &= \inf_{u(\cdot)\in\mathcal{U}} \left\{ \int_t^{t+\delta} L(x(s),u(s))\,ds + V_t(t,x)\delta + V_x(t,x)\cdot f(x,u(t))\delta \right\} + o(\delta), \\ &= \inf_{u(\cdot)\in\mathcal{U}} \left\{ \int_t^{t+\delta} L(x(s),u(s))\,ds + V_x(t,x)\cdot f(x,u(t))\delta \right\} + V_t(t,x)\delta + o(\delta). \end{aligned}$$

Dividiendo por δ

$$0 = \inf_{u(\cdot)\in\mathcal{U}} \left\{ \frac{1}{\delta} \int_t^{t+\delta} L(x(s),u(s))\,ds + V_x(t,x)\cdot f(x,u(t)) \right\} + V_t(t,x) + \frac{o(\delta)}{\delta}.$$

Haciendo tender δ a cero

$$0 = \inf_{u(\cdot)\in\mathcal{U}} \{ L(x,u(t)) + V_x(t,x)\cdot f(x,u(t)) \} + V_t(t,x),$$

de donde se deduce

$$V_t(t,x) + \inf_{u\in U} \{ L(x,u) + V_x(t,x)\cdot f(x,u) \} = 0$$

□

Observación

Consideremos el problema de control óptimo determinado por (6), (7) y (9). Entonces tenemos la nueva versión del Teorema 1

Teorema 2. *Si la función de valor V es C^1, entonces V es solución de la ecuación no lineal en derivadas parciales de primer orden HJB*

$$V_t(t,x) + H(x, V_x(t,x)) = 0, \quad (t,x) \in [t_0, T] \times \mathbb{R}^n, \tag{13}$$

donde

$$H(x,p) = \sup_{u \in U} \left\{ f(x,u) \cdot p + L(x,u) \right\}.$$

5 Aplicaciones

En esta sección daremos solución a algunos de los problemas planteados en la Sección 2 usando el método de programación dinámica.

5.1 Reacción química

Consideremos nuevamente el Ejemplo 3 presentado en la Sección 2 el cual consiste en una reacción química que produce un producto cuya calidad final depende del valor del pH. Si $x(t)$ denota la desviación del pH deseado, el sistema de control está dado por

$$\dot{x}(t) = \alpha\, x(t) + \beta\, u(t), \quad x(0) = x_0,$$

con las condiciones

(i) α y β son constantes positivas conocidas.

(ii) $x(0) = x_0$ es la desviación inicial del pH deseado.

(iii) El intervalo de tiempo $0 \leq t \leq T$ es previamente fijado .

(iv) El ingrediente de control u no es acotado.

La función de costo está dada por

$$J^{u(\cdot)}(t_0, x_0) = \int_0^T (K\, x^2(s) + u^2(s))\, ds,$$

donde K es una constante conocida.
La función de valor está definida por

$$V(t,x) = \inf_{u(\cdot) \in \mathcal{U}} \int_t^T (K\, x^2(s) + u^2(s))\, ds,$$

donde hemos considerado a la condición inicial como variable $x(t) = x$. De acuerdo con el Teorema 1, si la función de valor V es C^1, ella satisface la ecuación no lineal de primer orden en derivadas parciales HJB (11)

$$\begin{aligned} 0 &= V_t(t,x) + \inf_{u\in U} \left\{[\alpha\, x + \beta\, u] V_x(t,x) + K\, x^2 + u^2\right\}, \\ &= V_t(t,x) + \alpha\, x\, V_x(t,x) + K\, x^2 + \inf_{u\in U} \left\{\beta\, u\, V_x(t,x) + u^2\right\}. \qquad (14) \end{aligned}$$

Primero minimizamos la función cuadrática

$$u \longrightarrow u^2 + \beta\, u\, V_x(t,x),$$

cuyo mínimo es alcanzado en

$$u^* = \frac{-\beta V_x}{2}. \qquad (15)$$

Reemplazando el valor de u^* dado por (15) en la ecuación (14), obtenemos

$$\begin{aligned} 0 &= V_t(t,x) + K\, x^2 + \alpha\, x\, V_x(t,x) + \frac{\beta^2}{4} V_x^2(t,x) - \frac{\beta^2}{2} V_x^2(t,x), \\ &= V_t(t,x) + K\, x^2 + \alpha\, x\, V_x(t,x) - \frac{\beta^2}{4} V_x^2(t,x). \qquad (16) \end{aligned}$$

Ahora, asumimos que la solución de (16) es de la forma

$$V(t,x) = C(t) x^2$$

y tratamos de determinar $C(t)$. Las derivadas parciales son

$$V_x(t,x) = 2\, C(t)\, x, \qquad (17)$$

$$V_t(t,x) = C'(t)\, x^2. \qquad (18)$$

Reemplazando (17) y (18) en la ecuación (16) se obtiene

$$0 = C'(t)\, x^2 + K\, x^2 + 2\, \alpha\, C(t)\, x^2 - \beta^2\, C(t)^2\, x^2.$$

De aquí se obtine la ecuación de Ricatti

$$C'(t) = -K \;-\; 2\, \alpha\, C(t) \;+\; \beta^2 C(t)^2, \qquad (19)$$

con condición terminal $C(T) = 0$ que se deduce de $C(T)\, x^2 = V(T,x) = 0$. Combinando (15) y (17), el control óptimo está dado en términos de la solución de esta ecuación de Riccatti (19) por la forma

$$u^*(t) = -\beta\, x(t)\, C(t).$$

5.2 Modelo de propaganda de Vidale-Wolfe

Consideremos nuevamente el Ejemplo 5 presentado en la Sección 2 el cual consiste en tratar de maximizar las ganancias de una empresa mediante un control adecuado de los gastos en propaganda. El sistema de control según Vidale-Wolfe es

$$\dot{x}(t) = \alpha\,(1 - x(t))\,u(t) - \beta\,x(t), \quad x(0) = x_0, \tag{20}$$

bajo las condiciones siguientes:

(i) El estado $x(t)$ representa la tasa de ventas como fracción del mercado total.

(ii) El control , $u \geq 0$, representa la tasa de los gastos en propaganda.

(iii) El conjunto U donde los controles toman sus valores no es acotado.

(iv) $\alpha\, y\, \beta$ son constantes positivas conocidas.

(v) $0 < x_0 < 1$ (lo cual implica dada (20) que $0 < x(t) < 1$).

(vi) El intervalo de tiempo $0 \leq t \leq T$ es fijo.

La funcional de las ganancias está dada por

$$J^{u(\cdot)} = \int_0^T e^{-\rho t}\,[\gamma\,x(t) - u(t)]\,dt.$$

donde γ y ρ son constantes positivas conocidas. La función de valor está definida por

$$V(t,x) = \sup_{u(\cdot)\in\mathcal{U}} \int_t^T e^{-\rho s}\,[\gamma\,x(s) - u(s)]\,ds,$$

y satisface la ecuación diferencial ordinaria HJB (13) no lineal de primer orden

$$\begin{aligned} 0 &= V_t(t,x) + \sup_{u\in U} \left\{[\alpha\,(1-x)\,u - \beta\,x]V_x(t,x) + e^{-\rho t}\,[\gamma\,x - u]\right\} \\ &= V_t(t,x) - \beta\,x\,V_x(t,x) + \gamma\,x\,e^{-\rho t} + \sup_{u\in U} \left\{\alpha\,(1-x)\,u\,V_x(t,x) - u\,e^{-\rho t}\right\} \end{aligned} \tag{21}$$

El control óptimo resulta de maximizar la expresión

$$u \longrightarrow [\alpha\,(1-x)\,V_x(t,x)\; - \; e^{-\rho t}]\,u.$$

Dado que aquí $U = [0, \infty)$, el control óptimo es

$$u^*(t) = \begin{cases} 0, & \text{si } \alpha\,(1-x)\,V_x(t,x) \,<\, e^{-\rho t}, \\ \infty, & \text{si } \alpha\,(1-x)\,V_x(t,x) \,>\, e^{-\rho t}, \\ \text{indeterminado}, & \text{si } \alpha\,(1-x)\,V_x(t,x) \,=\, e^{-\rho t}. \end{cases}$$

Hay entonces tres casos que analizamos a continuación.
Para el caso en que u^* no fué determinado, es decir, si se tiene

$$\alpha\,(1-x)\,V_x(t,x) \,=\, e^{-\rho t}, \tag{22}$$

entonces de la ecuación (21) se deduce

$$V_t(t,x) + \gamma\, x\, e^{-\rho t} \,-\, \beta\, x\, V_x(t,x) = 0. \tag{23}$$

Asumiendo $V \in C^2$, diferenciando (23) con respecto a x y reemplazando los valores de V_x, V_{xx} y V_{xt} que se deducen de (22) se obtiene

$$\alpha\gamma(1-x)^2 - \rho(1-x) - \beta = 0. \tag{24}$$

Esta es una ecuación cuadrática para x con coeficientes que no dependen de t. Por lo tanto, una trayectoria para la cual (22) se satisfaga debe ser constante

$$x^*(t) \equiv x^c \;\; \text{(constante !)}. \tag{25}$$

Observe que como $x^*(t) \equiv x^c$ es una solución del sistema de control (20) con $u(t) = u^*(t)$, entonces

$$0 = \alpha\,(1-x^c)\,u^*(t) - \beta\, x^c,$$

de donde determinamos u^* para este caso como

$$u^*(t) = \frac{\beta\, x^c}{\alpha\,(1-x^c)},$$

el cual es también una constante.
En el caso en que el control óptimo es cero, las trajectorias de (20) son simplemente exponeciales decrecientes. Estas trajectorias son óptimas cuando $x_0 > x^c$ y hasta el momento en que $x^*(t)$ intersecta x^c.
Por último, si la condición inicial satisface $x_0 < x^c$ entonces el control óptimo resulta ser impulsivo. Para más detalles, ver [9].
Para concluir con la demostración de que los controles y trajectorias determinadas aquí son óptimas, es necesario aplicar el teorema de verificación. Para esto ver [11].

Referencias

[1] Bliss, G. A., *Calculus of Variations*, Mathematical Association of America, Carus Mathematics Monographs no 1, 1925.

[2] Bardi, M., Capuzzo, D., *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhäuser, Boston, 1997.

[3] Crandall, M.G., Ishii, H., Lions, P.-L., *User's Guide to Viscosity Solutions of Second Order Partial Differential Equations*, Bulletin of the American Mathematical Society, **32** (1994), 1322–1331.

[4] Crandall, M.G., Lions, P.-L., *Viscosity Solutions of Hamilton-Jacobi Equations*, Transactions of The American Mathematical Society, **277** (1984), 1–42.

[5] Dorroh, J. R., Ferreyra, G., *Optimal Advertising in Growing-Stabilizing Markets*, Optimal Control Applications and Methods, **14** (1993), 221–228.

[6] Dorroh, J. R., Ferreyra, G., *Optimal Advertising in Exponentially Decaying Markets*, Journal of Optimization Theory and Applications, **79**, 2 (1993), 219–236.

[7] Dorroh, J.R., Ferreyra, G., *A Multi-state, Multi-control Problem with Unbounded Controls*, SIAM Journal on Control and Optimization, **32**, 5 (1994), 1322–1331.

[8] Edie, R.D., *On the Optimal Control of the Vidale-Wolfe Advertising Model*, Optimal Control Appl. Methods, **18** (1997), 59–72.

[9] Ferreyra, G., *The Optimal Control Problem for the Vidale-Wolfe Advertising Model Revisited*, Optimal Control Appl. and Methods, **11** (1990), 363–368.

[10] Ferreyra, G., Hijab, O., *A Simple Free Boundary Problem in* $\mathbb{R}^d$, SIAM Journal on Control and Optimization, **32**, 2 (1994), 501–515.

[11] Fleming, W.H., Rishel, R.W., *Deterministic and Stochastic Optimal Control*, Springer-Verlag, New York, 1975.

[12] Fleming, W.H., Soner, H.M., *Controlled Markov Processes and Viscosity Solutions*, Applied Mathematics, Vol 25, Springer-Verlag, New York, 1993.

[13] Knowles, G., *An Introduction to Aplied Optimal Control*, Mathematics in Science and Engineering, Academic Press, **159**, New York, 1981.

[14] Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mischenko, E.F., *The Mathematical Theory of Optimal Processes*, Interscience, 1962.